RI15002, 9 November 2015 Computer Science

IBM Research Report

Blue Water: A Common Platform to Put Water Quality Data in India to Productive Use by Integrating Historical and Real-time Sensing Data

Sandeep Sandha, Sukanya Randhawa, Biplav Srivastava
IBM Research Division
IBM Research
New Delhi - 110070. India.

Blue Water: A Common Platform to Put Water Quality Data in India to Productive Use by Integrating Historical and Real-time Sensing Data

Sandeep S Sandha, Sukanya Randhawa and Biplav Srivastava
IBM Research, India
{ssandha1,Sukanya_R,sbiplav}@in.ibm.com

ABSTRACT

Water is unique in its role as a life preserver. It is important to all members of a society. However, if one is looking for quality data in India to make data-driven decisions, one is lost. This is surprising given that there is a rich history of data collection in the country and looks forward to adopting real-time water sensing in a big way. In this paper, we first explore issues around data management practices for water that have prevented their widespread dissemination and then, in response, propose a general data access and reuse framework which focuses on usage of water information for a purpose like irrigation, drinking or industrial need.We illustrate our approach through a prototype accessible as a mobile app and from a web site. This framework titled as 'Blue Water' would serve as a common data integration platform for historical as well as real-time sensing data, and for the latter, we also present details for development of a real-time Non-Stationery Water Quality Sensing Tool.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Decision support

General Terms

Water, Data Management, Sensing

Keywords

Water, Smart Cities, IoT, Data Management

1. INTRODUCTION

Water affects all stakeholders in a society due to its unique role as a life preserver. Apart from drinking, some of the other common activities that require water are domestic, agricultural, industrial and construction activites with different levels of acceptable water standards.

Pollution of water in India's rivers is a cause of major concern. A case in point is the largest river in India, Ganga, which is about 2525 km long. Its basin has an area of approximately 861,404 square km making it roughly one-fourth the size of the country. The water of Ganga and its numerous tributaries is used by over 400 million people, which is roughly one-third of India's population and about that of the United States. Unfortunately, the state of Ganga is quite poor [2] despite years of efforts spent in monitoring water quality monitoring and its reporting.

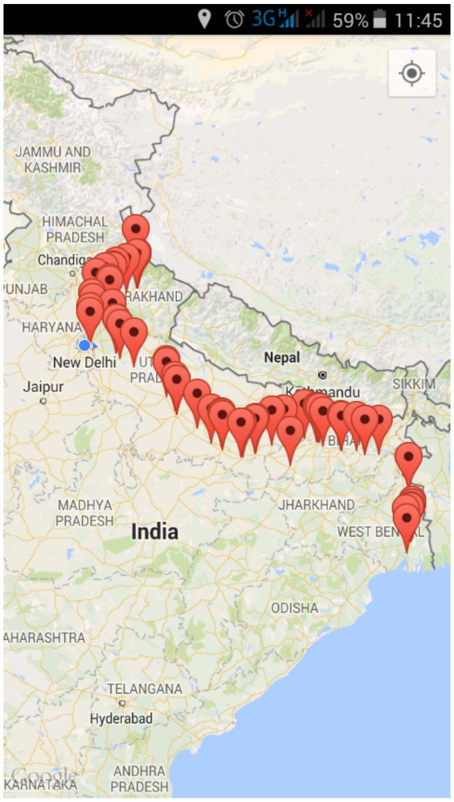
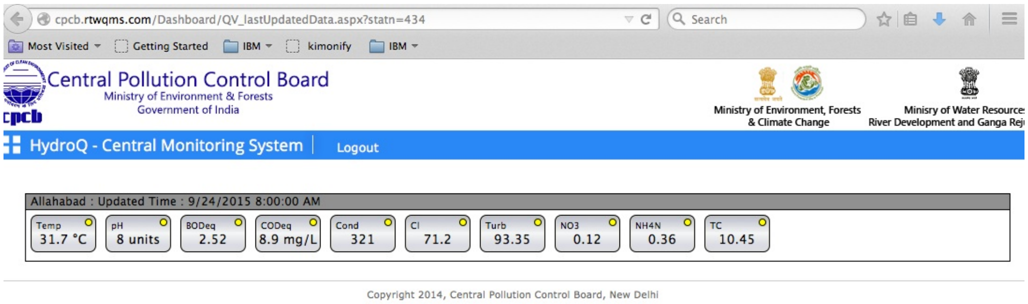


Figure 1: A snapshot of MyGanga app with spatial distribution of available data.

Furthermore, general public finds it very difficult to make decisions using water data. Abhay may want to take a bath in the river during a religious festival and would want to know which banks of the river are feasible to go without getting sick. Bina may want to fetch water for household activities. Chetan may want to use river water for irrigating his fields. Divya may wonder if fishing or vegetable growing



Allahabad : Updated Time : 9/24/2015 8:00:00 AM

Temp	pH	BODeq	CODeq	Cond	Cl	Turb	NO3	NH4N	TC
31.7 °C	8 units	2.52	8.9 mg/L	321	71.2	93.35	0.12	0.36	10.45

Figure 2: Data and visualization from CPCB site for Allahabad.

is promising on the river catchment area to supplement her family's earnings. Eashwar may want to help Ganga dolphins increase in strength so that more tourists can come to his river-side hotel. Farida, a doctor, may want to warn patients about water-borne diseases so that diseases in the area come down.

These and other are not able to discover any water relevant data to factor into the decisions they will anyway make. What they do not know is that water data in India has been collected by numerous agencies (usually government, academic and non-profits) and individuals for many years on different stretches of the river measuring different parameters at different times using different technologies. However, they as a whole, are not only insufficient in terms of quality and quantity but also inaccessible to wider public because of its uses focus on narrow technical water terms. Figure 2 shows the user-interface to access water data for rivers.

In this context, we want to promote wider usage of available data by increasing its access and ease of interpretation. Further, we want to enable a common platform for analyzing both historic and new, recent data, which are beginning to be collected with real-time sensing. We present an initial prototype of the platform called *Blue Water* in this direction. Figure 1 shows a screen shot of the *MyGanga* app showing all the available data along the Ganga basin culled from public and open sites[15]. A video is available too[14].

Apart from collating new data from various organizations (based on lab water testing), we go a step further to provide capabilities of collecting and integrating data from sensor platforms that are or would be deployed in the future for water quality monitoring in real-time. We provide details of how we developed and implemented a type of real-time sensing platform i.e the non-Stationery Water Quality Sensing Tool, designed for rivers and lakes monitoring and incorporated its data too in *Blue Water*.

Our contributions are that we:

1. review the state of availability of water data in India
2. outline use-cases where water data could be used and how to convert available data to usable form
3. present a prototype system which uses open water data and makes it readily available in the context of usage and relevant standards
4. describe data management practices for existing and new data that help improve water quality efforts in India

The rest of the paper is organized as follows: we begin with a survey of water data availability in public space and their issues. We then present our approach of collating data from multiple sources on a common platform and making it available via mobile and web. Next, we show how new sensing data can be organized together with historical data. We then present a walkthrough of *MyGanga* protoype, discuss advantages of our approach in context of prior work and concluding with pointers to future work.

2. DATA CHARACTERISTICS

We begin by discussing data availability in India. Table 1 lists a snapshot of water data sources from different organizations. It is representative but not exhaustive.

The first source is India's open data portal[12]. It collects data from different government departments and releases them under the NDSAP policy. It ensures that data does not have syntactical issues and provides access is a range of formats (e.g., CSV, JSON, APIs). Although this is a good start, the data may still have semantic quality issues.

The second source is a prominent portal maintained by a Non-Government Organization (NGO). It is rich in content but the content formats (e.g., PDFs) make reuse hard for follow-on or correlative analysis with other datasets.

The third and fourth sources are from a government pollution monitoring agency. They release data but again their supported formats and access mechanisms do not facilitate content reuse.

The fifth source has data from monitoring carried out by several academic institutions and research organizations namely: PCRI, BHEL - Haridwar, CPCB - Lucknow and Delhi, IIT Kanpur, IIT Roorkee, Patna University, Bidhan Chandra Krishi Vishwavidyalaya - Kalyani, Thapar Institute of Engineering and Technology - Patiala, Bangalore University, ITRC - Lucknow, SJIC of Engineering - Mysore, Bapuji Institute of Engineering and Technology - Davangere, Shimoga University and Anna University - Chennai. The data is available in text format which posses data integration issues with other sources.

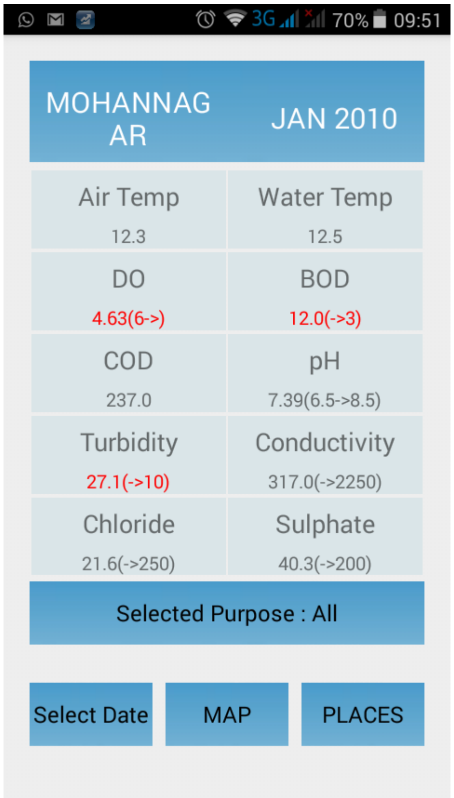
2.1 Purpose-driven Water Data Usage

As noted in the introduction, water information can be useful for invidual stakeholders for different purposes. In Figure 3, we illustrate a few examples further highlighting the datasets needed and off-the-shelf analytical techniques which can be used.

In [4], CPCB prescribes the parameters that 22 different industries should focus on for real-time sensing (in Table

No.	Organization	Description	Comments
1.	data.gov.in	India's data portal	Data released under NDSAP policy; sourced from multiple government agencies. E.g., https://data.gov.in/catalog/water-quality-data-river-ganga
2.	India Water Portal	Non Government Organization	Data and articles maintained by an NGO, mostly in PDFs At: http://indiawaterportal.org/
3.	Central Pollution Control Board	Government	Data from India's main pollution monitoring agency At: http://data.gov.in/sites/default/files/water_quality_2012_river_ganga.csv
4.	Central Pollution Control Board	Government	Real-time data from India main pollution monitoring agency. At: http://cpcb.rtwqms.com/, login needed
5.	Ministry of Environment and Forest	Government	Data from academic institutions and research organizations At: http://wqmrivers.nic.in/, login needed

Table 1: A snapshot of water data availability in India.



MOHANNAG AR JAN 2010

Air Temp 12.3
Water Temp 12.5
DO 4.63(6->)
BOD 12.0(->3)
COD 237.0
pH 7.39(6.5->8.5)
Turbidity 27.1(->10)
Conductivity 317.0(->2250)
Chloride 21.6(->250)
Sulphate 40.3(->200)

Selected Purpose : All

Figure 4: A snapshot of pollutant information in MyGanga app for a location and time shown in the context of permissible limits and purpose.

3). The safe limit for these parameters vary by purpose and are given by different organizations. Specifically, drinking standard is by BIS[1] and many others are by CPCB[5]. An effort is necessary to harmonize these different acceptable water standards themselves.

2.2 Data Challenges

As we see, there is a rich history of collecting river pollution data. However, they face some challenges:

1. The collection and measurement process is subject to dispute. Many readings are once per month or quarter and the process was only recently standardized.
2. There is a wide diversity in the number of parameters measured. While CPCB recommends more than 30 parameters [3, 4], the actual parameters vary from 2-25 with most datasets having less than 10 parameters.
3. Data is not easily accessible. This includes data being shown as a plot but raw numbers not being available to download, data behind a portal that cannot be compared with othe data sets.
4. The standards to judge parameters for different purposes need to be harmonized for consistency and ease in implementation.

3. SOLUTION APPROACH

We now discuss the technical details of BlueWater system.

3.1 Architecture

The system architecture of *BlueWater* is client-server based as shown in Figure 5. Here, data is stored in a database on a server while it is accessed and manipulated using mobile and web clients.

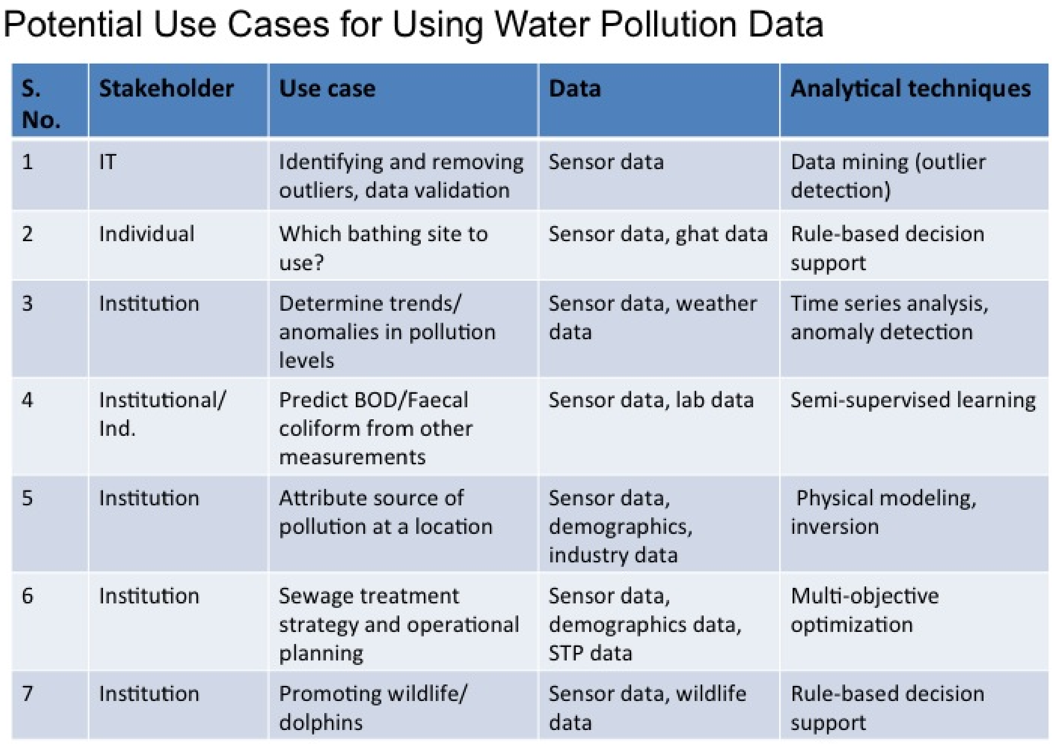
A web based portal is used for data uploading. We have provided functionalities to upload data from different sources having different parameters. The server-side stores, processes, and manages the collected data.

Client can be any android mobile device like a smartphone or a tab in which the collected water data is visualized. The system is developed using Java on the server side, and Android based apk for the client. We have used Cloudant as the database. The system is deployed using the IBM Bluemix.

3.2 Data Management

The data from different sources has different structure. The data available from the sources discussed in Table 1 is either in csv, excel or in simple text form. We extract the meta data and water quality parameters available in the data sources. The basic information required for each data row is the place info (latitude, longitude) and any one of the parameter value.

The data has lot of diversity and is semi-structured in nature. Some data sources have very few parameters whereas some data sources provides a large set of parameters. The parameters provided by the dataset depends largely on the use case addressed by the monitoring source. To handle the diversity and to make it more flexible to incorporate new parameters for the furture we have used the Cloudant database which stores data in semi-structured JSON format.



Potential Use Cases for Using Water Pollution Data

S. No.	Stakeholder	Use case	Data	Analytical techniques
1	IT	Identifying and removing outliers, data validation	Sensor data	Data mining (outlier detection)
2	Individual	Which bathing site to use?	Sensor data, ghat data	Rule-based decision support
3	Institution	Determine trends/ anomalies in pollution levels	Sensor data, weather data	Time series analysis, anomaly detection
4	Institutional/ Ind.	Predict BOD/Faecal coliform from other measurements	Sensor data, lab data	Semi-supervised learning
5	Institution	Attribute source of pollution at a location	Sensor data, demographics, industry data	Physical modeling, inversion
6	Institution	Sewage treatment strategy and operational planning	Sensor data, demographics data, STP data	Multi-objective optimization
7	Institution	Promoting wildlife/ dolphins	Sensor data, wildlife data	Rule-based decision support

Figure 3: An illustration of usage of water pollution data.

3.3 Data Upload And Visualization

The information from different sources is represented in common format as shown in Table 2. The common format can be directly uploaded in the form of file using a web portal. The metadata is used to identify the spatial temporal features used to present data in the visual interactive way to the user. The safe limits on different kinds of parameters are also maintained online on the server which allows us to do analysis of the different water quality parmeters and to address different use cases. The *MyGanga* application allows user to visualize the data in the form of map as well as highlighting the parameters which are not within safe limits as shown in Figure 1 and Figure 4.

The webportal and the functionalities supporting different types of data uploads is shown in Figure 6. Places data including place name, description, latittude and longitute, range can be uploaded using *Places File Upload*. To upload the water quality parameters we have provided several different upload features. *Data File Upload* is used to upload the water quality data available in JSON format. *CSV File Upload* is used to upload the water quality data available in CSV format and to upload the excel data file we have provided *Excel File Upload*. The safe and unsafe limits on different water qulaity parameters can be uploaded using *General Limit File Upload*. The currently available places data can be visualized using the *Places Data*, by clicking the submit.

A snapshot of server hosted on IBM Bluemix is shown in Figure 7. The IBM Bluemix console allows us to decide the allocated memory and number of instances of server to run.

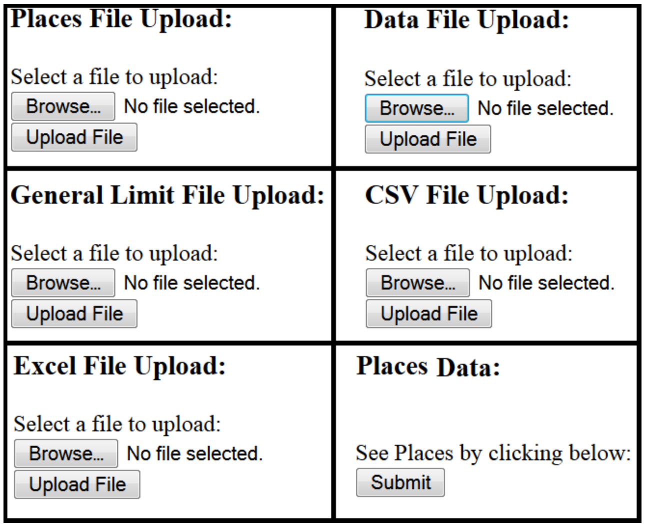


Figure 6: Web Portal for Data upload.

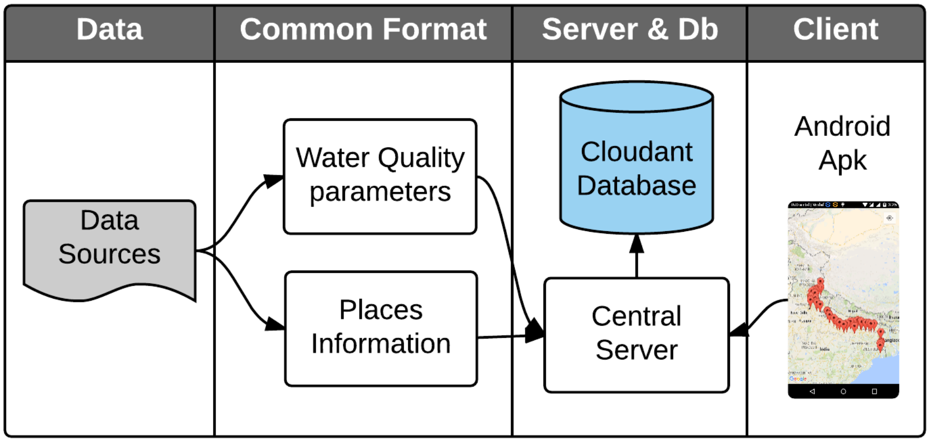


Figure 5: System architecture for data upload and visualization.

No.	Information	Description
1.	Place meta data	The place name, latitude and longitude time of collection, source of data.
2.	Water quality parameters	All parameters like pH, BOD, COD, heavy metals in [4, 5]

Table 2: Information present in common format.

The data stored in Cloudant database is also shown in figure 7. We have three databases. The database *limits_general* is used to store safe and unsafe limits. It has one document stored in JSON format. This document is accessed by android client at runtime to check the limits. Storing limits in cloud database allows us to modify the limits independent of the client application. The water quality parameters data is stored in *places_data*. As shown it has 118 documents, each document stores the water quality data for several parameters in JSON format along with its spatial and temporal metadata. The metadata is used to identify the location or place from *places_ganga* database to which this data belong and the time of collection of data. The place details are stored in *places_ganga* database which are used to identify the data belonging to a particular place from *places_data*. In the *places_data* currently we have 64 different places, each stored in separate document.

The currently supported queries on Cloudant are:

- Listing water quality parameters by location.
- Listing water quality parameters by purpose. We currently consider three different purposes which are drinking, irrigation and bathing. Depending on the purpose, the water quality parameters which are important and their safe limits vary.
- Listing water quality parameters of a location during a certain time period or duration .
- Listing safe/unsafe water quality parameters of a particular location.

The design is general purpose and more queries can be seamlessly supported.

3.4 Discussion

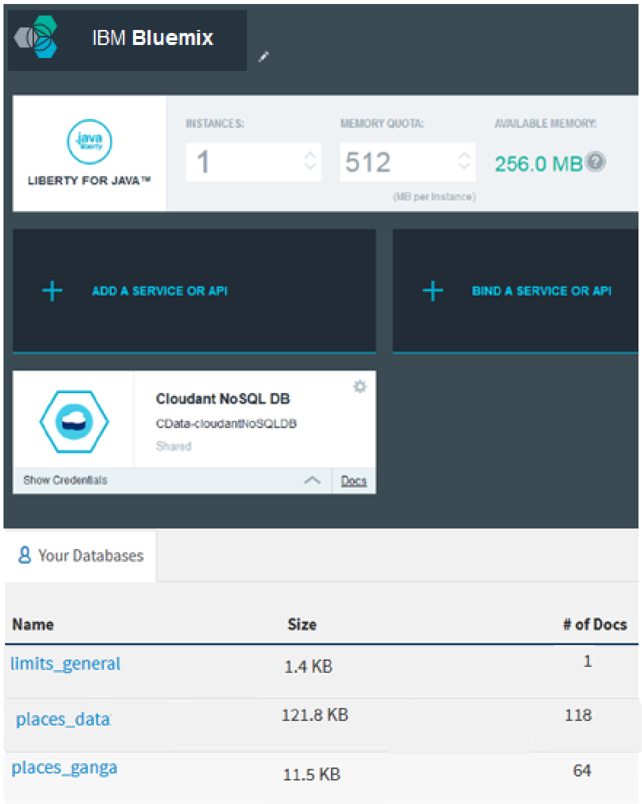
We highlight some of the salient features of *BlueWater*. The integration of data at a common platform was a challenge. The system was designed in order to allow the integration of different data sources having diverse parameters. We used the Cloudant database due to the lack of structure in data. In order to allow the large data sets uploading, we provided the capabilites to upload files after converting them to the common format by segregating the meta data and water quality parameters. The server is deployed on IBM Bluemix which allows us to dynamically control the dedicated processing power and resources to the application. The Android based visualization provides access, anytime and everywhere to the entire available data.

4. WORKING WITH RECENT SENSING AND PAST HISTORICAL DATA

Much of historical data collection data in India was based on taking water samples and then testing in the labs. There is a new push to measure water using real-time sensors. In this section, we describe how both recent and historical data can be used together on the same platform.

4.1 Integration of Real-time Sensing Data

Recently, the concept of real-time sensing of water resources has gained popularity on a global level. From a data perspective, this practice allows continuous monitoring of water quality parameters(physical, chemical and biological) within a reasonable time, thereby opening new opportunities to devise strategies for processing and validation of raw data, in order to convert them into useful informa-



INSTANCES: 1
MEMORY QUOTA: 512 (MB per Instance)
AVAILABLE MEMORY: 256.0 MB

Cloudant NoSQL DB
CData-cloudantNoSQLDB
Shared

Name	Size	# of Docs
limits_general	1.4 KB	1
places_data	121.8 KB	118
places_ganga	11.5 KB	64

Figure 7: Server on Bluemix and Cloudant databases.

tion. Real-time sensing platforms are promising for preservation of water resources as access to high-frequency spatial and temporal information facialitates in real-time event detection. This offer many advantages compared to sporadic testing in the context of various use-cases such as:

- Allows development of early warning detection systems to detect pollution events for a water body and prevent the spread of water-borne diseases due to human or animal consumption of infected water for different purposes.
- Allows monitoring industrial and urban waste water discharges, increased contamination levels due either treated or untreated water.
- Allows data-driven inspection to detect pollution violations and drive compliance[7].

However, for successful implementation of the points mentioned above, water quality data needs to be monitored continously over a reasonable interval of time in order to aid in enhanced decision-making and formulating new control strategies or procedures for water usage applications. As a first step to integrate the data from real-time sensing platforms, one needs to look at the ways in which the sensor data is transmitted to a server.

We will discuss this specifically in context to our efforts to develop a non-stationery sensing platform which has been built to collect high granularity data across the entire stretch (or a desired section) of a water resource like a river or lake. Figure 8 shows a snapshot of the sensing setup for surface water quality testing done at Ulsoor Lake, Bangalore. The setup consists of a floating buoy (white) which is tied to the boat end. To this buoy the sensor logging probe is fixed at the centre with the help of tightened screws for increased stability as the boat moves. The logging probe holds multi-sensors in its body that measure different water quality parameters such as pH, Temperature, Pressure, ORP (Oxidation-Reduction Potential), D.O (Dissolved Oxygen), Nitrate, EC (Electrical conductivity) and Turbidity. The logging probe connected to a multimeter via a 20m cable was purchased from Hanna Instruments (Model: HI 9829). The apparatus allows for real-time monitoring of all the water quality parameters (mentioned above) with a desired sampling interval time ranging from one second to hours. To avoid errors in these values due to bubbles or sticking of large particulate matter to the sensors, the logging probe is provided with a rugged outer cover with perforated holes that allow water to come into contact with the sensors while keeping any weeds or big particles away for sensor area. Moreover, for enhanced accuracy and stability of readings, they can be averaged upto 20 data points if required. However, in our case since the boat was moving with a fixed velocity(/ 5Km/h), averaging was not found to be necessary. The HI9829 system offers many logging options where one could choose between logging via meter or via probe independently. We chose the meter logging option as it also provides GPS locations for the measured data. The logged data was then transferred to a device/PC via company proprietary software which allows data to be imported into most spreadsheet programs. The device/PC is then interfaced

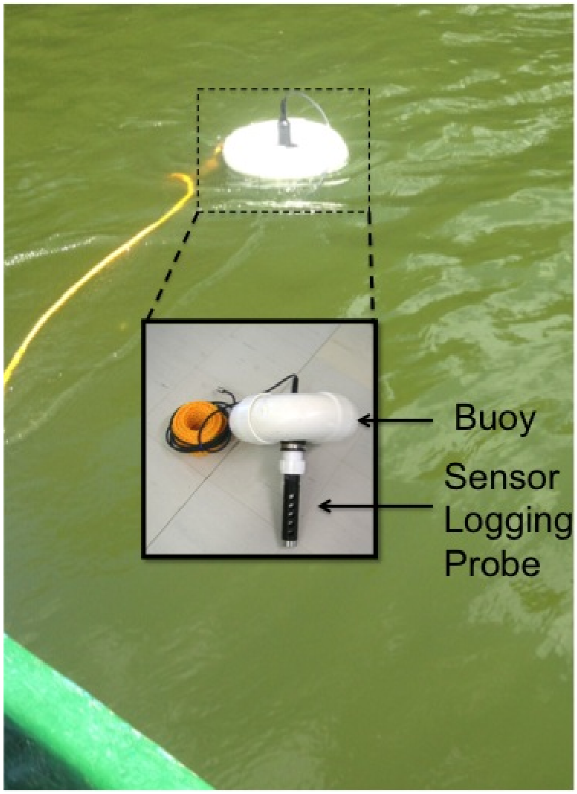


Figure 8: A snapshot of a Non-Stationery Sensing Platform with a Buoy attached.

with the cloud platform (Bluemix by IBM).

Typically, this architecture could also be extended to any device (or board) for data acquisition and transfer to cloud as long it is capable of interfacing with the meter/probe on one hand and the cloud platform on the other (via a WiFi dongle or Ethernet port or GSM/GPRS shield). As also shown in Figure 9, for communicating data onto cloud platform, a protocol for gateway interface(for e.g. MQTT) is required, that allows for publishing and subscribing to the server or other devices. The cloud platform is typically intergrated with a web server framework for handling requests and a robust client-side library, databases like Cloudant & different visualization applications supported by web/phone. By this way, various plug-in engines could be integrated for enhanced functionalities that include various data analytics engines as well. This way a specific application can be tailored with respect to the use-case targeted and allows for a controlled data- flow from real-time sensors to application server side.

Cloud based platforms faciliatates the implementation of a moving platform like in the case of our non-stationery sensing setup attached to a boat that allows for instant monitoring of water quality data even for remote locations. However, further development of real-time moving sensing idea requires stable mechanical platforms for improved data stability and accuracy from water sensors. From our initial experiments we learned that bouy based solutions could get very unstable in water when the boat velocity exceeds 5 km/h. Also, monitoring water quality at surface presents new challenges given the high amount of aeration that gets triggered due to bubbling caused by the movement of the boat. This leads to an increased amount of oxygen which is picked up specially by the D.O sensor and is not representative of the actual water data. To overcome these issues, we would propose new and more stable platform as part of our future work to ensure accurate water quality measurements by non-stationery sensing platforms.

4.2 Loading Recent Data and Incorporating in Platform

The real-time sensing data collected by Hanna Instruments (Model: HI 9829) is stored in excel file format. In the web based portal we have provided the capability to directly upload this excel file. The server extracts the meta data and the water quality parameters from the file. However with the real-time sensing the latitude and longitude of the data reading changes with almost each data reading due the movement of the data collection platform. In order to visualize and analyze the data of particular place we need to identify the data points belonging to a particular place. We addressed this issue by associating a range (which can vary from few metres to several kilo-metres) with each place in the database. All the points lying within the range defined for the place are found using the *Haversine* formula given in Equation 1 are now considered to belong to that place. The definition of *Haversine* is given by Equation 2. *Haversine* is actually used to calculate the distance between two points on a sphere (earth surface) from their latitudes and longitudes. More formally, for two points on a sphere (of radius ξ) with latitudes ϕ_1 and ϕ_2, latitude separation $\Delta\phi = \phi_1 - \phi_2$, and longitude separation $\Delta\lambda$ and the distance d between the two points. In our case distance d is the range defined for a particular place.

$$haversin(\frac{d}{\xi}) = haversin(\Delta\phi) + cos(\phi_1)cos(\phi_2)haversin(\Delta\lambda) \tag{1}$$

$$haversin(\theta) = sin^2(\frac{\theta}{2}) \tag{2}$$

This way in order to do detailed spacial anaysis of the data we can define many different places having small ranges and when we need to analyze the entire city it can be done by defining large range of several kilo-metres. The real-time data had few parameters which were being measured and Since we were using semi-structured database, they were incorporated easily in the platform.

5. DISCUSSION AND FUTURE WORK

The need for better river water quality data has been felt for a while and especially in cases of severe pollution like Ganga in Kanpur[9] and Hindon[11]. The environment courts in India have expressed lack of satisfaction with existing data and taken drastic actions like closure of hundreds of tanneries[10]. The government and World Bank are planning to put 100+ real-time measuring stations for rivers in India but only a handful are operational despite years of efforts.

In [3] and [4], India's pollution regulator has prescribed guidelines for undertaking lab-based water and real-time water sensing, respectively. The guidelines cover measurement procedures for over 30 pollutants but it is not clear how

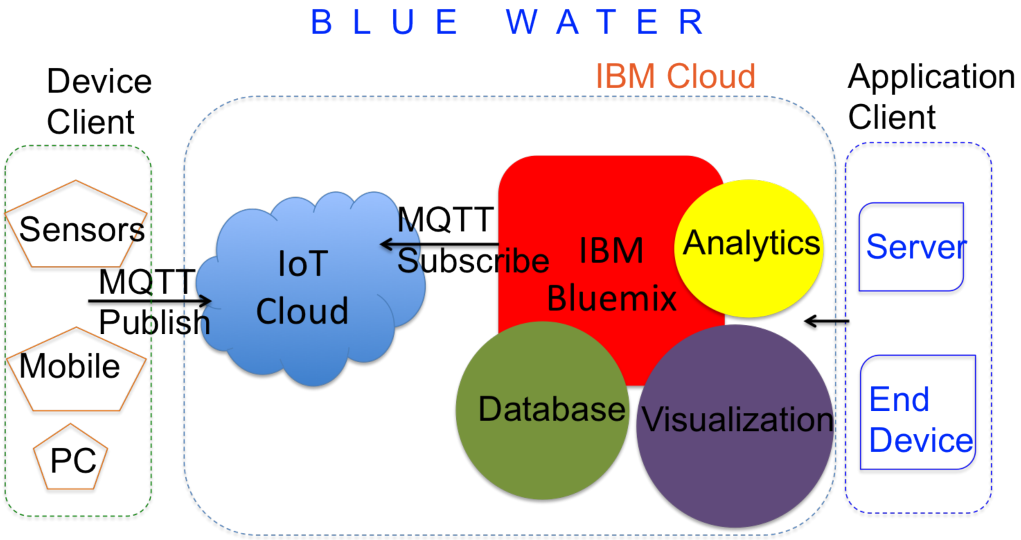


Figure 9: An overview of cloud architecture for real time sensors.

many water monitoring organizations are actually following them as reflected in the data in public domain for rivers (recall Table 1) and status reports for rivers like Ganga[2].

In this paper, we argued that existing data was not being put to maximum usage due to poor data management practices. We want to promote wider usage of available data by increasing its access and ease of interpretation based on usage needs of different stakeholders. Further, we want to enable a common platform,for analyzing both historic and new, recent data, which are beginning to be collected with real-time sensing.

A walkthrough of *BlueWater* platform and its working via the *MyGanga* app is available in a video[14]. The user can select a location from a map or a list and latest information available is shown for that place with default purpose. There are 60+ locations listed at present. The user can select a different time period or purpose and see the available water pollution parameter values and their corresponding permissible limits.

In future, one can improve data storage practices even further. Specifically, current data is stored in a common database but it can be stored over multiple instances managed by independent organization connected together using linked-data technology for seamless analysis and limitless scalability. Second, one can build more expressive geo-spatial queries to give a rich interpretation of river condition. Third, one can try more alternative real-time river measurement technologies and incorporate their data.

With the given increase in numerous technologies that utilise the concept of mobile based sensing of water quality based on smart phone platforms with attached sensor units[13, 8, 6], we would also like to extend the framework of our platform to integrate such emerging sensing methods. This type of feature integration provides us with the unique advantage of having data that is ubiquitous given the widespread ease of data capture and transmission via mobile and wireless technologies[6]. Furthermore, this allows us to connect many inaccessible and remote areas together with the mainstream locations to a centralized server and in return facilitate instant data processing in terms of early warning detection and reporting for these remote locations.

As a long term solution, bringing together data from such varied sources to a common platform facilitate virtual connection of the physically interconnected different water resources over a very large area, such integrated knowledge playing a pivotal role in understanding the broader effects and their causes on the overall water ecosystem.

6. CONCLUSION

In this paper, we presented a general, common platform to integrate historical and real-time sensing data for water called *BlueWater*. The system can uses existing data from any site making it available for reuse (*open data*) as well as new sensing data. A web based interface faciliates seamless loading of data as well as access via programming interfaces (APIs). The system organizes the data by location, time and parameters and shows them in the context of permissible limits for different usage. The existing data and insight can be access via a mobile app called *MyGanga* and web. The system exemplifies the potential of putting water data in India to use by diverse stakeholders that hither-to had no other alternative.

7. ACKNOWLEDGEMENTS

The authors will like to thank Karthik Visweswariah and Supratik Guha, for overall guidance; Theodore G van Kessel, Hendrik Hamann and Bharat Kumar for technical inputs; and V. Rajagopalan for explaining the Indian water ecosystem.

8. REFERENCES

[1] BIS. Indian standard drinking water specification (2nd rev). In *Available at: http://cgwb.gov.in/Documents/WQ-standards.pdf*, 2012.

[2] C. P. C. Board. Pollution assessment: River ganga. In *Available at: http://cpcb.nic.in/upload/NewItems/NewItem_203_Ganga_report.pdf*, 2013.

[3] C. P. C. Board. Guide manual: water and wastewater analysis. In *Available at: http://cpcb.nic.in/upload/NewItems/NewItem_17_guidemanualw&wwanalysis.pdf*, 2014.

[4] C. P. C. Board. Guidelines for online continuous monitoring system for effluents. In *Available at: http://www.inpaper.com/FinalGuidelinse.pdf*, 2014.

[5] CPCB. National water quality monitoring programme. In *Available at: www.cpcb.nic.in/divisionsofheadoffice/pams/NWMP.pdf*, 2014.

[6] A. Faustine and A. Mvuma. Ubiquitous mobile sensing for water quality monitoring and reporting within lake victoria basin. In *Wireless Sensor Network*, 2014.

[7] B. Ford, B. Srivastava, and M. Tambe. Can game theory help clean the ganga? randomized inspection is a case in point. *https://data.gov.in/community/can-game-theory-help-clean-the-ganga-randomized-inspection-is-a-case-in-point/*, 2015.

[8] G.L.Liu. Mobosens. In *At http://nanobionics.mntl.illinois.edu/mobosens/*, 2013.

[9] S. Gupta, R. Gupta, and R. Tamra. Challenges faced by leather industry in kanpur. Report, 2007.

[10] D. Jainani. Kanpur leather industry in danger as ngt cracks whip on pollution, 2015.

[11] H. Lewis. Hindon river: gasping for breath. *Meerut: Janhit Foundation Publisher*, 2007.

[12] NIC. Indian open data. In *At http://data.gov.in. Accessed 6 May 2015*, 2015.

[13] TernUp. Caddisfly. In *http://caddisfly.ternup.com/*, 2015.

[14] VideoDemo. Myganga app. In *Available at: https://www.youtube.com/watch?v=eB8zs7bKyhk*, 2015.

[15] WWWF-ODI. Open data barometer report, second edition 2015. In *At http://barometer.opendataresearch.org//. Accessed 6 May 2015*, 2015.