ISSN No. 0976-5697

Volume 4, No. 3, March 2013 (Special Issue)

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Geo Location Big Data Based Collaborative Crowd Sourced Data Mining Architecture for Environmental Monitoring and Vegetation Management Systems

Samrat P. Patel
B.E. in computer Engineering
.( Database Developer, IT consultant),
Thadomal Shahani College Of Engineering Mumbai,
Computer Society of India, Mumbai University
Mumbai, India
Samratppatel@gmail.com

Sumeet S. Deshmukh
Bsc Information Technology.
(Analyst/SAS Programmer, Datawarehouse Developer)
Karnataka State Open University, Mysore
NIIT,Mumbai.
Mumbai, India
sumeetcos@gmail.com

Anil R Rajbhar
Associate Scientist- Ph.d Research Scholar (Botany)
St' Xavier's College Mumbai
Mumbai, India, Mumbai University ,
taxonomy.rajbhar@gmail.com

***Abstract*** The proposed paper gives an Idea regarding the Geo Location based vegetation data storage repository for purpose of smart environmental management. The proposed system overcomes the shortcomings in the current systems and software tools used for the said purpose of environmental management and analysis [28]. The system proposed can function as a full fledge Environmental Management and Audit System and abide by operating principles and rules of any traditional EMS i.e Plan-Do-Check-Act Cycle [8][14][15].
Our system is extension of the above mentioned concept and proposes a collaborative model for community based crowd sourced environmental data storage and management with help of big data storage coupled with Data Mining & Analysis tools[28[29]. It is used along with highly scalable database management and storage of historical and highly dynamic environmental monitoring data in order to tackle serious problems related to the environment by studying data and the factors affecting the environment for that location .Effective management of the environmental resources is possible by mapping of real time vegetation data and various other forms of environmental based data. this will help the corporate ,government and the concerned authorities to take smart and informed decision also simultaneously educating the masses using web as a medium of propagation of information and promoting a social sense of responsibility for every citizen to contribute for the betterment of the environment of the city.

***Keywords:*** Location Based ,Database ,Big Data Storage and Analysis, Environmental Management System Architecture, Data Mining, Data Analysis, Geospatial, Crowd Sourced Monitoring Architecture, Remote Sensing, Sensor Networks ,Geo Information Systems, GPS, Vegetation Database, Environmental Planning, Mobile Application, Smart Phone, Tablet

## I. INTRODAUCTION

**Urbanization** or **urban drift** is the physical growth of urban areas as a result of rural migration and even suburban concentration into cities, particularly the very largest ones. The United Nations projected that half of the world's population would live in urban areas at the end of 2008.Urbanization is closely linked to modernisation, industrialisation, and the sociological process of rationalisation[15]. So the term urbanisation can represent the level of urban relative to overall population, or it can represent the rate at which the urban proportion is increasing. More importantly, it's about a complete change from rural to urban style in terms of industry structure, employment, living environment and social security [1]. The rapid urbanization of the world's population over the twentieth century is described in the 2005 Revision of the UN World Urbanization Prospects report. The global proportion of urban population rose dramatically from 13% (220 million) in 1900, to 29% (732 million) in 1950, to 49% (3.2 billion) in 2005. The same report projected that the figure is likely to rise to 60% (4.9 billion) by 2030[1] [2] [3].

In regard to future trends, it is estimated 93% of urban growth will occur in developing nations, with 80% of urban growth occurring in Asia and Africa Urbanization rates vary between countries. The United States and United Kingdom have a far higher urbanization level than China, India, Swaziland or Niger, but a far slower annual urbanization rate, since much less of the population is living in a rural area[1].
With the growing urbanization surely it has its effect on the environment best example is the deforestation of the forest areas which are being converted for human habitation due to which wild animals are escaping the forests and entering the civilian space .The phenomenon of Urban heat islands has become a growing concern. Incidence of this phenomenon as well as concern about it has increased over the years. An urban heat island is formed when industrial and urban areas are developed resulting in greater production and retention of heat. A large proportion of solar energy that affects rural areas is consumed evaporating water from vegetation and soil. In cities, where there is less vegetation and exposed soil, the majority of the sun's energy is absorbed by urban structures and asphalt. Hence, during warm daylight hours, less evaporative cooling in cities results in higher surface temperatures than in rural areas [4]. Vehicles and factories release additional city heat, as do

industrial and domestic heating and cooling units [5]. As a result, cities are often 2 to 10 °F (1 to 6 °C) warmer than surrounding landscapes [6]. Impacts also include reducing soil moisture and a reduction in re-uptake of carbon dioxide emissions [5]. In today's world forests are considered as an asset for the rich biodiversity.

There is need to keep track of the growth and expansion of the forest regions with the help of identification and tracking the growth of various plants spices and real time demarcation of individual trees present in the forest ecosystem. There is also a need to manage the forest resources in case of disaster in the forests like natural calamities and forest fires [23].

Overall the environmental pollution today has assumed a serious and gigantic proportion. Not only this, but also the very existence of human society is threatened. The socio-economic development has reached at a point, its speed has so increased and its destructive power is so enhanced that it is decimating the environment rapidly. Therefore it has become necessary to find-out such new paths and methods of development which will not destroy ecology and produce pollution [15] [16].

Environmental management precisely addresses above mentioned problems. If one can manage environment properly, one can reduce pollution or even put a stop to it.

### *A. Few of the factors that add to environmental pollution and degradation*

a. Lack of awareness amongst the public, industrial and the government authorities
b. Water pollution due to disposal of waste in the rivers , ponds , sea [25] [26]
c. Clogged sewage pipelines and manholes
d. Unmanaged solid wastes and other form of wastes i.e. biodegradable or non degradable products
e. Traffic congestions at places
f. Cell tower radiation at various locations
g. Deforestation due to urbanization
h. Increased population density for the location and lack of space
i. Pollution due to burning of fire crackers
j. Pollution due to mining activities
k. Noise pollution due to traffic and industrial machinery
l. Pollution of the coastal lines due to non degradable waste such as plastic bags , and plastic articles and immersed idols during the festival seasons
m. Pollution of the oceans and coastal regions due to crude oil spills and harmful chemicals

### *B. Drawbacks of the current system and supporting software tools:*

The current system accepts the data in only particular format i.e text, csv, xls etc where as the incoming data can be in video , image files ,audio etc. Limitations in accepting various data formats narrow downs the scope of the data analysis which is required for decision making by the organizations. More the data more accurate will be the analysis and smarter will be the final decision.

The current system does not allow collaboration of various sources of data and promote collaborative knowledge sharing between the organizations, individuals and the government.

Since the current traditional systems does not encompass the data from various other sources responsible for environmental resource degradation the decision and planning process can be affected as a result as they i.e (factors that lead to pollution ) can be connected or there may exist some degree of correlation between each other.

Since the data regarding the vegetation and ecosystem for a particular location is very dynamic in nature it requires monitoring from time to time and surveys need to be conducted frequently by the organizations in order to maintain the quality of the data time to time and since the current system is not collaborative in nature it can become a humongous task for the organizations to conduct the surveys regularly in such a situation there is a possibility of generation of stale data entry within the system as the data is not update on a regular basis wide gaps in the ecology monitoring survey sessions can lead to insufficient data for that location.

There are also chances that there is data redundancy and the data is repeated and data can also be erroneous due to human error or some or the other problem that may arise in the survey systems ( GPS , Databases , Web Portal for Survey , GIS etc.) at that point of time.

### *C. Description Of The Proposed System*

Conserving the environment has become a global issues and it has become a responsibility for every citizen and the government authorities to do the needful procedures to make Our planet habitable for the years to come. The below diagram describes proposed collaborative system architecture for environmental management and modeling which can help in real-time management of the ecological resources in the given locations and help the government bodies and the concerned authorities to take smart decisions based on the real time dynamic historic and present data and help to mitigate the risks involved due to increase in the pollution or conditions that lead to damage of environmental resources with the help of Data Mining technologies [28] [29] . The proposed system helps in mitigation and management of the risks by means of identifying the area where there is need for improvement or controlling of the usage of the environmental resources.

The proposed architecture has following main components which form the crux of the system.

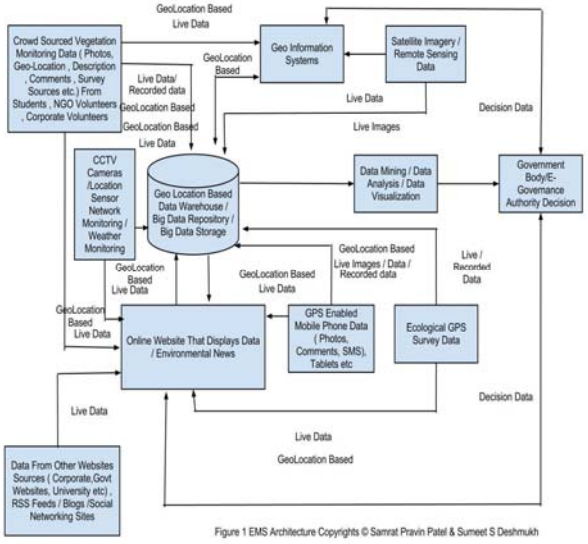


Figure 1:





CONFERENCE PAPER
II International Conference on
"Advance Computing and Creating Entrepreneurs (ACCE2013)"
On 19-20 Feb 2013
Organized by
2nd SIG-WNs, Div IV & Udaipur Chapter , CSI , IEEE Computer Society Chapter India Council , IEEE Student Chapter Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

### *a. Data Acquisition and Secured Big Data Storage:*

The Geo Location Based Structured and Unstructured data arrives from various sources. The data formats can be files, videos, emails; text etc. is stored in the Data Warehouse Big Data Storage System for further analysis and usage [9] [10]. The sources of geo location based data acquisition can be

a) Crowd sourced Data from the Volunteers from NGO's for environment
b) CCTV Surveillance Cameras based at various locations
c) Crowd sourcing Data from public (Students, Office goers, Pedestrians etc.)
d) UAV and Surveillance Drone's
e) Geo Location Based Live Traffic Data from road traffic authorities for particular location to keep the track of the pollution build up passing the real time data from the traffic control centre or on duty
f) Various Database Sources , Local Survey Data and Centralized Storage Data etc
g) QR Codes
h) Social network communities based in that particular area or communities or groups based in location of interest for environmental development
i) Data contributed from various mobile devices such as smart phones , laptops, tablet pc's
j) Satellite Imagery
k) GIS for decision making for traffic conditions, population density , waste management etc
l) Location Based Sensor Networks for environmental monitoring ( Measure poisonous gases and Carbon Monoxide (CO) and Carbon Dioxide (CO2) emission)
m) Public News Websites, Government Web Portals, Blogs , RSS feeds etc.
n) Public Survey or census Data

### *b. Data Mining and Analysis:*

It includes the collection real-time data and instantaneously analyze that data and deploy strategies that can help in maintaining the ecological balance. Some of the goals that fall in the scope of Data Mining are [28] [29]

a) soil and groundwater remediation and analyzing how cost can be reduced by detecting the leaks in underground water tank storage [12]
b) Greenification of the industrial belt by way of tree census of various species analysis and soil sample analysis [28]
c) Survivability analysis of plant species and which green plantation could be the best in that particular location measuring stress due to draught [19] [28]
d) Health of green cover
e) Vegetation pattern analysis and detection of any factors affecting the environment and affecting the growth of trees or destroying them [18]
f) Balancing the ratio of green cover in the city to curb the effects of pollution due to traffic or the electromagnetic radiations from the cell phone towers. Also deciding upon which plants can improve the air quality inside homes [13]
g) Better planning of the city and industrial infrastructure
h) Finding which is the best technology for communication or IT Services that can sustain without harming the environment

### *c. Decision Support and Planning:*

With the help of data mining and data visualization the government authorities will be able to take smart decisions and help managing natural resources. This will aid the government to reform various environmental policies and form rules and regulation for the masses to prevent the degradation of the ecological balance with the help of time to time updates received from various sources for data acquisition.

### *d. Public Outreach Program For Environmental Education:*

This stage involves public participation and knowledge sharing amongst the public and the government authorities through the medium of news updates or timely surveys that can be done in conjunction with public , government , NGO's and the corporate . The data can be made available for the education purpose to the masses via various means of communication such as dynamic data web portals, web applications etc**.** The data collected through the public outreach program can also serve as key factor for decision making by the government authorities

### *e. How Proposed Collaborative Model Is Better:*

The current collaborative model encourages data acquisition and storage from various sources and the data is stored in structured format once it arrives

a) Social Networks
b) Crowd sourced Public Data from NGO Volunteers, Corporate, Students, Government Officials, and Associations etc.
c) Various Web Services, web content from blogs , emails, documents etc.
d) Sensor Data from various locations
e) Survey Data , Contributed Data in various files hence there is easy management of structured and unstructured data
f) Data contributed from various mobile devices such as smart phones , laptops, tablet pc's
g) User contributed content such as news, images, comments
h) User recorded voice / videos etc

Reduction in cost for storage infrastructure as the data shall be stored as Big Data instead of storing the data in traditional Relational Database Management Systems with proper classification of the data from various data sources [9] [10].

Widened scope for data analysis as the data is from various sources stored in a centralized repository and widened scope of analyzing relationships between to data thus helping in accurate decision making for the organizations and the concerned authorities Data updated and filtered time to time so that errors and redundancy can be mitigated so there will be lesser outliners and better will be the analysis for taking the right decision Easily implementable by any organization or by the government authorities for formation of their environmental policies and for planning of environmental strategies for the management of the natural resources that can be a boon to the benefit of mankind Multipurpose application in various branches of environmental management right from social to ecological purpose .The proposed can help in restoring the balance in the ecosystem by proper natural resource management [17].

## II. APPLICATIONS

a. Collaborative Crowd Source Monitoring For Forest, Vegetation in the city.
b. Education of masses and promoting eco-activism programs for environment protection
c. Exploring and monitoring the changes and degradation in the landscape cover and for purpose of generation of the green belt for industrial or barren regions with the help of statistical analysis and data mining tools [20][21][28][29]
d. Planning exploration for education programs for environmental research
e. Preservation of Tribes and community heritage and archaeological heritage sites [23]
f. Planning , Decision making and resource management during the time of disaster
g. Preservation of endangered species of plants and animals by monitoring and data analysis of their habitat and to ensure maintenance of viable population of wildlife [23]
h. Monitoring the forest regions and preventing the encroachment and deforestation of forest space by the civilians [23]
i. Wetland management in case of construction of industrial structures constructed on a wetland [24].
j. To protect key watershed areas of Lakes for drinking water [25] [26].
k. Waste land management for disposing of the waste.
l. Coastline conservation and management during the oil spills [22] [27]
m. Management and conservation of surrounding areas near the mining areas
n. Managing of the vegetation with area wise sectors and demarcation techniques
o. Identifying hotspots for green plantation / green belt development / pollution control due to traffic, noise, electromagnetic pollution due to mobile phone network access points, non biological waste, waste water management, drinking water management, and green energy generation resource locations etc. with the help of public participation.
p. To help the government for formation of new ecology related policies , rules and regulations to protect the environment
q. Precision farming applications / industrial infrastructure development and decision making
r. Community development and medical help camps for awareness
s. Planning public safety and preventing casualties of public by identifying the accident zones or factors that lead to heavy public loss
t. To promote and encourage tourism whereby throwing open certain areas to the tourist for aesthetic, educational and recreational purposes for creating awareness amongst the tourist by interpreting the unique features of the wildlife nationals park through various media
u. For transportation planning and logistics especially when in case of the transportation routes pass through the eco-sensitive zones such as forest, lakes, heritage sites etc.

## III. BENEFITS

a. The model is open to all the groups within the society by allowing real time knowledge sharing with increased flexibility in data collection methods and techniques with improved management of structured and unstructured data.
b. Real Time monitoring of the vegetation cover with help of public participation.
c. Accurate Location based vegetation data repository due time to time updating and monitoring by the public
d. Advanced Geo Location Based Environmental Intelligence Decision Support System
e. Less resources are used for the infrastructure of the system
f. Improved EHS/ emergency response training, awareness, involvement & competency [11]
g. Increased management and operational efficiency for the organizations and the volunteers
h. Improved relationships with regulators and the community
i. Reduced waste production & energy consumption
j. Increased transparency and ability to respond to community requests for information
k. Demonstration of management control over highly complex and varied activities

## IV. CONCLUSION

With the growing rate of urbanization it has become necessary for the government authorities to plan the natural resources and industrialization of the locations intelligently. The proposed system presents a flexible model cum architecture for collaborative participation of the public, corporate as well as the government to help in real time monitoring, planning and conserving the natural resources of the ecosystems and to maintain the balance with the help of education outreach and real time knowledge sharing of the natural resources present for at various locations. The system proposed abides by the rules of a traditional Environmental Management System Architecture and also at the same time is multipurpose, highly adaptable system to sudden changes in data and can be also used by the government and the concerned authorities for environmental decision support by combining results and reports from various sources with the help of Big Data Storage technology.

## V. REFERENCES

[1]. http://en.wikipedia.org/wiki/Urbanization
[2]. "UN State of the World Population". UNFPA. 2007.
[3]. Ankerl, Guy (1986). Urbanization Overspeed in Tropical Africa. INUPRESS, Geneva. ISBN 2-88155-000-2.
[4]. Park, H.-S. (1987). Variations in the urban heat island intensity affected by geographical environments. Environmental Research Center papers, no. 11. Ibaraki, Japan: Environmental Research Center, The University of Tsukuba
[5]. http://gtresearchnews.gatech.edu/newsrelease/china-climate.htm
[6]. http://www.epa.gov/heatisland/

[7]. http://gtresearchnews.gatech.edu/newsrelease/china-climate.htm
[8]. http://en.wikipedia.org/wiki/Environmental_resources_management
[9]. http://webinfo.raidinc.com/hpc-unified-scalable-storage-white-paper
[10]. http://webinfo.raidinc.com/hpc-white-paper-cluster-parallel-file-systems
[11]. www.lgean.org/documents/EMSGuidebookPA.pdf
[12]. www.iwaponline.com/jh/009/0107/0090107.pdf
[13]. http://www.ehow.com/how_6673509_protect-yourself-cell-towers.html
[14]. Environmental quality, environmental management systems and environmental performance related to procured printed circuit boards of an original equipment manufacturer Electronics and the Environment, 1999. ISEE -1999. Proceedings of the 1999 IEEE International Symposium on Date of Conference: 1999 Author(s): Nagel, M.H. Fac. of Ind. Design Eng., Delft Univ. of Technol
[15]. Industrial waste minimization and environment management system in MXIC Date of IEEE Conference: 16-17 Jun 1998 Author(s): Chang, K. Macronix Int. Inc., Hsun-Chu Huang, L.M. ; Huang, G.J. ; Shyu, J.L.
[16]. Implementation of Environmental Management Systems in Swedish industry Date of IEEE Conference: 1-3 Feb 1999 Author(s): Backmar, J. Dept. of Machine Design, R. Inst. of Technol., Stockholm Ritzen, S. ; Norell, M.
[17]. Urban Innovation Based on the Environmental Management This paper appears in: Environmental Science and Information Application Technology, 2009. ESIAT 2009. International Conference on Date of IEEE Conference: 4-5 July 2009 Author(s): Yang Tong Sch. of Manage., Xian Polytech. Univ., Xi'an, China Wang Nengmin
[18]. Midlevel Riparian Vegetation Mapping Using Data-Mining Techniques: Support of Coalbed Methane Leasing Analysis Denise Laes, Paul Maus Remote Sensing Applications Center Salt Lake City, UT Jeff Dibenedetto Custer National Forest Billings, MT
[19]. A Data Mining Approach to Monitor Vegetation Stress Due to Drought : integrating satellite, climate, and biophysical data over the U.S. Central Plains Dr. Tsegaye Tadesse National Drought Mitigation Center, University of Nebraska-Lincoln, NE 68583-0749 www.case.ibimet.cnr.it/MTPprogramme/data/doc/4ClimateAnalysisTools/DataMiningApproach.pdf
[20]. Land Degradation Detection, Mapping, and monitoring in the Northwestern part of Hebei Province, China, Using RS and GIS Technologies (Map Asia 2004) Ayad Mohammed Fadhil Al-Quraishi, Guang Dao Hu, Jian Guo Chen http://www.gisdevelopment.net/application/environment/conservation/ma04153.htm
[21]. Environmental Change Monitoring - A Case Study in the Region of Yinchuan, Ningxia, China (Map Asia 2004) Weicheng Wu http://www.gisdevelopment.net/application/environment/conservation/ma04031.htm
[22]. A Digital Oil Spill Sensitivity Atlas for Mauritius using GIS H. Runghen, M. Bhuruth and S.D.D.V. Rughooputh http://www.gisdevelopment.net/application/environment/conservation/env_con001.htm
[23]. Global Positioning System GPS and its application in Forestry (Map India 2003) Maitreyi Mandal http://www.gisdevelopment.net/application/environment/conservation/mi03169.htm
[24]. Vegetation Mapping of Kushiro Wetland in Northeast Hokkaido, Japan: Application of SPOT images, Aerial Balloon Photographs and airborne color near infrared ( CNIR ) Images for classification http://www.gisdevelopment.net/application/environment/wetland/envwm005.htm
[25]. Geospatial data for wetland mapping and capacity building in southern Africa (Map Africa 2006) Lisa-Maria Rebelo, Max Finlayson, Matthew McCartney http://www.gisdevelopment.net/application/environment/wetland/maf06_10abs.htm
[26]. Arsenic Mapping for North 24- Pargana District of West Bengal using GIS and Remote Sensing technology (Map India 2003) Balen Basu, Samik Sil http://www.gisdevelopment.net/application/environment/water/mi03204.htm
[27]. A study on industrial growth along coast line of Orissa and its undesirable effect on environment using GIS - an olive ridley (sea turtle) perspective Vibhu Sinha ,Ateet Vatan Bahmani http://www.gisdevelopment.net/application/environment/wildlife/mi08_250.htm
[28]. Data Analysis in Vegetation Ecology, Otto Wildi WSL Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland
[29]. Data Mining as a Tool for Environmental Scientists Jessica Spate a , Karina Gibertb , Miquel S` nchez-Marr` c , Eibe Frankd , Joaquim Comase , Ioannis Athanasiadisf , Rebecca Letcherg