IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Responsive Information generation system for Kanhan River, an effective information system for river modeling

Dinesh A. Lingote[1*], Girish S. Katkar[2], Ritesh Vijay [3], R. B. Biniwale[4]

[1]CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur, India
[2]Department of Computer Science and Application, Art, Commerce & Science College, Koradi, Nagpur, India
[3]CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur, India
[4]CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur, India

[*]*Corresponding Author: da_lingote@neeri.res.in, Tel.: +91-7773952770*

Available online at: www.ijcseonline.org

Accepted: 09/Dec/2018, Published: 31/Dec/2018

***Abstract*—** River is main source of water for drinking and domestic usage. Over exploitation and discharge of surface water in river stream has ecologically stressed the rivers. In view to manage river health, it is essential to carry out river engineering periodically. River follows complex structure; goes through dense forest and valleys. Therefore, measuring water quality and several other parameters associated with the river is a discouraging job. Mostly river follows longest path and generating data for such large geographical area is very challenging. Scientific study of the river requires data for several consecutive years. Having such large data requirement and expecting data generation simply through field-work is highly burdened and never ending process. Therefore, in this paper we introduced auto data generation techniques like: data extraction, data generation through public-partnership, data estimation and data generation using GIS (Geographic information system) based utility software. Lastly, we illustrate complete data generated by using these auto data generation techniques.

Keywords— Information system, River modelling, Kanhan River; Geo-mapping, River engineering, Water quality

I. INTRODUCTION

Rivers are very important for living organism; they cover covers large geographical area. Sudden increase in industrialization, excess exploitation of natural resources, domestic & industrial discharge, and enormous human activities has significantly deteriorated the river water [1]. India is rapidly developing country; this has ecologically stressed the Indian rivers. Dam constructed on the rivers provides several benefits like: Social, economic, flood control and electricity generation, but on the flip side it has significantly altered the river system [2]. Specifically, if Kanhan River is considered then this river also involves in various agriculture activities, human interference and urban activities. Although Kanhan River is majorly used water resource for Nagpur district, surprisingly this river is not notified as river by state Government. Urbanization has badly affected the Kanhan River. Such environmentally stressed river requires immediate engineering for its rejuvenation, along with it is also important to disseminate river water quality data for the health management. R&D team requires ready data to carry out river engineering. Data is a key entity in conducting any R&D activity; its non-availability discourages the researcher. Although data is available, then its incomplete form is observed a significant hurdle. If data is generated, then its all-time availability for the researcher also needs to be ensured. If research study for large geographical area required to be carried out, then data generation simply by means of forming a team of the researchers and collect data won't be enough sufficient. Therefore, it is very essential to explore different possible methods, which will be effective in generating data for large geographical area.

Considering such factors, developed information system is comprised of five methods for data generation that are: Data generation through field work, extract data from web (HTML), word, excel, pdf and text documents and free GIS based software, generate data through public-partnership and lastly, estimate some missing data using simple statistical data estimation techniques [3]. This information system depicts data on geographical map, thus facilitates look and feel of the data. Being a web-based application, it facilitates great shareware, ensures all time availability of the data for the research team and also facilitates public portal, which benefits common people in observing water quality before its usage.

© 2018, IJCSE All Rights Reserved

II. RELATED WORK

This is first work for Kanhan River of its kind. However, CSIR-NEERI research team has already carried out Kanhan river water quality study and publish research paper "Qualitative evaluation of Kanhan River and its tributaries flowing over central Indian plateau" Environ Monit Assess. 2008 Dec;147(1-3):83-92. Epub 2007 Dec 22. This work motivated research team to develop such information system.

III. METHODOLOGY

This information system is developed for the Kanhan River. The Satpura hills is the widest part of peninsular India, Kanhan River originates from these hills situated in Chhindwara districts of Madhya Pradesh, river progresses to Nagpur district of Maharashtra and confluences the Wainganga River. Some tributaries like: Pench, Kolar, Nag and Sur rivers including some minor water resources joins the river [4]. To implement this information system for Kanhan River, around 300Kms of Kanhan River stretch from Amla Dam, Madhya Pradesh to Gosekhurd Dam, Maharashtra, India (figure 1) is considered along with the water sources confluences the river. Research team is kin in conserving this river, hence developed a web-based information system to drag common people awareness.

Figure 1. Considered Kanhan River path

Various physicochemical parameter like PH (Potential of Hydrogen), turbidity, water conductivity, total alkalinity, hardness, arsenic, BOD (Biochemical Oxygen Demand), Calcium, chloride, chlorine, COD (Chemical Oxygen Demand), DO (Dissolved Oxygen), ESP (Exchangeable Sodium Carbonate), FC (Faecal Coliform), Hydrogen Carbonate, Iron, Magnesium, Nitrate, Phosphate, Potassium, RSC (Residual Sodium Carbonate), SAR (Sodium adsorption ratio) , Sodium, Sulfate, TC (Total Coliform), Fluoride, Lead, Copper and Zinc are recorded for assessing river water quality, which helps mainly in river engineering. To unfold many facts associated with the river, decades of data need to be considered. To accomplish such requirement, an information system is designed and developed which generate central repository of the essential data along with data security [5].

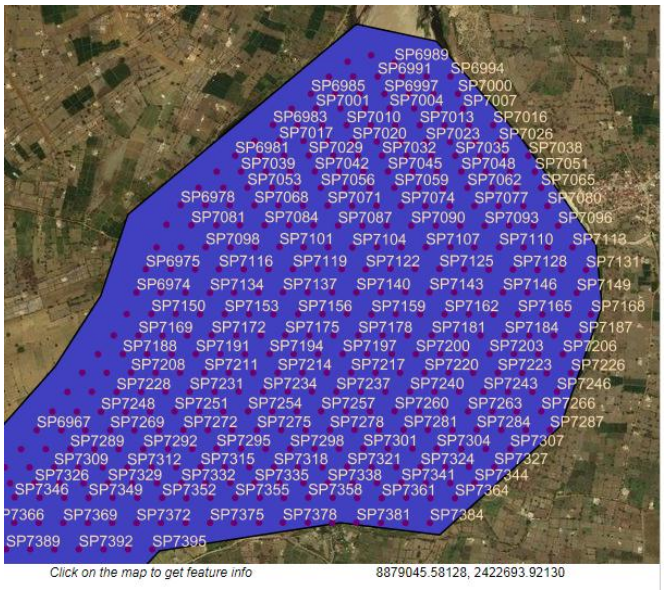

Figure 2. Study point imposed on the river path.

In view to record information along with its geo-position, SPs (Study points) are imposed (figure 2) on the river stretch at 100-meter distance. Bing Map is used as a base map, which is overlaid with system's basic maps namely: River-Path (depicts Kanhan River path), SPs (recorded water quality values depicted on or closed to SPs), and landmark (gives details of the geo-graphical position). Google Earth is used for creating polygon, point vector layer in KML (Keyhole Markup language) and KMZ (Keyhole Markup Language Zipped) format these files imported in QGIS (Quantum Geographic Information System) to create shape files. PostgreSQL 9.2 has been used for storing information and shp2pgsql utility of PostGis is used for converting shape file into PostgreSQL table. Geosever 2.11 is used for publishing the layers. Tomcat webserver is used for publishing map information and webpages. Java technology and OpenLayer programming is used for development of map features and dynamic webpages. Information system is mainly categorized in two information portals, public portal and private portal respectively. Public portal is open to all; user can observe water quality of the Kanhan River at different geographical locations before consuming river water or using river water domestic purposes. Apart from this, common people can also register as a guest-user and contribute valuable data (Since data is also getting generated through public-partnership it is termed as citizen assisted

© 2018, IJCSE All Rights Reserved

system). Private portal is accessible for authorized users termed as R&D team belongs to authorized research and development organizations. Authorized users of private portal can upload, download, extract, generate, observe and estimate data. Admin user of the system is privileged with managing and layer publishing on GeoServer. Figure 3 shows system architecture used for the development of the information system.

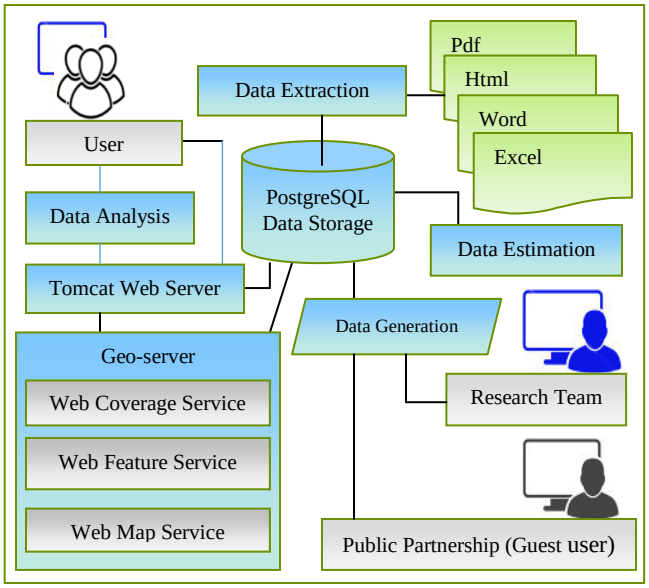

Figure 3. System architecture.

Database design:

PostgreSQL database with GIS support is used for the storing map/layer information. To identify inputs made through different modes, separate flags (Boolean) are set which is described in table-1. Whereas, table-2 depicts the table design used to store data for alkalinity layer. Similarly, several other information layers are created which follows more or less same pattern for storing information.

Table 1. Flag system for identifying data version

Sl. No.	*Flag*	*Description*
1.	guestentry	If true, data entered by guest user
2.	generation	If true, then data generated by registered R&D team.
3.	estimation	If true, then data generated by data estimation method
4.	confirmed	If true, then only data will be visible to public portal generated by guest user.
5.	extraction	If true, then data generated through data extraction method

Table 2. Table structure to store alkalinity information layer.

Gid	Name-Data	Geom	guestentry	generation	estimation	Confirmed	extraction
1	7.3	01010000E0E6100000A1D6A7316188534058 0853B31DEB354	T	F	F	T	F
2	7.5	01010000E0E610000097D5156869895340BC 80E64E0EEA354	T	F	F	T	F
3	7.5	01010000E0E610000083C74D FCFE895340B E33C8AC74EA354	T	F	F	T	F
4	7.6	01010000E0E610000039E399 A08E895340726F244B74EB354	T	F	F	T	F

1. DATA EXTRACTION

River modelling mostly require numerical data; hence only structured data is considered for the extraction. Having such defined frame, customized algorithm is developed for the data extraction. Whereas general available free libraries / utilities are used for data extraction from pdf, word, and excel. Customized algorithm follows the extraction techniques as below.

1.1 Data extraction from text file:

Text files are available in comma (,), semicolon (;), and colon (:) separated forms; several utilities are already available to convert such text files into required format, hence these files are not in scope of this extraction algorithm. Developed algorithm follows extraction technique, which can extract data from non-delimited files. Algorithm collects column-widths from the user for each column and generate excel file. Algorithm assumes fix width for each column, thus requires fix width; If not, received null-string will be replaced with zero value for the respective columns. Algorithm extract data line by line progressively till the end of file.

1.2 Data extraction from html file:

Web-page designer mostly follows different style for designing web documents; as a result of this, data extraction from HTML is tough job. There are no specific recommendations for designing web-pages, style attributes of webpage can either be provided along with HTML tags or separate CSS (Cascading Style Sheets) file can be used for designing web documents. Every algorithm faces a trouble with data collection and follows different approach in collecting data [6]. Extracting data from webpages mostly

© 2018, IJCSE All Rights Reserved

requires human intervention; expected accuracy in the extraction can't be maintained, if complete auto extraction followed. Considering such limitations, developed customized algorithm extracts data occurs in between outer <Table> </Table> tags of HTML, as only structured data extraction is considered.

Developed algorithm follows pre-order tree traversal technique, it traverse input file top to bottom. Ignores HTML tags till the occurrence of <table> tag. Counts number of lines occurs in between outer <table> and </table>, these counted lines are the required data, simple filtration techniques are applied to obtain essential data and subsequently output is stored to excel file. Microsoft excel provides various data filtration and data tuning features which is utilized for obtaining refined data. Images provided in <td> tag are not counted as a structured data, hence <img> tag is ignored while extracting data. Schematic diagram (Figure 4) along with tree structure gives general concept followed for the development of algorithm.

Sample Html Code

```
<html>
<head>
<title></title>
</head>
<Body>
<table>
 <TR>
   <td>
      <table>
        <tr>
<td>AA</td><td>AB</td>
        </tr>
        <tr>
              <td>BB</td><td>BC</td>
        </tr>
      </table>
   </td>
 </TR>
  <TR>
       <td>CC<td>
  </TR>
   .
   .
  <TR>
     <td>xy</td> …. <td>yz</td>
  </TR>
</Table>
</Body>
</html>
```

Ignored tags

Ignored tags

	A	B	C	D	E
1	AA	AB			
2	BB	BC			
3	CC				
4	XY	YZ			

Sample Output:

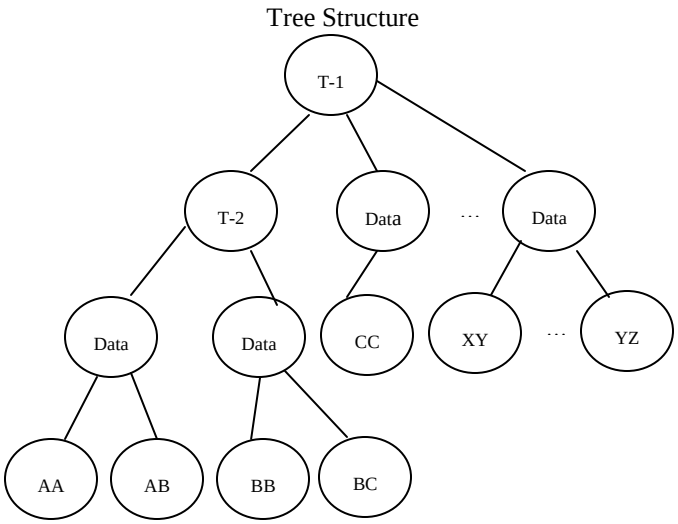

Figure 4. General data extraction concept.

2. Data estimation

Research team collects data by means of carrying out field work. While completing field work, most of the missing SPs gets emerged where data is missing. Besides, if considered R&D work, such missing entries causes hurdle as R&D needs data in complete form. R&D work cannot be kept awaited till data completeness is attained. Therefore, such missing data essentially need to be best estimated.

Maximum Likelihood Estimation (MLE) method has been used for missing data estimation. Where x_1, x_2 ...x_n are the sampling values for the zone where estimation of sampling value required to be carried out. MLE is calculated for each value x_1, x_2 ...x_n and missing value x decided for xi (any value from the samples) for which MLE calculated is maximum. Most commonly used Mean method has not been used for data estimation as this method does not consider variance in the data and mostly calculated mean may not be a value from the samples.

(Mean) $$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{N}$$

(Variance) $$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{N-1}$$

(MLE) $$P(x; \bar{X}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{-\frac{(x-\bar{X})^2}{2\sigma^2}}$$

Where N is total number of samples, π = 3.14, x is sample value for MLE need to be calculated

© 2018, IJCSE All Rights Reserved

3. Data Generation:

R&D team usually follows random sampling technique for collecting water samples. To assess Kanhan River water quality, samples can be collected and various physicochemical factors can be measured at the different geographical locations. Instrument like GPS (Global Positioning system) can be used for getting global positioning of the SPs and accordingly uploaded on the information system. Developed information system has a feature wherein data can be uploaded in the system point by point or complete excel file also can be uploaded in the system. However, data uploaded through excel file needs administrative privileges for its confirmation.

4. Data Generation (Through public-partnership):

Most of the physicochemical parameters can be measured in the presence of water, hence simple but most valuable feedback like: presence or non-presence of water at particular geographical point can be recorded in the software using social or public partnership. This valuable feedback is very effective in generating data at various places across the river bed. Data contributed by public-partnership may have low significance in R&D activities; but on the other side, it is very emphatic in validating generated data. Public portal of the system facilitates the end-user to create guest user account and contribute data. Data submitted by guest user can be authenticated and confirmed by the system administrator of the portal. To identify data contributed by the guest user, software set guestuser flag to TRUE and by default sets confirmed flag to FALSE. System administrator can set confirmed flag to TRUE on confirming data generated by the guest user, which further gets reflected on public portal and also gets available for the R&D team.

5. Data extraction using GIS (Geographic information system) based utility software.

Google earth provides a 3D earth satellite imaginary, on which several information vector layers are imposed and sensible information can be shared among the users. Google earth provides referral latitude and longitude where points, polygons can be depicted along with associated information (Ex: polygon of locality may indicate population; whereas polygon of pond may indicate water body). Such information extracted by means of KML/KMZ file provided by Google earth. Free QGIS software is used to change KMZ/KML file into shape files which extracted into PostgreSQL using shp2pgsql utility and then published on Geo-server. Using such methodology and referral latitude & longitude provided by Google earth utilized for global positioning of data on the map.

Data analysis

While working on R&D projects, it's primarily aim to generate quality data. Team generates essential data using this web-based interactive information systems; system presents data beautifully so that it talks with the research team. Collected huge data and have nothing to reveal means generated data has no importance. To ensure data accuracy; data trend, correlation and validity need to be monitored progressively and periodically. R&D team contributes most and generates valuable data keeping various approaches in mind. Developed information system facilitates two modes for observing and validating generated data. Firstly, generated data is mapped as information layer and overlaid on satellite imaginary, this helps in estimating possible values observing nearby geographical area. Secondly, system illustrates graphical representation of generated data by means of column graph, line graph, scattered graph, area graph and bar graph; this helps in observing data trend and validate the data.

If observed column graph depicted in figure 5, it can be significantly noted that Magnesium value for sampling point lies close to 200 has Magnesium value 48 mg/L, which is higher than common trend followed by the data. Therefore, it needs special attention from the research team to validate such data entry.

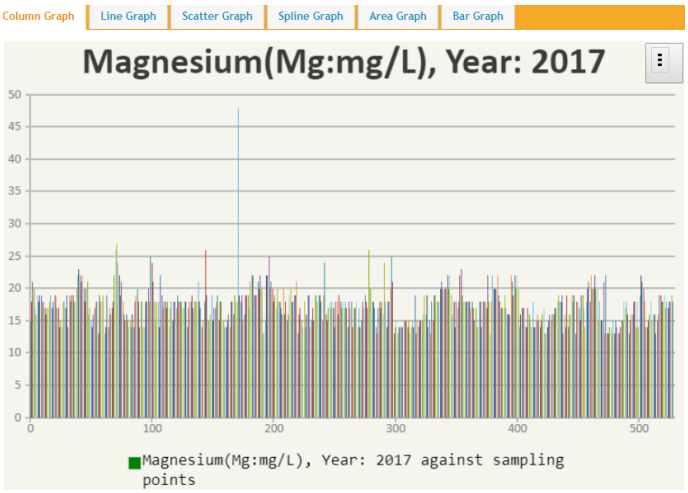

Figure 5. Sample Magnesium values at different SPs. (Column Graph).

Figure 6 to 10 shows different graph generated by the information system for observing and validating data.

© 2018, IJCSE All Rights Reserved

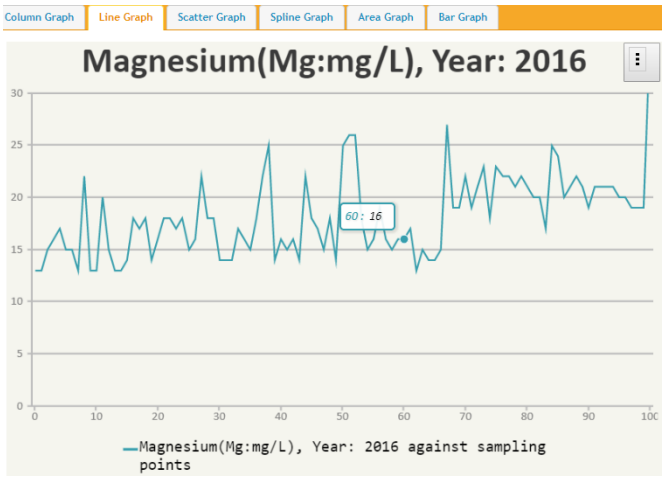

Figure 6. Sample Magnesium values at different SPs (Line Graph).

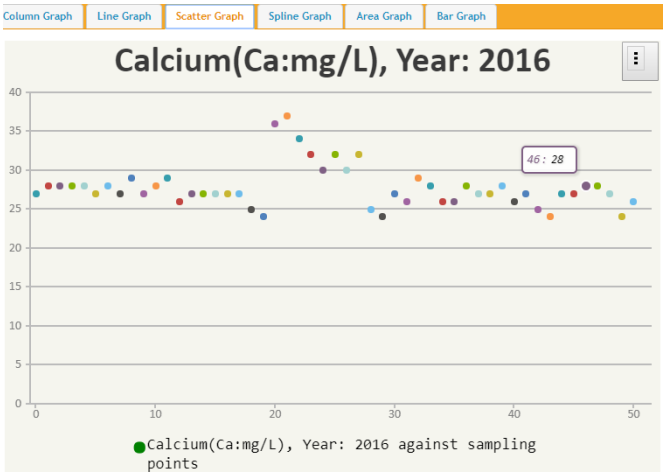

Figure 7. Sample Calcium values at different SPs (Scatter graph).

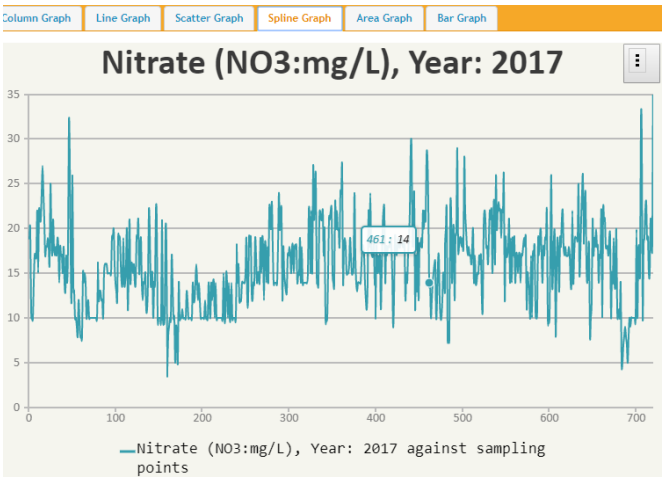

Figure 8. Sample Nitrate values at different SPs (SP line graph).

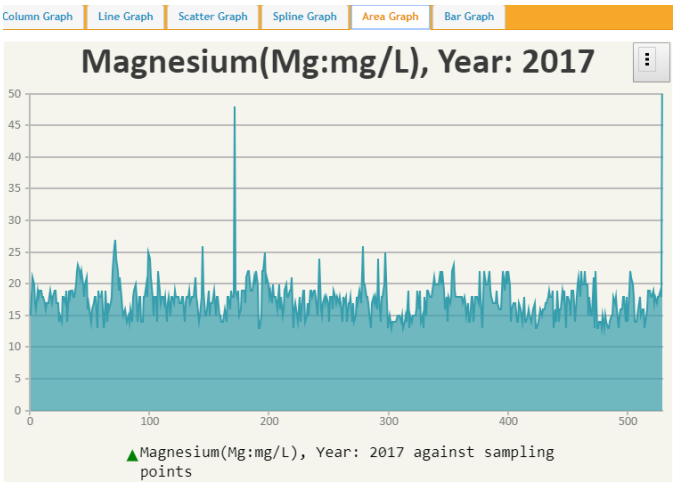

Figure 9. Sample Magnesium values at different SPs (Area Graph).

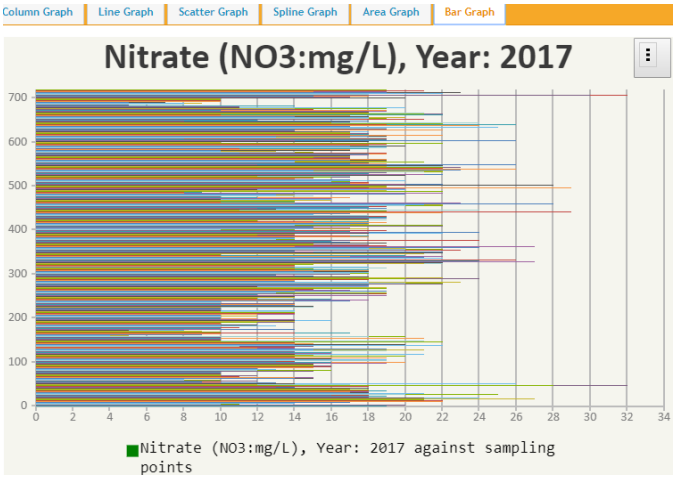

Figure 10. Sample Nitrate values at different SPs (Bar Graph).

IV. RESULT AND DISCUSSION

This research work explores different methods for data generation; using these techniques data generation and data completeness is attained. Figure 11. shows sample data recorded on or closed to SPs at Gosekhurd dam, figure 12. shows Geo-mapped data at Kanhalgaon (MH) along with River-Path, figure 13. Shows TC at Gosekhurd dam and Figure 14. Shows complete data at recorded at Gosekhurd dam. R&D requires year-wise data, hence year-wise central data repository is maintained.

Data Extraction: There is no specific recommendation for designing web pages, hence web-page designer follows own style for designing web-pages. This consequently burdens data extraction algorithm as variation in web-page designing need to be coded in data extraction algorithm. If data

© 2018, IJCSE All Rights Reserved

extraction technique is not enough smart, then it ultimately burdens the data filtration and tuning techniques. Therefore, data extraction algorithm need to be enough smart and shall bear due flexibility so that appropriate data extraction can be done.

Data Estimation: While using MLE research team need to identify the data segments (Ex: group of around 50 SPs) which follows quite similar pattern and then apply MLE for estimation. From the general study, it is inferred that population(N) value shall range from 10 – 50 for best estimation. In support to this, it is also observed that the segment considered for the estimation shall have around 50-60 percent known values and then apply MLE for missing value for the better result.

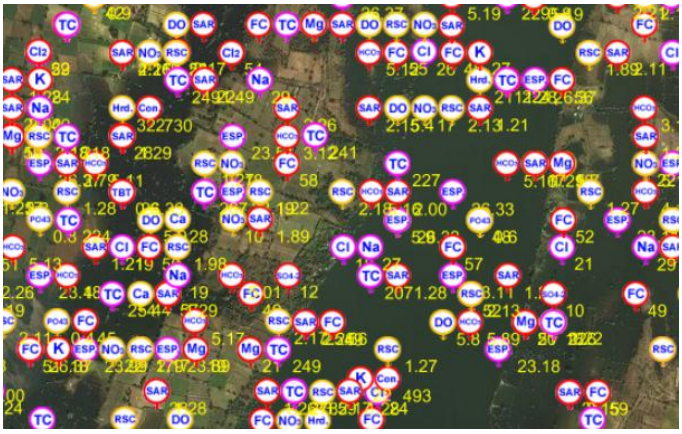

Figure 11. Geo-mapped data at Gosekhurd dam

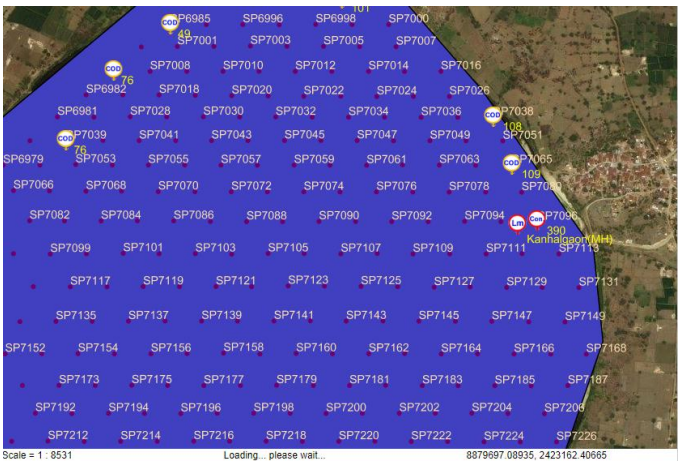

Figure 12. Geo-mapped data at Kanhalgaon (MH) with River-Path

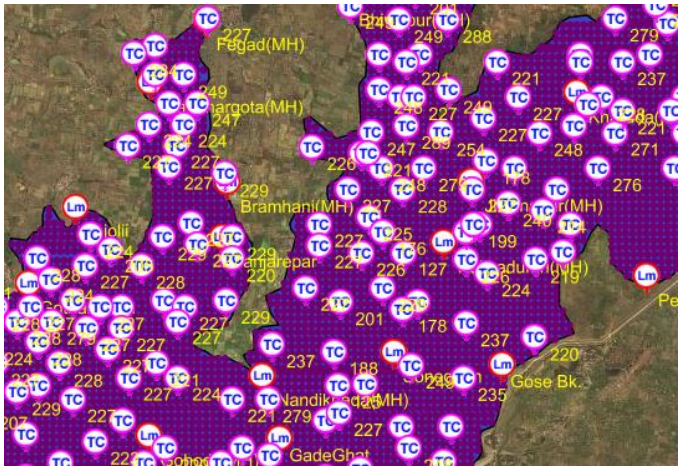

Figure 13 geo-mapped data at Gosekhurd dam Indicates TC

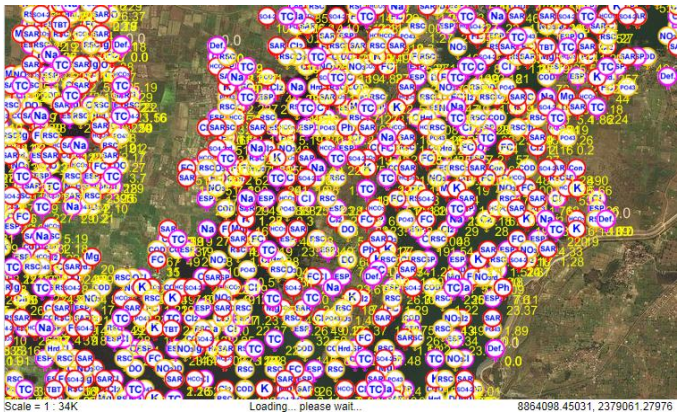

Figure 14. Geo-mapped data at Gosekhurd dam depicts complete data

Figure 15. Default layer with null data at Gosekhurd dam along with Landmark

V. CONCLUSION AND FUTURE SCOPE

Excess exploitation of natural resources is big threat to the environment, so happening with the kanhan River. This

research work helps to drag awareness in conservation of natural resources. Generally, R&D activities take adequate time to hit public domain and benefit them, whereas public portal of this information system immediately shares approved river water quality data on its public portal.

Used techniques has great scope in automated data generation, provides seamless features to generate and validate variety of data. This enables research team to plan variety of R&D projects; which gives a scope where one research activity concludes, may be beginning for several other R&D activities.

This information system bears great expandability. To extend information, system provides default layer. Figure-15 shows default layer, which holds null values. Default information layer can be cloned for new information generation. Introducing such feature, several information layers can be created and seamless information can be generated. Currently this system has been designed for Kanhan river, which can be extended for any river by incorporating few changes. Following are the challenging areas wherein such developed data can be utilized effectively.

- Bio-diversity study
- River life forecasting
- Reverse engineering
- River water and health management
- Average rainfall and water carrying capacity of the River
- Dams and its impact on bilateral river and its tributaries
- Run-off river water contamination
- Policy decision making

A. Abbreviations and Acronyms

R&D	Research and development
GIS	Geographic information system
SPs	Study points
KML	Keyhole Markup language
KMZ	Keyhole Markup language Zipped
QGIS	Quantum Geographic Information System
PH	Potential of Hydrogen
BOD	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
ESP	Exchangeable Sodium Carbonate
FC	Faecal Coliform
RSC	Residual Sodium Carbonate
SAR	Sodium adsorption ratio
TC	Total Coliform
GPS	Global Positioning system
CSS	Cascading Style Sheets
HTML	Hypertext Markup Language
MLE	Maximum Likelihood Estimation

ACKNOWLEDGMENT

I wish to express my sincere gratitude to Dr Rakesh Kumar, Director CSIR-NEERI for providing me an opportunity to complete Ph.D. work.

I sincerely thank to Dr Girish S. Katkar, Professor and Head, Department of Computer Science and Application, Art, Commerce & Science College, Koradi, Nagpur for his guidance and support to carry out this research work.

I sincerely thank to Dr R. B. Biniwale, Sr. Principal Scientist and Head, Cleaner Technology and modelling division, NEERI for his encouragement and support to carry out Ph.D. work.

Lastly, I would like to express my gratitude to the scientist and engineers working with me for their guidance and support.

REFERENCES

[1] Soumita Mitraa, Swayambhu Ghosh, Kamala Kanta Satpathy, Bhaskar Deb, Bhattacharya, Santosh KumarSarkar, PravakarMishra, P.Rajae, "Water quality assessment of the ecologically stressed Hooghly River Estuary, India: A multivariate approach", 126(2018) 592-599

[2] Stuart E. Bunn, Angela H. Arthington, "Basic Principles and Ecological Consequences of Altered Flow Regimes for Aquatic Biodiversity1", DOI: 10.1007/s00267-002-2737-0

[3] Nilamadhab Mishra, "Internet of Everything Advancement Study in Data Science and Knowledge Analytic Streams" International Journal of Scientific Research in computer science and Engineering, Vol.6, Issue.1, pp.30-36, Feb (2018)

[4] Dr. G. K. Khadse, P. M. Patni, P.S. Kelkar, S. Devotta, "Qualitative evaluation of Kanhan River and its tributaries flowing over central Indian plateau", Environ Monit Assess. 2008 Dec; 147 (1-3):83-92. Epub 2007 Dec 22.

[5] Poonam Devi, "Attacks on Cloud Data: A Big Security Issue", International Journal of Scientific Research in Network Security and Communication, Volume-6, Issue-2, April 2018

[6] P.V. Nikam, D.S. Deshpande, "Different Approaches for Frequent Itemset Mining", International Journal of Scientific Research in computer science and Engineering, Vol.6, Issue.2, pp. 10-14, April (2018)

Authors Profile

Mr. Dinesh A. Lingote pursed B.Sc. (Computer Science) from Amravati Univeristiy, pursed M.C.A. from Amravati university and currently pursuing Ph.D in RTMNU, Nagpur, Under the Guidance of Dr. Girish Katkar, Department of Computer Science and Application, Art, Commerce & Science College, Koradi, Nagpur. He is currently working as Sr. Scientist in CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur 440020. CSIR-NEERI is a Constituent laboratory of CSIR(Council of Scientific and Industrial Research, Anusandhan Bhawan, 2 Rafi Marg, New Delhi, Delhi 110001) The Renouned Institutttion is emerged as significant solution provider in the area of environmental science and engineering for the sustainable development. CSIR-NEERI have highly qualified and experienced

© 2018, IJCSE All Rights Reserved

scientist and engineers, who have developed several technologies and published quality research papers in environmental sicence and engineering area. He is not only involved in research and development, but also administers IT infrastructure of the Institution. He is currently working on auto data generation and extraction techniques to generate large data for environmental science and engineering.

Dr Girish S. Kakar, pursed M.Sc., PGDCS, MCM, He has pursed Ph.D. in computer science from RTMNU, Nagpur. He is currently working as Professor and Head of Department of computer science and application, art, Commerce & Science College, Koradi, Nagpur. He has guided several Ph.D. scholars, and published several quality research papers. He is currently working on data mining.

Dr. Ritesh Vijay pursed B. E., M. Tech Environmental Engieering and pursed Ph.D. in Environmental Engineering. He is currently working as Principal Scientist and Head, Mumbai Zonal Center of CSIR-NEERI. He has worked on several prestigious projects. He has published several quality research papers. He has commendable awards on his name. His specialised area is application of remote sensing and GIS in environmental science and engineering.

Dr. Rajesh Biniwale pursed M. Tech in chemical engineering, pursed Ph.D. in Chemical engineering and pursed Post. Doc in chemical engineering. He is currently working as Sr. Principal Scientist and Head, Cleaner Technology and Modeling division of CSIR-NEERI. He has worked on several prestigious projects. He has published several quality research papers. He has developed creditable technologies, He has commendable awards on his name.

© 2018, IJCSE All Rights Reserved