Environmental Modelling & Software 58 (2014) 86–94

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration

Zhanming Wan[a,b], Yang Hong[a,b,*], Sadiq Khan[a,b], Jonathan Gourley[c], Zachary Flamig[b,c], Dalia Kirschbaum[d], Guoqiang Tang[e]

[a] *School of Civil Engineering and Environmental Sciences, University of Oklahoma, OK, USA*
[b] *Hydrometeorology & Remote Sensing Laboratory (HyDROS) and Advanced Radar Research Center (ARRC), National Weather Center, Norman, OK, USA*
[c] *NOAA/National Severe Storms Laboratory, National Weather Center, Norman, OK, USA*
[d] *NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*
[e] *Department of Hydraulic Engineering, Tsinghua University, Beijing, China*

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
7 April 2014
Accepted 8 April 2014
Available online

Keywords:
Flood
Cloud
Crowdsourcing
Web GIS

ABSTRACT

Flood disasters have significant impacts on the development of communities globally. This study describes a public cloud-based flood cyber-infrastructure (CyberFlood) that collects, organizes, visualizes, and manages several global flood databases for authorities and the public in real-time, providing location-based eventful visualization as well as statistical analysis and graphing capabilities. In order to expand and update the existing flood inventory, a crowdsourcing data collection methodology is employed for the public with smartphones or Internet to report new flood events, which is also intended to engage citizen-scientists so that they may become motivated and educated about the latest developments in satellite remote sensing and hydrologic modeling technologies. Our shared vision is to better serve the global water community with comprehensive flood information, aided by the state-of-the-art cloud computing and crowd-sourcing technology. The CyberFlood presents an opportunity to eventually modernize the existing paradigm used to collect, manage, analyze, and visualize water-related disasters.

1. Introduction

Flooding is one of the most dangerous natural disasters globally, frequently causing tremendous loss of life and economic damages. According to the International Federation of Red Cross (IFRC) and Red Crescent Societies (RCS), almost half of the natural disasters that happened between 2002 and 2011 were floods. During this period, natural disasters caused approximately 1.1 million fatalities worldwide, affected approximately 2.7 billion people, and led to economic losses totaling approximately $1.4 trillion USD. Of these damages, approximately 57,000 (5%) of the fatalities, 1.2 billion (44%) of the affected, and $278 billion USD (20%) of the economic damages were attributed to floods alone (Zetter, 2012).

The significant global impact of recurring flooding events leads to an increased demand to have comprehensive flood databases for flood hazard studies. There are several existing flood databases, such as the International Disaster Database (EM-DAT), ReliefWeb (launched by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA)), the International Flood Network (IFNET) and the Global Active Archive of Large Flood Events (created by the Dartmouth Flood Observatory (DFO)). However, there is often a lack of specific geospatial characteristics of the flooding impacts or a failure to enlist all flood events due to variable entry criteria. Moreover, these data warehouses lack interactive information sharing with the communities affected by the flood events. Therefore, a methodology developed by Adhikari et al. (2010) utilized valuable flood event information from the aforementioned sources, specifically the DFO, and synthesized these data with media reports and remote sensing imagery in order to provide a record of flooding events from 1998 to 2008. The digitized Global Flood Inventory (GFI) gathers and organizes detailed information of flood events from reliable data sources, defines and standardizes categorical terms as entry criteria for flood events (e.g. severity and cause), and cross-checks and quality controls flood

* Corresponding author. Advanced Radar Research Center, National Weather Center, HyDROS Lab, HyDrometeorology and RemOte Sensing, 4610 Suites, 120 David L. Boren Blvd., Norman, OK 73072, USA. Tel.: +1 405 325 3644; fax: +1 405 325 4217.

E-mail addresses: zwan@ou.edu (Z. Wan), yanghong@ou.edu (Y. Hong).

URL: http://hydro.ou.edu

http://dx.doi.org/10.1016/j.envsoft.2014.04.007
1364-8152/© 2014 Elsevier Ltd. All rights reserved.

event information (e.g. location) to eliminate redundant listings. These characteristics make GFI an appropriate starting point to develop a unified, global flood cyber-infrastructure. However, one limitation of this database is that GFI only contains flood events through 2008. Although it is possible that flood events after 2008 can be collected manually, as was done in Adhikari et al. (2010), it can be incomplete and inefficient since this process only involves a limited number of resources and people. Recently, technological advances in social media have tremendously improved data gathering and dissemination, especially with the development of world-wide web technologies. Built on the platform of social media, crowdsourcing has become a versatile act of collecting information from the public.

Crowdsourcing is a term that generally refers to methods of data creation, where large groups of potential individuals generate content as a solution of a certain problem for the crowdsourcing initiator (Estellés-Arolas and González-Ladrón-de-Guevara, 2012; Hudson-Smith et al., 2009). In theory, crowdsourcing is based on two assumptions described by Goodchild and Glennon (2010). First, “a group can solve a problem more effectively than an expert, despite the group's lack of relevant expertise”, and second, “information obtained from a crowd of many observers is likely to be closer to the truth than information obtained from one observer.” Based on the definition and assumption of crowdsourcing, it has the ability to collect a considerable amount of information from its randomly distributed participants. The nature of crowdsourcing accommodates data collection in numerous forms, including questionnaires, phone calls, text messages, emails, web surveys and other paper-based, mobile phone-based, and web-based methods. Moreover, crowdsourcing can be embedded with location-based information by using GPS-enabled devices, IP (internet protocol) addresses, or participants' awareness of their current locations. Crowdsourcing offers new opportunities to expand the information available to impacted communities and provide a “two-way” street for the same affected populations to communicate with the global community.

The data collected from crowdsourcing will be used in a cloud computing framework for information sharing that includes data processing and visualization. Gong et al. (2010) adopted cloud computing technology in geoprocessing functions to provide elastic geoprocessing capabilities and data services in a distributed environment. Behzad et al. (2011) used cloud computing in addition to a cyber-infrastructure-based geographic information system to facilitate a large number of concurrent groundwater ensemble runs by improving computational efficiency. Huang et al. (2012) integrated cloud computing in dust storm forecasting to support scalable computing resources management, high resolution forecasting, and massive concurrent computing. As defined by the National Institute of Standards and Technology (NIST), cloud computing is a model for supporting elastic network access to a shared pool of configurable computing resources (Mell and Grance, 2011). The nature of cloud computing assures that it can (a) reduce the time and cost during implementation, operation, and maintenance of the global flood cyber-infrastructure, (b) provide an interface for collaboration at both global and local scales, and (c) conveniently share data in a secure environment. These advantages make cloud computing an attractive technique in the global flood cyber-infrastructure that can maximize the efficiency and data safety during collaboration, while minimizing time and expense spent on the system.

Several studies have already used cloud-based services with water-related management and monitoring. The study of Sun (2013) presented a collaborative decision-making water management system using a cloud service provided by Google Fusion Table. The author describes the migration of the management system from a traditional client-server-based architecture to a cloud-based web system, revealing the potential to fundamentally change a water management system from its design to the operation. Another example is the Namibia flood SensorWeb infrastructure, which was created for rapid acquisition and distribution of data products for decision-support systems to monitor floods (Kussul et al., 2012). The decision-support system utilizes the Matsu Cloud to store and pre-process data through hydrological models, eliminating the latency when clients select specific data.

This study proposes a cloud-computing service provided by Google to establish the global flood cyber-infrastructure, to share the GFI, to provide statistical and graphical visualizations of the data, and to expand the breadth and content of the GFI by collecting new flood data using crowdsourcing technology (i.e. CyberFlood). The next section focuses on the architecture of the cloud computing system designed for global flood monitoring, analysis, and reporting. Section 3 demonstrates the system's functionality, and a summary is provided in section 4.

2. Cyber-infrastructure design for flood monitoring

The global flood cyber-infrastructure consists of four components: the GFI data source, cloud service, web server, and client interface (Fig. 1). The GFI is pre-processed before being imported into the cyber-infrastructure, as explained later in this section. The cloud service, which significantly improves the performance and decreases the burden on the web server, handles all data queries, data visualization, and data analysis. The web server simply deals with sending requests and responses between clients and the cloud. The client interface is mainly built with hypertext markup language (HTML) and JavaScript. Since all the data are processed before being imported into this cyber-infrastructure, the client side only sends operational requests from users and renders responses from the cloud service.

As previously mentioned, GFI standardizes categorical terms as entry criteria for flood events. In other words, every data column in GFI is carefully designed so that each entry strictly follows the criteria of the corresponding data column (Fig. 2a). GFI was pre-processed before being successfully imported into a Google Fusion Table. Python code, which is a cross-platform, extensible, and scalable programming language (Sanner, 1999), was written to

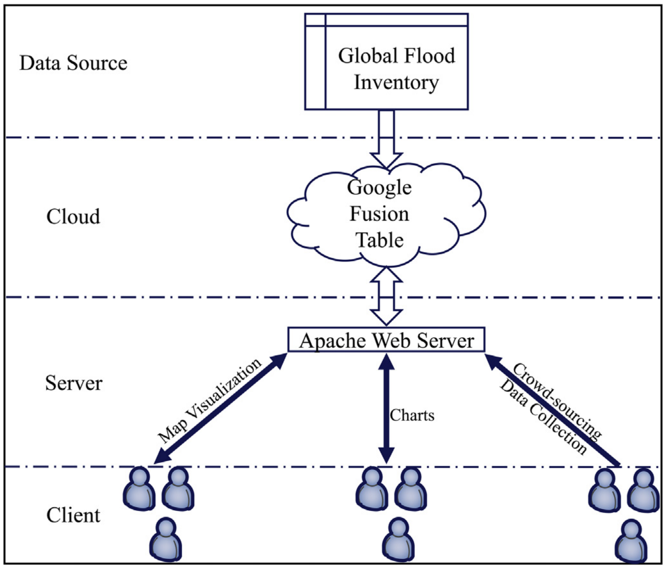


Fig. 1. The global flood community cyber-infrastructure framework.

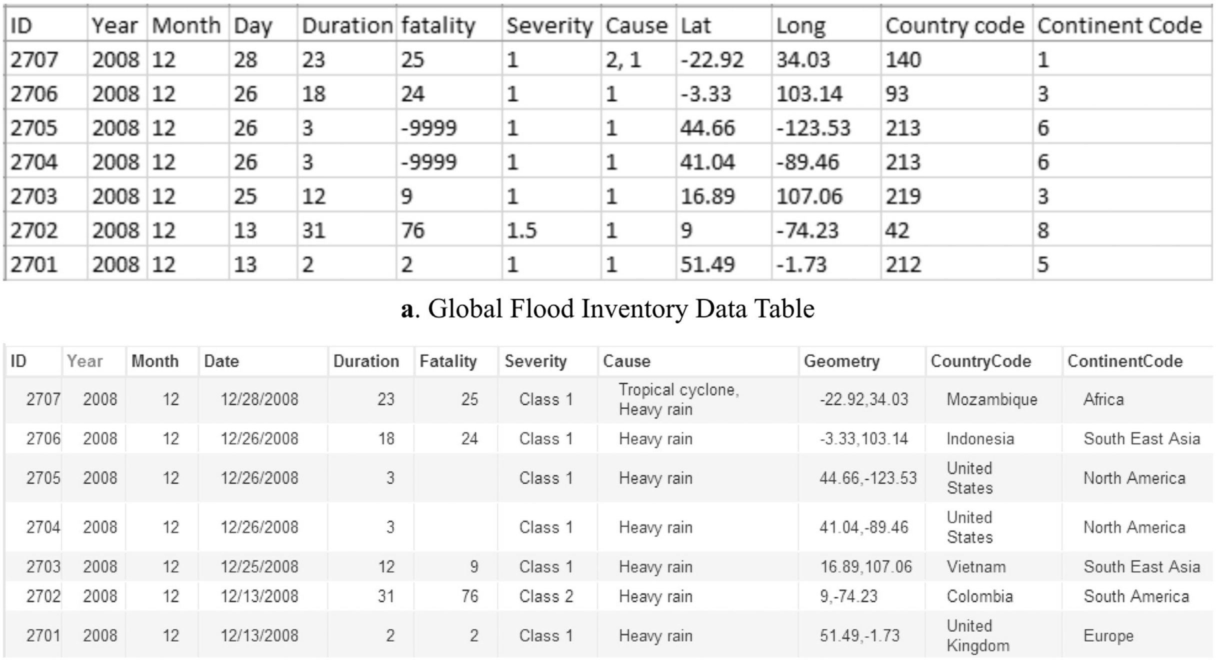
do the data conversion. The purpose is to maintain data consistency, making the converted data readily readable and reducing the data conversion load on the client side. In this process, cells containing −9999, which represent no value in GFI, are removed because they are not consistent with empty cells that also represent no value. Data columns of flood severity, cause, country, and continent are filled with numbers to indicate certain meanings in GFI. A look-up table was used to convert the numerical codes into text. For example, "1" means "heavy rain" in the column pertaining to flood causes, whereas it means "Africa" in the column pertaining to continents (Fig. 2a and b). In other words, if the GFI with numbers are imported into the Fusion Table and used directly by the cyber-infrastructure, the numbers have to be converted to the corresponding texts each time during the refresh on the client side. As a result, text is assigned to severity, cause, country, and continent during this process. Location, the most important information for map visualization in this cyber-infrastructure, is described in two columns representing latitude and longitude in GFI. However, if one flood event involves more than one location, then there will be multiple data records, and only the first data record has shared information such as event ID and date. To improve this data structure and for better visualization, multiple data records representing the same flood event are combined into a single record, while location is presented as MultiGeometry using Keyhole Markup Language (KML) (Wilson, 2008).

An example of a flooding event in New Hampshire in October 2005 is illustrated in Fig. 3. Fig. 3a shows the event as stored in the original GFI covering events from 1998 to 2008. Five locations were associated with this event. Cells are left blank if they share the same record as in the first row. Fig. 3b shows the same flooding event as in Fig. 3a but converted into a Google Fusion Table. This table also includes all five locations that are now represented in the geometry column with KML. Fig. 3c illustrates the visualization of this event, showing the severity as well as the specific locations impacted. Additional layers such as rivers, roads, and topography can also be included during this step to ascertain the spatial extent of inundation.

The processed GFI, now converted to a Google Fusion Table (Fig. 2b) belongs to a "Software as a Service" (Yang et al., 2011) type of cloud-based service for data management and integration (Gonzalez et al., 2010). Google Fusion Table was created to manage and collaborate with tabular datasets in which geospatial fields can be included to provide location information. These geospatial fields can be in the form of latitude and longitude in two separate columns, latitude and longitude pairs in one column, or KML strings in one column. Fusion Table accepts many different tabular formats of files as its data source. Any text-delimited files such as comma-separated values (CSV) files, KML files, and spreadsheets can be imported directly into a Fusion Table. Since Google Fusion Table is a part of Google Drive, users can simply select an existing spreadsheet from their Google Drive and import it into a Fusion Table. Cloud computing is embedded to provide rapid responses to requests from users for data querying, summary, and visualization. Moreover, data security and sharing is already implemented in Google Fusion Table.

The steps required to import data into a Google Fusion Table are straightforward. First, the data must be in one of the supported formats (tabular or text-delimited data such as CSV files, excel spreadsheets, and other similar types.). A wizard then provides easy-to-follow instructions describing how to upload the data. Fusion Table looks like a common table in a spreadsheet, whereas it supports structured query language (SQL) to operate the table. Keywords, such as "SELECT", "INSERT", "DELETE" and "UPDATE", can be used to manipulate Fusion Table, which is similar to how a table is handled in a database. Fusion Table provides application programming interface (API) to programmatically perform SQL-based, table-related tasks through using hypertext transfer protocol (HTTP) requests (Google, 2013). By combining with other Google-provided APIs, the capability of Fusion Table can be extended to not only manipulate the data in the table, but also to visualize the data through thematic mapping and analytic charts.

Fusion Table, which plays an important role in this global flood cyber-infrastructure, provides data storage, data sharing, and fast data access. However, since the infrastructure is functioning from



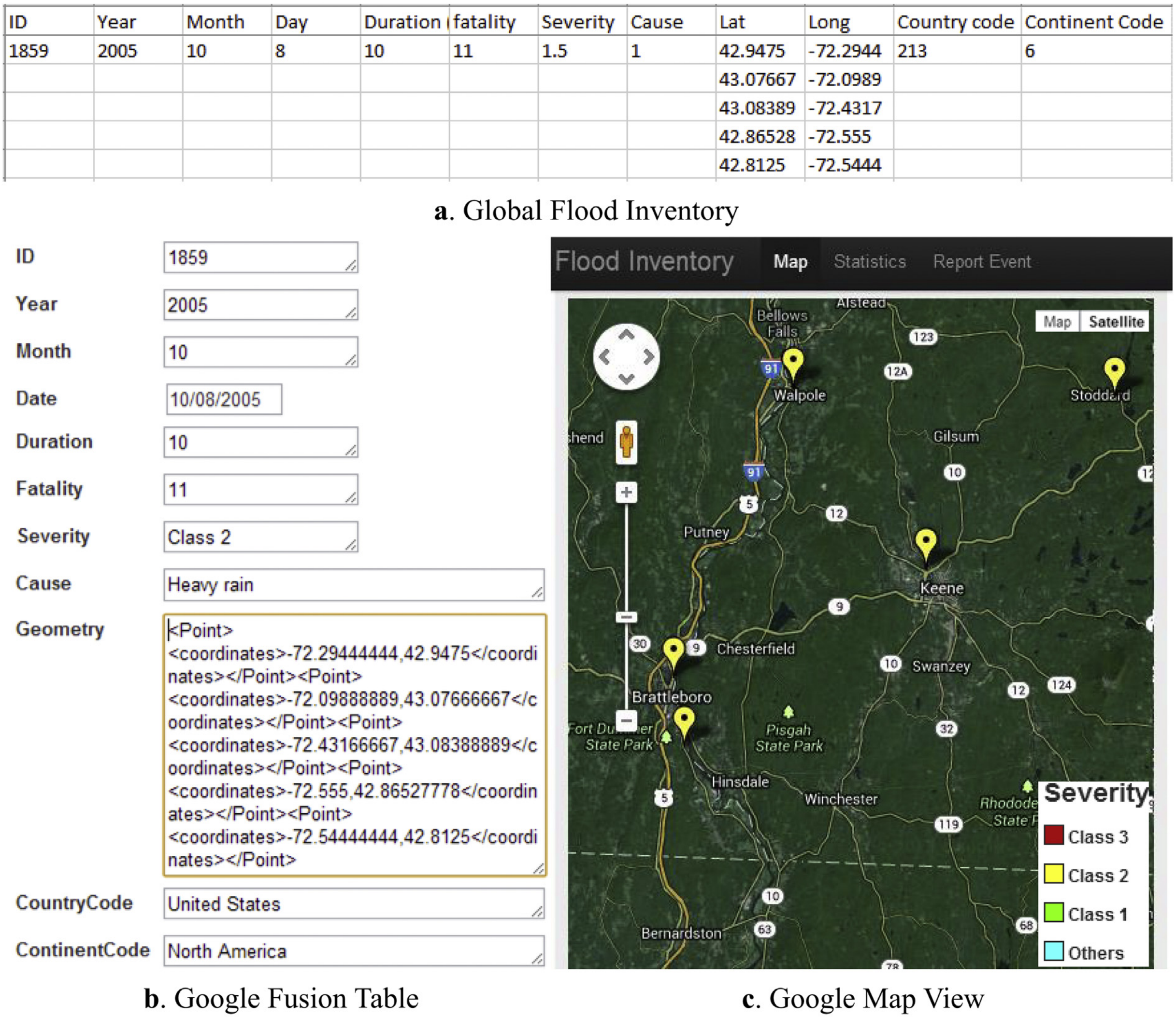
ID	Year	Month	Day	Duration	fatality	Severity	Cause	Lat	Long	Country code	Continent Code
2707	2008	12	28	23	25	1	2, 1	-22.92	34.03	140	1
2706	2008	12	26	18	24	1	1	-3.33	103.14	93	3
2705	2008	12	26	3	-9999	1	1	44.66	-123.53	213	6
2704	2008	12	26	3	-9999	1	1	41.04	-89.46	213	6
2703	2008	12	25	12	9	1	1	16.89	107.06	219	3
2702	2008	12	13	31	76	1.5	1	9	-74.23	42	8
2701	2008	12	13	2	2	1	1	51.49	-1.73	212	5

a. Global Flood Inventory Data Table

ID	Year	Month	Date	Duration	Fatality	Severity	Cause	Geometry	CountryCode	ContinentCode
2707	2008	12	12/28/2008	23	25	Class 1	Tropical cyclone, Heavy rain	-22.92,34.03	Mozambique	Africa
2706	2008	12	12/26/2008	18	24	Class 1	Heavy rain	-3.33,103.14	Indonesia	South East Asia
2705	2008	12	12/26/2008	3		Class 1	Heavy rain	44.66,-123.53	United States	North America
2704	2008	12	12/26/2008	3		Class 1	Heavy rain	41.04,-89.46	United States	North America
2703	2008	12	12/25/2008	12	9	Class 1	Heavy rain	16.89,107.06	Vietnam	South East Asia
2702	2008	12	12/13/2008	31	76	Class 2	Heavy rain	9,-74.23	Colombia	South America
2701	2008	12	12/13/2008	2	2	Class 1	Heavy rain	51.49,-1.73	United Kingdom	Europe

b. Google Fusion Table

Fig. 2. Comparison of data tables a) global flood inventory and b) Google fusion table.



ID	Year	Month	Day	Duration	fatality	Severity	Cause	Lat	Long	Country code	Continent Code
1859	2005	10	8	10	11	1.5	1	42.9475	-72.2944	213	6
								43.07667	-72.0989		
								43.08389	-72.4317		
								42.86528	-72.555		
								42.8125	-72.5444		

a. Global Flood Inventory

ID	1859
Year	2005
Month	10
Date	10/08/2005
Duration	10
Fatality	11
Severity	Class 2
Cause	Heavy rain
Geometry	<Point><coordinates>-72.29444444,42.9475</coordinates></Point><Point><coordinates>-72.09888889,43.07666667</coordinates></Point><Point><coordinates>-72.43166667,43.08388889</coordinates></Point><Point><coordinates>-72.555,42.86527778</coordinates></Point><Point><coordinates>-72.54444444,42.8125</coordinates></Point>
CountryCode	United States
ContinentCode	North America

b. Google Fusion Table

c. Google Map View

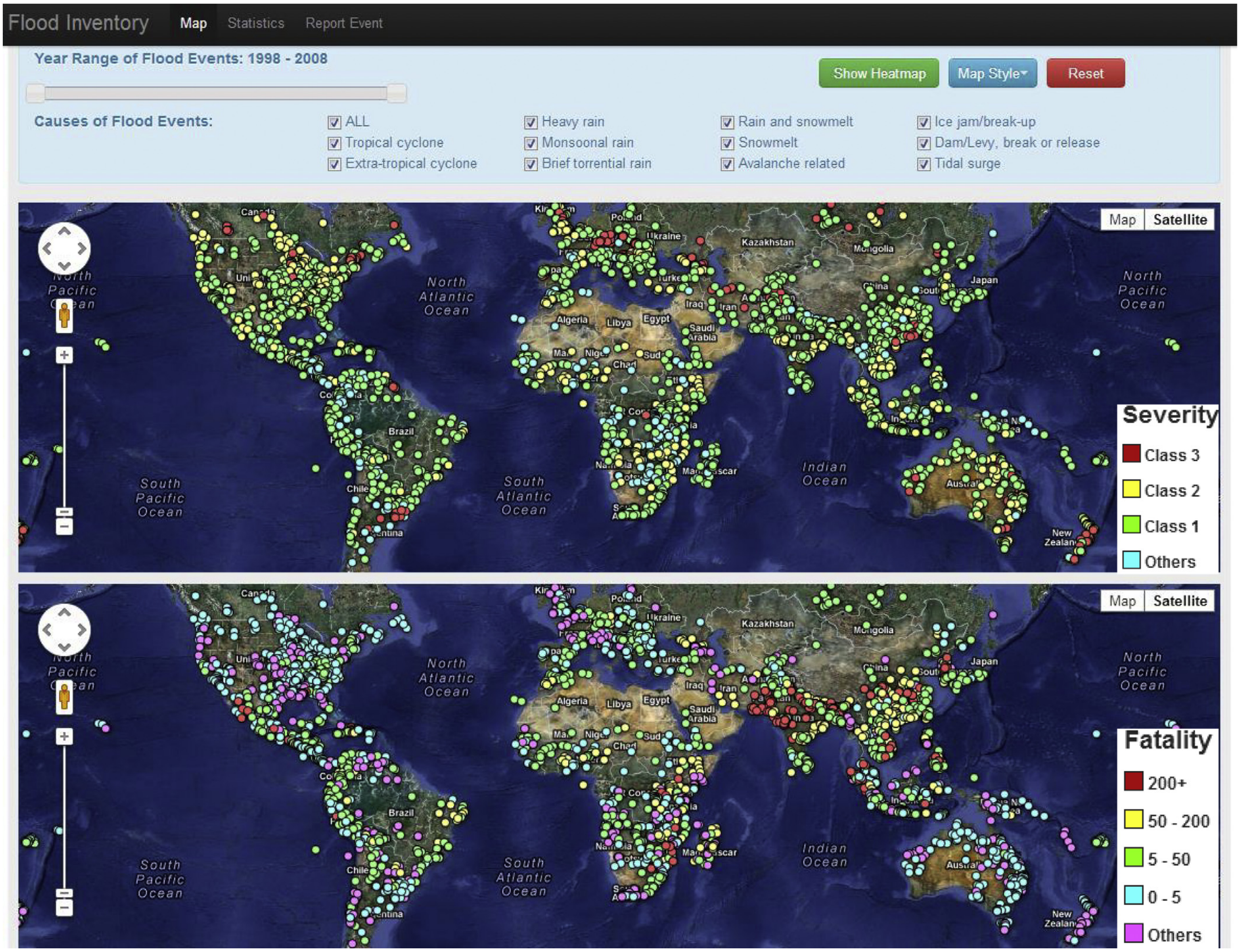
Fig. 3. Flood event over Northeast U.S. in New Hampshire of October 2005 a) global flood inventory, b) Google fusion table attributes, and c) Google map view.

the backend, users cannot benefit from this service unless a traditional server and client components are included for interaction. Since all the computing loads are on the cloud, the web server only serves as a "middleware" dealing with requests and responses between the cloud and clients. The web server also protects the Fusion Table on the cloud from being accidentally modified by clients. Google provides two kinds of API keys for programmers to develop applications. One of the keys is a string, which grants permission to applications to select items from the Fusion Table. The other key is a special file which should be stored securely with the application on the web server. This type of key grants permission to the application from the specific web server to insert, update, or delete items from the Fusion Table. The client side is programmed with HTML and JavaScript, along with several APIs from Google, to send requests through the server to the cloud, receiving responses for location-based and analytic visualization.

3. Demonstration

The global flood cyber-infrastructure is currently running at http://eos.ou.edu/flood/ (Fig. 4). An Apache web server is deployed to host the frontend web interface. Google Map has been integrated to map the locations of flood events after querying the Fusion Table using the Google Map API. All the points representing locations of flood events are color coded by severity or fatalities associated to the flood event. Severity is classified into classes 1, 2, and 3, with "Class 1" being least severe and "Class 3" the most severe. Fatalities are categorized into four classes based on the value. Users are allowed to select a range of years and causes of flood events from the provided controls. Each selection will lead to a new query from the Fusion Table, which means that the desired data will be plotted on the map and can include event details that have just been uploaded in real-time. In addition to visualization of the data using information stored in the Fusion Table, a Google Chart API is utilized to create analytic charts for statistical analysis of the flood events (Fig. 5). Variables such as the year, month, severity, cause, continent, and country, can be analyzed in a chart and a table. Variables can be summarized by the count of the variables, sum of fatalities, or average of fatalities. For instance, Fig. 5 demonstrates the summary of flood events by year and severity. Flood events with Class 1 severity are in a blue color (in web version) on the chart, with about 270 of the flood events in 2003 occurring with such a severity class.

In order to expand and update the existing GFI, now stored as a Fusion Table, crowdsourcing from public entries is implemented in this cyber-infrastructure by providing a flood events observation report form (Fig. 6). Most of the fields are the same as the existing GFI. However, photo URL and source URL fields are appended to the



Year Range of Flood Events: 1998 - 2008

Causes of Flood Events:

☑ ALL
☑ Tropical cyclone
☑ Extra-tropical cyclone
☑ Heavy rain
☑ Monsoonal rain
☑ Brief torrential rain
☑ Rain and snowmelt
☑ Snowmelt
☑ Avalanche related
☑ Ice jam/break-up
☑ Dam/Levy, break or release
☑ Tidal surge

Severity
🟥 Class 3
🟨 Class 2
🟩 Class 1
🟦 Others

Fatality
🟥 200+
🟨 50 - 200
🟩 5 - 50
🟦 0 - 5
🟪 Others

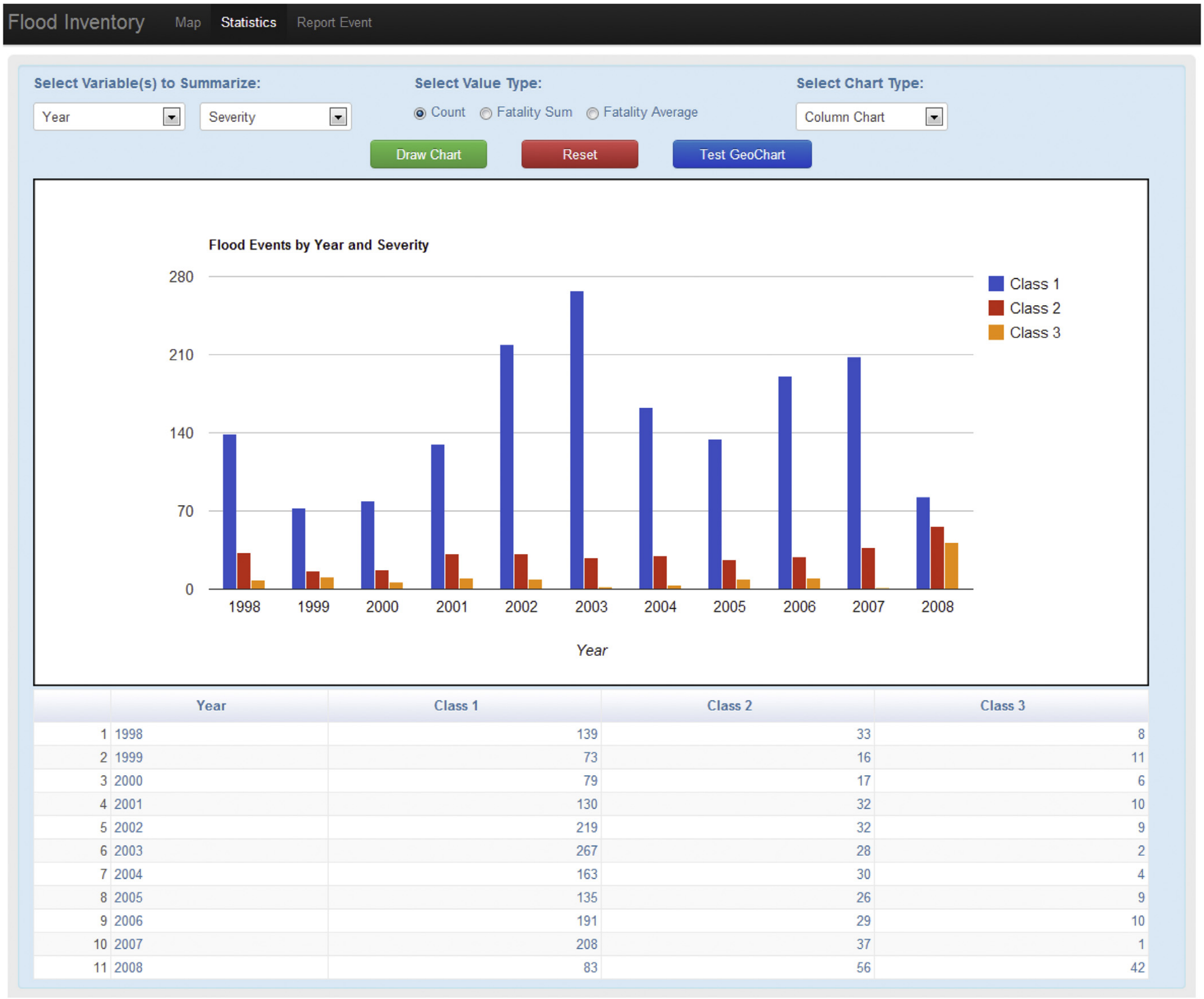
Fig. 4. The map visualization of global flood cyber-infrastructure. The top and bottom maps are color coded by severity and fatalities respectively.

Fusion Table to store additional details about the submitted flood event. This means that users are able to upload one photo per submission and provide a URL of the web source as a proof or supplemental information of that flood event. The current date will be retrieved from the users' operating system by default to submit present flood events. Users can also select any date between 1998 to present if past events are reported. Since reported events will be displayed on the map in real-time immediately following submission, location is a required field in the report form. Location will be automatically retrieved if a location service is allowed by the client's browser or the uploaded photo is geo-tagged. This report form is submitted directly into the Fusion Table through the server, and this process is protected by Google Account Authentication and Authorization Mechanism to secure data on the Fusion Table. A two-way quality control approach of data from crowdsourcing is implemented. First, when a user submits a report of flood events, the system will automatically check if each field is correctly formatted. For example, fields of latitude and longitude can only be numeric values. Fields of day, month, and year are restricted to certain numbers which can only be selected by users. Instructions have also been created for first-time users and they can learn what each field means and how to retrieve current location to help them submit correct information. Secondly, after submission, the data will be manually checked with different sources, including news reports, flood reports from other major disaster data sources, and satellite imagery. Checking data sequentially is not an efficient way of quality control. However, it is effective in this case since the number of data received so far is limited. Newly submitted events following post-processing will be assigned IDs according to the number of milliseconds from 1970/01/01 to the time of the submission. For example, a flood event reported at 12/18/2013 23:35:15.199 will be assigned an ID of 1387431315199. Sequential IDs will be assigned to newly submitted data after quality control is complete. If crowdsourced data submissions increase in frequency in the future, automated data quality control procedures will be developed to check the spatial and temporal consistency with other flood reports. Other automated procedures can cross-check the reports with global flood forecasts available from http://eos.ou.edu/Global_Flood.html. A crowdsourcing way to control the quality of crowdsourced flood events reports are under consideration. A mechanism could be established to grant permission to qualified users and students who have expertise in flood monitoring and validation to check the data quality in the Fusion Table.

4. Discussion

4.1. Advantage

Although CyberFlood does not directly solve flooding problems, this work is expected to be able to help advance flood-related



Select Variable(s) to Summarize: Year, Severity

Select Value Type: ● Count ○ Fatality Sum ○ Fatality Average

Select Chart Type: Column Chart

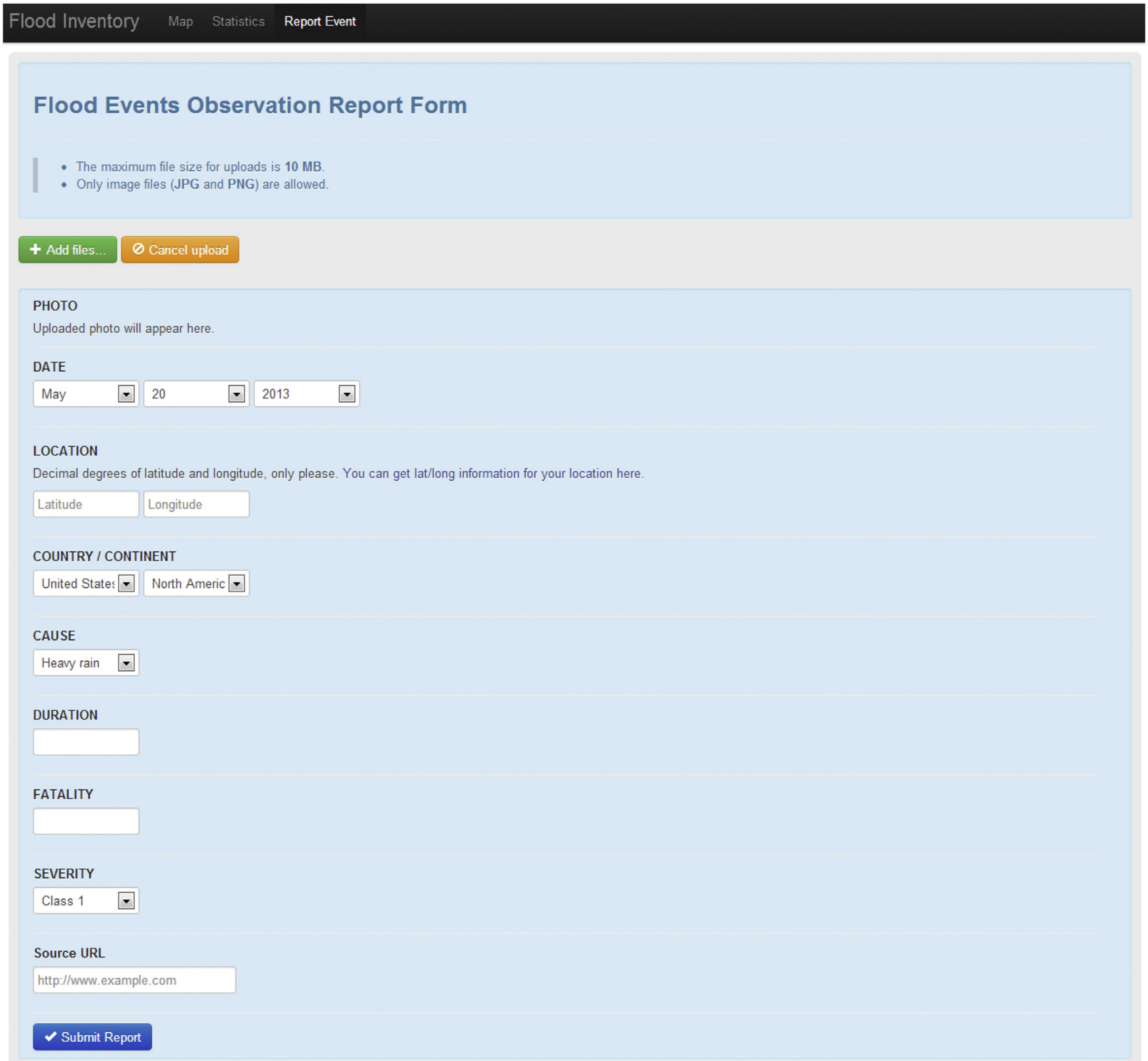
	Year	Class 1	Class 2	Class 3
1	1998	139	33	8
2	1999	73	16	11
3	2000	79	17	6
4	2001	130	32	10
5	2002	219	32	9
6	2003	267	28	2
7	2004	163	30	4
8	2005	135	26	9
9	2006	191	29	10
10	2007	208	37	1
11	2008	83	56	42

Fig. 5. The statistical chart and table of global flood cyber-infrastructure.

research areas such as hydrologic model evaluation, flood risk management, and flood awareness. Both the public and research community can use the resources provided by this cyber-infrastructure to analyze retrospective flood events and submit their witness accounts of previously unreported flood events. Therefore, this approach is useful for flood monitoring and validation research. The long-term database could also help generate flood climatology of occurrences and damage and therefore could potentially lead to better flood risk management for zoning and other flood-related decision-making purposes. Public engagement using crowdsourcing and cloud-based techniques could potentially raise flood awareness around the globe and provoke citizen-scientists to consider careers in the natural sciences, engineering, and mathematics.

CyberFlood has been created to be used by anyone with internet access. In order to access the flood resources, a web-based interface is provided and is becoming accessible through iOS apps for mobile users. As CyberFlood becomes more accessible through these apps, more people will use it to view retrospective flood events, monitor current flood events, and contribute to the flood community by submitting their reports of flood events. CyberFlood has been created to adapt the idea of Volunteered Geographic Information (VGI), which is described as tools to create, assemble, and disseminate geographic information provided voluntarily by individuals (Goodchild, 2007), for compiling flood events by involving map-based visualization and utilizing human sensors to collect useful data globally.

Compared with the traditional server-client structure, the cloud computing service provided by Google Fusion Table enhances the performance of the global flood cyber-infrastructure in terms of the speed during data query and data visualization. By providing a Fusion Table API, the complexity of the global flood cyber-infrastructure is significantly reduced. This benefits both programmers and clients since they are able to focus more on the actual functions they need to implement and use, not on the logistics with the cloud itself. Rather than using the traditional server-client based structure, this simplified cloud-based framework makes it easier to develop scalable applications. Furthermore, taking into consideration of data sharing and collaboration, Fusion Table provides a comprehensive solution to keep data secure while making seamless communications between collaborators and Google servers for data updates, queries, and visualization.



Flood Events Observation Report Form

- The maximum file size for uploads is **10 MB**.
- Only image files (**JPG** and **PNG**) are allowed.

PHOTO
Uploaded photo will appear here.

DATE
May ▾ 20 ▾ 2013 ▾

LOCATION
Decimal degrees of latitude and longitude, only please. You can get lat/long information for your location here.
Latitude Longitude

COUNTRY / CONTINENT
United States ▾ North Americ ▾

CAUSE
Heavy rain ▾

DURATION

FATALITY

SEVERITY
Class 1 ▾

Source URL
http://www.example.com

Fig. 6. The flood events observation report form.

4.2. *Performance experiment*

An experiment was developed to compare the speed of reading data and geographically displaying data using Google Maps API with a Google Fusion Table and a MySQL database respectively, both of which contain the same dataset. Google Maps API provides two ways to display markers on Google Map. The traditional way is by using google.maps.Marker class. The more efficient way is to utilize google.maps.FusionTablesLayer class which can only be employed by data from the Google Fusion Table. As a result, the data in the Google Fusion Table is visualized by google.maps.FusionTablesLayer class while the data in the MySQL database is visualized by google.maps.Marker class in this experiment. The query speed of both Google Fusion Table and MySQL database are rapid, taking a few milliseconds. However, the speed advantage becomes predominant when using data from the Google Fusion Table with google.maps.FusionTablesLayer class. Table 1 demonstrates the results of this performance experiment. The first 1000, 5000, 10,000, 50,000, and 100,000 records are retrieved from the dataset. The average time of reading and displaying different size of data is calculated from 5 consecutive measurements. When data records increase from 1000 to 100,000, the average elapsed time for using the Google Fusion Table with google.maps.FusionTablesLayer class is always low (less than 10 ms) while the average elapsed time for using the MySQL database with google.maps.Marker class is much higher (more than 1000 ms) and increases significantly to more than 3000 ms when displaying 100,000 records.

4.3. *Limitation and scalability*

Fusion Table has some limitations on storage and usage. Each user can import data files no more than 100 MB into each Fusion

Table 1
Performance comparison results.

	Google Fusion Table (google.maps.FusionTablesLayer)					
Test order	1	2	3	4	5	Average
1000 records	17	8	8	7	8	9.6
5000 records	9	7	6	7	9	7.6
10,000 records	12	7	8	6	7	8.0
50,000 records	8	9	8	7	7	7.8
100,000 records	14	10	9	8	8	9.8
	MySQL (google.maps.Marker)					
Test order	1	2	3	4	5	Average
1000 records	1052	1039	1041	1049	1048	1045.8
5000 records	1128	1138	1143	1111	1116	1127.2
10,000 records	1202	1194	1230	1233	1240	1219.8
50,000 records	1842	2145	1915	1867	2211	1996.0
100,000 records	3050	3332	2938	2895	3123	3067.6
				Unit: Milliseconds (ms)		

Table, and each Google cloud account can contain data no more than 250 MB. The Google Fusion Table is an experimental product, which does not have a payment option for increasing the storage space. However, the data inside the Google Fusion Table is text-based which takes up very little space. When data are inserted into the Google Fusion Table, efforts have been made with additional code/scripts to save space by normalizing each field and trimming unnecessary spaces. Currently, there are 2730 records in the Fusion Table, which takes up 657 KB out of 250 MB. This means approximately 1 million similar data records can be stored with just this one Google cloud account. Furthermore, photo submissions are uploaded to a separate server with terabyte-level shared storage space and only the URLs linked to the photos are stored in Fusion Table.

The situation when the dataset grows beyond the limit of approximately 1 million records has also been taken into consideration. One solution is to have the data stored in multiple Fusion Tables of multiple Google accounts and perform a cross-table query. Another way is to use other cloud-based services, such as Google Cloud SQL and BigQuery, Amazon EC2, and Windows Azure. Google services will be our first choice because it is usually straightforward to develop applications with other Google products, such as Google Maps/Earth and Google Chart.

When inserting a data record into the Fusion Table, the record should be less than 1 MB, and a maximum of 25,000 requests per day can be sent to one Google account with free Fusion Table API access. However, the number of maximum requests per day can be increased by request through Google.

As a result, there is a trade-off between using Fusion Table resources directly and consuming a small portion of the resources from clients. In order to reduce the times in querying the Fusion Table, data from the prior queries are stored on the client side in the global flood cyber-infrastructure. If the next operation from the client side returns the same result as the previous operation, no request will be sent to the Fusion Table. It will use the stored data instead.

4.4. Data sharing

Although Google Fusion Table API does not provide a way to download raw data programmatically, as a shared cyber-infrastructure, the data in the Fusion Table of CyberFlood is free to download. A link can be provided to the actual Fusion Table from where users can view raw data and download them as a CSV or KML file. After the raw data have been made accessible, it is possible for users to adapt the raw data to visualize flood events in their own way and gain more discovery.

4.5. Sustainability

In order to involve people, some poster presentations about CyberFlood have been given at several conferences. Meanwhile, iOS apps for iPad and iPhone are under development, providing functions for people to view map and chart visualization of flood events and submit their witness accounts of flood events. Plans are made to advertise the CyberFlood through non-traditional media, such as social media Facebook and Twitter. We have also developed the mPING (Meteorological Phenomena Identification Near the Ground: http://www.nssl.noaa.gov/projects/ping/) app which includes flood entries (4 levels of severity) and uses crowdsourcing technique to obtain data. Given that the mPING has more than 200,000 active users today, this app will also be utilized to advertise our CyberFlood system.

Since only limited entries from crowdsourcing during the 2009–2013 period are obtained, locally recruited students are compiling flood events from multiple sources for that period with manual quality control. Data for these years will be available in CyberFlood.

5. Conclusion

The global flood disaster community cyber-infrastructure (CyberFlood), with cloud computing service integration and crowdsourcing data collection, provides on-demand, location-based visualization, as well as statistical analysis and graphing functions. It involves citizen-scientist participation, allowing the public to submit their personal accounts of flood events to help the flood disaster community to archive comprehensive information of flood events, analyze past flood events, and get prepared for future flood events. This cyber-infrastructure presents an opportunity to eventually modernize the existing methods the flood disaster community utilizes to collect, manage, visualize, and analyze data with flood events.

In the future, data describing the flood reports in this cyber-infrastructure will be linked to real-time and archived satellite-based flood inundation areas, observed stream flow, simulated surface runoff from a global distributed hydrologic modeling system, and precipitation products. These datasets will be beneficial both as method to validate the crowdsourced flood events and to help educate, motivate, and engage citizen-scientists about the latest advances in satellite remote-sensing and hydrologic modeling technologies. Given the elasticity of a cloud-based infrastructure, this cyber-infrastructure for global floods can be applied to other natural hazards, such as droughts and landslides, at both global and regional scales.

Acknowledgment

The first author would like to acknowledge the seed money support from Hydrometeorology & Remote Sensing Lab (HyDROS) and Advanced Radar Research Center (ARRC) at the University of Oklahoma.

References

Adhikari, P., Hong, Y., Douglas, K., Kirschbaum, D., Gourley, J., Adler, R., Robert Brakenridge, G., 2010. A digitized global flood inventory (1998–2008): compilation and preliminary results. Nat. Hazards 55 (2), 405–422.

Behzad, B., Padmanabhan, A., Liu, Y., Liu, Y., Wang, S., 2011. Integrating CyberGIS gateway with Windows Azure: a case study on MODFLOW groundwater simulation. In: Proceedings of the ACM SIGSPATIAL Second International Workshop on High Performance and Distributed Geographic Information Systems. ACM, pp. 26–29.

Estellés-Arolas, E., González-Ladrón-de-Guevara, F., 2012. Towards an integrated crowdsourcing definition. J. Inf. Sci. 38 (2), 189–200.

Gong, J., Yue, P., Zhou, H., 2010. Geoprocessing in the Microsoft Cloud Computing Platform-Azure. In: Proceedings the Joint Symposium of ISPRS Technical Commission IV & AutoCarto, p. 6.

Gonzalez, H., Halevy, A., Jensen, C.S., Langen, A., Madhavan, J., Shapley, R., Shen, W., 2010. Google fusion tables: data management, integration and collaboration in the cloud. In: Proceedings of the 1st ACM Symposium on Cloud Computing. ACM, pp. 175–180.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69 (4), 211–221.

Goodchild, M.F., Glennon, J.A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. Int. J. Digital Earth 3 (3), 231–241.

Google, 2013. Google Fusion Table API. https://developers.google.com/fusiontables/ (accessed 28.06.13.).

Huang, Q., Yang, C., Benedict, K., Chen, S., Rezgui, A., Xie, J., 2012. Utilize cloud computing to support dust storm forecasting, 6 (4), 338–355.

Hudson-Smith, A., Batty, M., Crooks, A., Milton, R., 2009. Mapping for the masses accessing Web 2.0 through crowdsourcing. Soc. Sci. Comput. Rev. 27 (4), 524–538.

Kussul, N., Mandl, D., Moe, K., Mund, J.-P., Post, J., Shelestov, A., Skakun, S., Szarzynski, J., Van Langenhove, G., Handy, M., 2012. Interoperable infrastructure for flood monitoring: sensorweb, grid and cloud. Selected topics in applied earth observations and remote sensing. IEEE J. 5 (6), 1740–1745.

Mell, P., Grance, T., 2011. The NIST definition of cloud computing (draft). NIST Spec. Publ. 800, 145.

Sanner, M.F., 1999. Python: a programming language for software integration and development. J. Mol. Graph Model 17 (1), 57–61.

Sun, A., 2013. Enabling collaborative decision-making in watershed management using cloud-computing services. Environ. Model. Softw. 41, 93–97.

Wilson, T., 2008. OGC KML, Version 2.2. 0. Open Geospatial Consortium.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M., Fay, D., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? Int. J. Digital Earth 4 (4), 305–329.

Zetter, R., 2012. World disaster report: focus on forced migration and displacement. Int. Fed. Red Cross Red Crescent Soc., 260–267.