Disaster Scenario and Record Capture System

Heng-Chih Hsieh, Hung-Fu Chen, and Jane Win Shih Liu, Life Fellow, IEEE
Institute of Information Science
Academia Sinica
Taipei, Taiwan
{hunter205,jeff, janeliu}@iis.sinica.edu.tw

Abstract—**Typical reports on past disasters contain valuable lessons that can help us improve our preparedness and response strategies and operations, especially with the help of modern analysis, simulation and visualization tools. However, disaster historical records are typically written in natural language and for human consumption. They often contain inconsistent data and information, which take time and effort to uncover. Worse yet, further analysis based on report information is almost impossible. DiSRC (Disaster Scenario and Record Capture) system is being designed as a solution of these problems. The system contains a cloud-based authoring system together with software tools running on mobile devices for capturing observational data records in a machine-readable form. The quality of the captured records is controlled and assured by both the capturing tools in real-time and by tools in the cloud. The authoring system can assist authors to write reports based on the captured data and tools for processing the data such as searching, analyzing, etc.**

Keywords—disaster; quality; mobile; database;

I. INTRODUCTION

Today, disaster historical records are typically written in natural language text together with tables, graphs, and images. In their current form, disaster records have two shortcomings. Data and information may be inconsistent and incomplete, and such defects take time and effort to uncover. Typical reports on past disasters contain valuable lessons that can help us improve our preparedness and response strategies and operations, especially with the help of modern analysis, simulation and visualization tools. However, it is nearly impossible to extract from the existing reports machine-readable input needed to use the tools. Moreover, quality assurance of observation data is often overlooked [1]. Up until 2013, only 17% of databases have applied some quality control procedure, and some of them are verified on a semi-annual basis [2]. Without data quality control, the validity and reliability of vulnerability and risk assessment based on data are questionable.

This fact motivated the proposed DiSRC (disaster scenario and record capture) system. The system contains as components mobile Disaster Record Capture Tool (DiReCT) and a cloud-based authoring system. Primary users of DiReCT are government agents who are responsible for contributing data to be included in disasters records and volunteers trained by the responsible agencies to do the work. They use DiReCT to capture observational data and related information on the dynamics and development of the disaster and its effects on places and people in a machine-readable form. The authoring system can assist authors to write reports on the disaster based captured data and information. Disaster records produced with the help of the system can be read and processed by search, analysis, simulation and visualization tools.

In addition to producing disaster records in machine-readable formats, a distinguishing capability of the proposed system is quality assurance and control of data in disaster historical records. DiReCT can detect in real-time basic errors (such as typos, incorrect capture time and location, and poor time and space resolutions) and take follow-up measures, such as reminding the user to repeat the observation, make revisions of the capture record or marking the record as defective. Once the record is uploaded to the authoring server, the system will make global consistency check by comparing records captured by different users during the current disaster event and records captured during similar past events. The authoring system provides a library of diverse tools to enable the integration and analysis of all observational records uploaded from all DiReCTs and maintained the records for purposes of developing better response strategies, providing decision support during emergencies, post-disaster analysis, and so on.

Following this introduction, Section II presents related work. Sections III, IV and present an overview of DiSRC design rationales and details on DiReCT and the authoring systems, respectively. Section V presents the current status and future work to complete DiSRC system.

II. RELATED WORK

Over the past few years, a growing number of tools and digital data capture systems had been developed to eliminate shortcomings of paper-based disaster information. These tools [3-20] support the collection, storage, and analysis of observational data. Using these tools, project administrator can easily generate questionnaires with the help of web-based user interface and send them to data collectors' mobile devices. After data collectors complete their forms, the observational data contained in them are stored on platforms where data can be viewed online. Through analysis and visualization tools provided by the digital data capture system, data are presented on maps or in charts, and can be easily analyzed. Some of them are hazard-specific [13, 14, 18, 19] and others can be used on any type of disasters depending on the questionnaires used by data collectors.

Most of the tools offer custom field-based verification and basic workflow for data quality control. Project administrator can specify input type of each field, such as integer, string, or multi-choice, and also the time when questions should be visible or invisible during questionnaire design process. These tools will automatically check whether the record meets specified quality requirements, including whether each answer is in a right data type or any missing answer exists. Fulcrum [15] has a more powerful quality control mechanism: Administrator can program their own custom quality assurance logic through REST APIs and code libraries. In the information flows of GDACSmobile [20], observation submitted by moderated users will enter a quality control loop, while observation submitted by non-moderated user will not enter the loop. Through the loop, moderator can manually review the observation and, if needed, give to the user a response such as extending the questionnaire, clarifying uncertain answer, or asking for further information.

To the best of our knowledge, DiSRC is the first platform that can capture observational data in a machine-readable format together with customizable and rigorous data quality assurance and control. Together, the system of DiReCTs and the authoring system support pre-collection data quality assurance, real-time quality control during collection, post-collection quality as specified by a quality-control SOP (standard operating procedures). Best practices during data collection are also specified and enforced through SOPs which are downloaded from the authoring system to DiReCTs prior and during the collection process along with historical records for data validation usage. Specifically, the SOPs define the DiReCT application state-machine workflow, user interface, and data quality control logic. With the design of SOP-based execution, DiReCT can be made flexible and extensible, capable of turning itself to a hazard-specialized data capturing system based on any disaster that SOP is designed for.

III. Design Rationales

As shown in Fig. 1, the DiSRC system contains as components a large number of DiReCT (Disaster Record Capture Tool). Each DiReCT is a system of software tools running on a tablet, smart phone or some other mobile platform. The other major component is the authoring system (AS) that runs on a cloud-based server. DiSRC does not crowdsource data collection. However, as the upper-right corner of the figure illustrates, the system and its users can access crowdsourced sensor data, such as earthquake intensity data, surveillance sensor data, and so on.

As stated earlier, an objective of DiSRC is to make captured data directly readable by tools. The system collects and organizes disaster historical data and information from DiReCT and crowdsourced sensor data. Authors who write disaster historical records and reports for human reader are able to access data via web-services, visualization tools, file systems, and so on. The system provides them with tools for translating data into forms (e.g., spread sheets, tables and graphs) that can be included directly as parts of their reports.

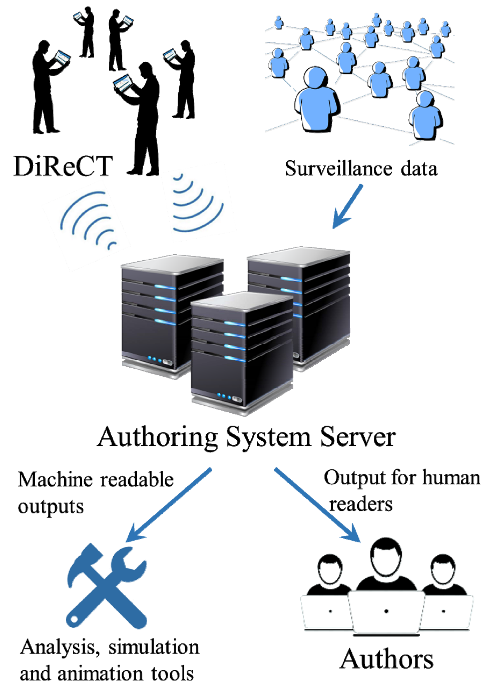

Fig. 1. DiSRC system overview

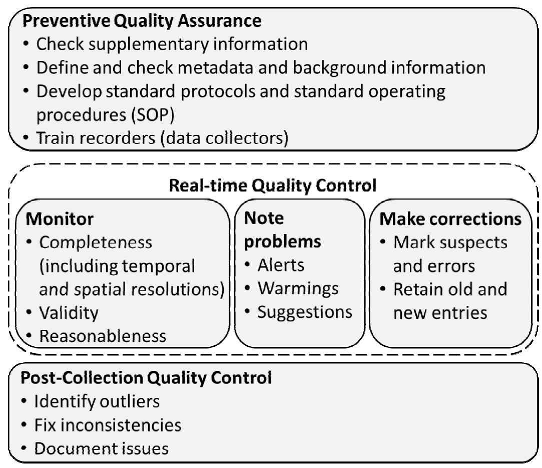

Fig. 2. Data quality assurance (QA) and quality control (QC) for disaster data recording

Also stated earlier, data quality control is a distinguishing feature of DiSRC. Following published recommendations such as [21-23], DiSRC divides data quality into quality assurance (QA) and quality control (QC) as illustrated by Fig. 2. The former refers to activities that take place before data collection begins; the latter refers to activities that take place during and after data collection. To be specific, the system provides supports including monitoring the capture of observation data, detecting errors in real-time, alerting the user of the errors and overseeing corrections are made, checking inconsistencies between records, and so on.

IV. DISASTER RECORD CAPTURE TOOL (DIRECT)

Each DiReCT is a mobile system of advanced IT tools. It is used by one user (referred to as a recorder in Fig. 2) at a time during disaster preparedness and response times to capture data and information on the dynamics and development of the disaster and its effects on places and people in a machine-readable form. As its requirements, the mobile system has sufficient storage capacity and compute and communication power to be used standalone or in collaboration with the cloud-based authoring system.

A. Support Data/Information and Tools

Each DiReCT can provides its user with local data and contextual information as shown in Fig. 3 to enable him/her to enter observation data and record occurrences of events and phenomena (e.g., rain stopped, debris flow starts, etc.) in context-specific ways: Background (e.g., census data), scientific, and GIS (Geographical Information System) information are types of information about the region and the type of emergency event which the recorder may need. Historical disaster records are data on similar and related types of disasters that have occurred in the region and their impacts on the region. Metadata are data about disaster data type and format and parameters for data quality control, as well as data and information needed to provide the user with on-line helps at all times.

SOPs are one or more standard operating procedures that define and enforce best practices during the data capture process and how RTQC (real-time quality control) is conducted. They were downloaded from the cloud during initialization of the device and in response to change notifications from the authoring system as a part of quality assurance. In particular, each DiReCT monitors the capture process and captured data and automatically alert the user of anomalies (e.g., missing data or conflicting data) to be tracked down, resolved, and corrected as much as possible in ways specified by the SOPs. Tools such as location/navigation tools, calendar time and timers, and several hardware sensors provide essential data/information for record data capturing and quality control.

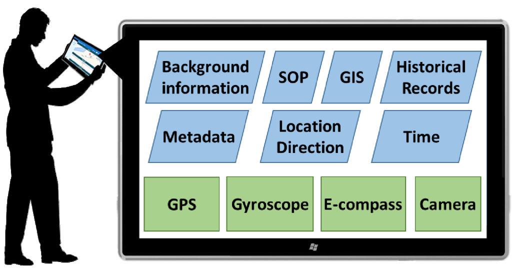

Fig. 3. Local data/information and tools held by DiReCT

B. Major Components and Software architecture

The block diagram in Fig. 4 depicts the major components and software structure of DiReCT. The user interacts with DiReCT via its User Interface (UI): This is where the user enters observational data and carries out associated tasks. A dedicated UI thread is used to keep the system responsiveness. In addition, the interaction between UI and other components in DiReCT are all asynchronous so as to avoid blocked user interface. Specifically, the UI is supported by User Interface Generator (UIG) that dynamically generates and sequences the displays presented to the user according to the UI configuration file and SOP configuration file.

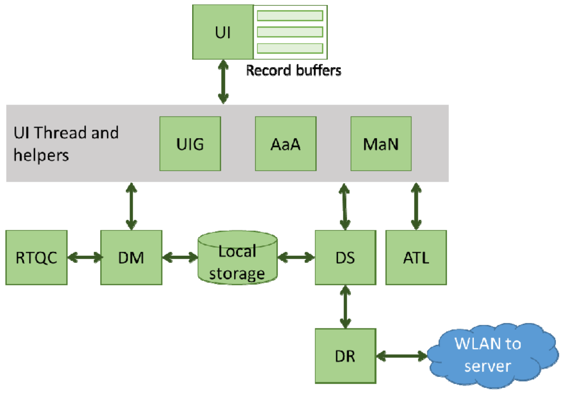

Fig. 4. Software architecture of DiReCT

Authentication and Authorization (AaA) module authenticates the user during login. It checks the user for authorization to download event data and upload record data to and from DiReCT. Monitor and Notification (MaN) is a runtime monitoring tool. It enforces the SOPs governing time and spatial resolutions of the capture observational record. In addition, it provides in capability of the real-time monitoring, capture and analysis of events and conditions that indicate the potential for occurrences, or actual occurrences, of errors, and issuing alerts and notifications to trigger error handling or prevention actions. It also provides notification functions other modules can call to send reminders and alerts. The Real-time Quality Control (RTQC) module examines the record metadata and input data during the data collection. When it detects defects in real-time, it may alert the user of the errors and oversee the corrections, or mark the record as defective, or even ignore the defect, and on in ways defined by the SOP governing how RTQC is required to be done.

Observational records captured by the user are stored, maintained and transferred under the control of Data Manager (DM) module. It functions include overseeing the examination by RTQC of each record stored temporarily in the UI record buffers and putting records found free of defects in one or more collections of records captured by the user (and handling defective records in ways specified by SOP) and then storing the collections in the local storage. All other modules access the event data, user data and record data in the local storage using the API functions of the DM module. In particular, it works with the Data Synchronizer (DS) to upload record collections stored in the local storage during or after data collection and download event data (work descriptions, UI configuration, SOP configuration, map data, user data, historical records, and quality control data) and user data (profile and settings) before data collection.

Data Refresher (DR) is responsible for data transmission between DiReCT and the AS (Authoring System) server. DR guarantees reliable data transmission and specified quality of service when network is available. When the network is down, it is responsible for the completion of all data transmissions requested during network outage as well as yet-to-be-completed transmissions requested before network outage. Finally, Assistant Tool Loader (ATL) load assistant plugins during the startup of the application. Examples of assistant tools include tools for estimating the number of people through camera image, accurate locations of victims or objects at unreachable places, boundaries and sizes of disaster-impacted areas, and so on.

C. DiReCT Prototype

In our very first DiReCT prototype, the configurability of the device is achieved mostly by providing it with a GUI composed of a WPF (Windows Presentation Foundation) plus Gmap.NET component and a state-machine workflow component. The former presents to the user with display and user interface elements, including input controls (e.g., checkboxes, buttons, list boxes and text fields), navigational elements (e.g., sliders, search fields, tags and icons), informational elements (e.g., icons, progress bar, notifications, message boxes, and modal windows) and containers. The DiReCT state-machine workflow defines formally user-device interactions (i.e., changes in display and available controls and actions taken by the device in response to user actions).

The pre-defined workflow (will be contained in SOPs in the future) is loaded during the initialization phase. The UI then operate following the workflow. Each module works asynchronously on threads independently and communicates through wrapping works into work items and inserting work items to work queues own by each module, and through raising events to each other.

V. AUTHORING SYSTEM

A purpose of AS is to help authors generate historical records in human-readable and machine-readable forms making use of observational data captured by all recorders during and after disaster events. AS also is where disaster-event-specific and user-specific SOPs governing data capture practices and RTQC are maintained, and preventive data quality assurance and post-collection data quality control are carried out. It aims to support downloading to DiReCTs SOPs and contextual data and information as a part of preventive data quality assurance and opportunistic upload of captured data from DiReCTs; integrate data uploaded from all DiReCTs and does post-collection data quality control; provides access to an extensible library of search, analysis, simulation, visualization and data mining tools and tools for translating captured data into input formats of the tools; provides a virtual repository of available disaster historical records with APIs and web services for data and information discovery and retrieval by human users, and an open data repository for maintaining future historical records generated by DiSRC system.

A. Components of Authoring System

As Fig. 5 shows, the authoring system (AS) has five major components. First, it presents to authors and tools a virtual repository of data and information on past disasters, geographical information, related scientific information, and real-time data and information on recent emergencies. Datasets containing these data and information reside physically in data centers operated by government agencies and non-government institutions in Taiwan and abroad, including Central Weather Bureau and GIS Center in Taiwan, NOAA in USA and numerous universities worldwide. Tools such as web crawler, format converters, access control and interoperability services shown in the upper-left box in the figure can significantly reduce the efforts required in accessing widely distributed data and make use of them.

The second major AS component is an open data database/repository. In the best-case scenario, AS will be adopted and widely used. It should be one of the sites where disaster history records will be maintained. This repository is being designed and implemented like an open data site for creation and maintenance disaster historical records in the future.

The third component is a database that holds libraries of SOPs guiding observational data capture process, data for quality assurance, and real-time quality control for each of the types of disaster/emergency event supported by DiReCT. The DiReCTs used for capturing data on each type of emergency by each government agency and department are configured and customized according to the requirement specification of the devices. Ideally, requirement specifications of DiReCTs are executable, in terms of workflows, for example. This database maintains the specifications and other parameters needed for configuration and customization of DiReCTs.

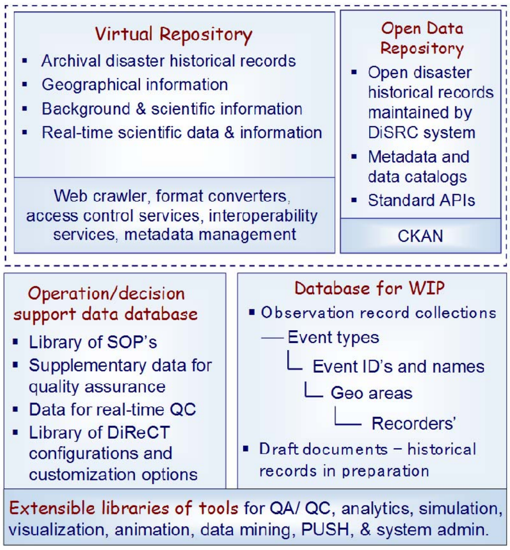

Fig. 5. Major components of authoring system

The fourth component is the WIP (Work-in-Progress) database that holds observation records uploaded from DiReCTs during and after emergencies. The records on each emergency event are held here until they are processed and used to generate history records and reports. At that time, parts of them will be moved to the open data repository and other archives depending on the choice of the government agency who owns the data. As Fig. 5 shows, within the observation record database, the records are organized hierarchically. For example, the tree of the record collections starts from event type, individual event, geographical region and time interval, and then collections of records contributed by individual recorders. Information on recorders can be stored here optionally. This database also maintains documents generated from the records, including capabilities version control and collaborative authoring.

The fifth component contains extensible library of tools. Fig. 5 lists examples. The system has capabilities for authors to use libraries of tools remotely, assisting them to write reports without the need of installing tools on their devices.

VI. AS PROTOTYPE

Currently, the prototype authoring system contains only the virtual history record repository and a local open data repository, i.e., the first two AS components mentioned above. Author can search data in these repositories. Search results are managed in Data Catalog. These data can be further searched by users via APIs, which provides data in JSON format, and web user interface, which visualizes data into table format. For example, the virtual historical record repository now provide access to data from P-Alert, a disaster record website developed by Institute of Earth Science, Academia Sinica and the CWB website. During a searching process launched by a user, the system searches data with specified filter via APIs provided by P-Alert and via HTML webpage analysis based on web crawler tools to obtain data that match the filter condition. In this way, data are presented to the user together as if the data were from the same database.

The local open data repository is built on CKAN [24], an open source data portal software for data management. Data that have been reviewed for data quality control issue can be uploaded, categorized, deleted, and more, by administrators of each events. When a user launches a search process via APIs or web user interface, the system will search through the local database by calling CKAN API, and then return the search results.

VII. SUMMARIZATION AND FUTURE WORK

In the previous sections, we described the proposed DiSRC system. A DiSRC system contains numerous DiReCTs and an authoring server system. The former enables the capture and recording observational data during disasters, while the latter provides functions to help author analyze and use the capture data to generate their reports. The proof-of-concept system demonstrates that the DiReCT application can executes based on pre-defined SOP. The current SOP includes a flood scenario state-machine workflow, simple data validation functions, and a questionnaire. The prototype of authoring system demonstrates the feasibility of the discovery and gathering data from multiple websites and databases, hiding the physical distribution of the data from the users. Our next step is to develop a more polish and complete DiReCT system and its assistant tool library. We will complete the authoring server system, in particular completing a basic part of the tool library containing methods for validating and merging inconsistent data to ease data managing and analyzing process, and tools for analysis, visualization, and simulation.

REFERENCES

[1] M. Pittore, M. Wieland, and K. Fleming, “Perspectives on global dynamic exposure modelling for geo-risk assessment,” *Nature Hazards*, vol. 86, no. 1, pp.7-30, 2017.

[2] UNDP, “A Comparative Review of Country-Level and Regional Disaster Loss and Damage Databases”, 2013. [Online]. Available: http://www.undp.org/content/undp/en/home/librarypage/crisis-prevention-and-recovery/loss-and-damage-database.html.

[3] CommCare. [Online]. Available: https://www.commcarehq.org/home/.

[4] Magpi. [Online]. Available: http://home.magpi.com/.

[5] EPICollect. [Online]. Available: http://www.epicollect.net/.

[6] Fieldata. [Online]. Available: http://fieldata.org/.

[7] Microsoft Data Gathering. [Online]. Available: https://www.microsoftdatagathering.net/home.

[8] Open Data Kit. [Online]. Available: https://opendatakit.org/.

[9] OpenXdata Kit. [Online]. Available: http://www.openxdata.org/.

[10] PoiMapper. [Online]. Available: http://www.poimapper.com/.

[11] Surveybe. [Online]. Available: http://surveybe.com/.

[12] Formhub. [Online]. Available: https://formhub.org/.

[13] DARMsys. [Online]. Available: http://qldreconstruction.org.au/about/darmsys.

[14] Rover. [Online]. Available: http://www.roverready.org/.

[15] Fulcrum. [Online]. Available: http://www.fulcrumapp.com/.

[16] Orion. [Online]. Available: http://www.orionprotected.com/.

[17] KoBoToolbox. [Online]. Available: http://www.kobotoolbox.org/.

[18] F. Ballio, D. Molinari, G. Minucci, M. Mazuran, C. Arias Munoz, S. Menoni, F. Atun, D. Ardagna, N. Berni, and C. Pandolfo, “The RISPOSTA procedure for the collection, storage and analysis of high quality, consistent and reliable damage data in the aftermath of floods,”. *Journal of Flood Risk Manage*, 2015.

[19] J. Bevington, R. Eguchi, C. Huyck, H. Crowley, F. Dell’Acqua, G. Iannelli, C. Jordan, J. Morley, M. Wieland, S. Parolai, M. Pittore, K. Porter, K. Saito, P. Sarabandi, A. Wright, and M. Wyss, “Exposure Data Development for the Global Earthquake Model: Inventory Data Capture Tools,” *Proceedings of the 15th World Conference on Earthquake Engineering, Lisbon*, no. 5057, 2012.

[20] D. Link, A. Widera, B. Hellingrath, T. De Groeve, G. Eidimtaite, and M. K. Limbu, “GDACSmobile - An IT Tool Supporting Assessments for Humanitarian Logistics,” *The International Emergency Management Society 2015 Annual Conference*, 2015.

[21] National Weather Service, “Observational Quality Control—General, National Weather Service Instruction 10-1305,” July 2014. [Online]. Available: http://www.nws.noaa.gov/directives/sym/pd01013005curr.pdf.

[22] Integrated Ocean Observing System, “Manual for Real-Time Quality Control of Water Level Data,” May 2014. [Online]. Available: http://www.ioos.noaa.gov/qartod/welcome.html.

[23] U.S. Geological Survey, “Data Management: Manage Quality.” [Online]. Available: http://www.usgs.gov/datamanagement/qaqc.php.

[24] CKAN. [Online]. Available: https://ckan.org/