Implementation of Data Warehousing Tool with Data Mapping for Environmental Research

Ana Elena L. Conjares[1], Bobby G. Gerardo[2] and Ruji P. Medina[3]

**[1] Philippine Nuclear Research Institute
Quezon City, Philippines**

**[2] West Visayas State University
Iloilo City, Philippines**

**[3] Technological Institute of the Philippines
Quezon City, Philippines**

Abstract

The interpretation of immense data generated from numerous environmental researches enables the generation of knowledge apart from those for which the data were originally gathered for. Hence, a data management system with capability for data integration, pre-processing, production of specific datasets for external data analysis, data mining and/or support for decision making tasks is essential. This paper implements a data warehouse design that serves as management platform for the scientific data generated in a nuclear research and development facility in the Philippines and the meteorological, radioactivity, and other relevant data obtained from environmental monitoring systems around the world. The datasets are loaded into the system through a parser module with metadata information attached to each dataset. Data transformation is achieved using a pre-processing module to generate customized datasets through a retrieval module that is compatible with data analysis software such as WEKA.

Keywords: *Data Management, Data Warehousing, Metadata, ETL, Parser.*

1. Introduction

In a scientific research organization, where various research studies of the environment involving the application of nuclear/radioisotope techniques are being undertaken, data management should be central to their implementation. A well-structured storage of data is necessary for easy accessibility, transformation, retrieval and data analysis.

Research projects vary from environmental radioactivity monitoring, air pollutant source identification and apportionment, harmful algal bloom studies, studies on the availability and sustainability of freshwater and long-term radon measurements, to name a few. These projects generate volumes of data. However, it is still a common practice that these data are often not managed in a way that they could be used for other purposes other than that of the original objectives of the study and/or for sharing with other groups. Researchers are normally concerned only about the outcome of their projects, publishing the results of their studies and thereafter archiving their data in spreadsheets stored in their personal computers. Long-term availability and reusability are not usually considered or implemented from proposal to project finalization [2].

Inmon etal. [3] discussed that it is common practice for researchers to generate their own data and perform their own analysis. Thereafter, the data used sits in a dormant, protected state without any chance of being reused and further analyzed by other researchers. Hence, researchers spend time and resources collecting and refining data while time and resources for data analysis and/or knowledge discovery becomes limited. With reusability of research data, however, analysis could be done much more quickly and much less expensively [3]. Moreover, researchers could efficiently create more opportunities without the burden of data collection and repetition of efforts [4]. Having access to these data, especially those collected over a lifetime, researchers could creatively innovate from archival data sets and promote new meaning from existing data sets [4]. It is, therefore, important for organizations or institutions conducting researches in the fields of earth and/or environmental sciences to have a data management infrastructure in place that provides for an integrated and organized repository of research data including functionalities for easy access for the purpose of data sharing and collaboration with other researchers, internally or externally. This will only be possible if the existence of individual/group research project datasets are made known and made available to a wider community.

It is on this context that this study was undertaken focusing on the development of the initial functionalities of a data warehouse as a data management platform for environmental data that have potential use not only for the purpose that they were gathered for but also for future data analysis and decision making. The best foundation for relevant decisions is the data in a well-designed warehouse.

The prototype presented herein is an environmental data management system as an initial implementation of a data warehouse called Data Warehouse for Environmental Research, or simply, DW-ERes. Data warehouse is a subject-oriented, integrated, non-volatile and time-variant collection of data in support of management's decision [3]. DW-Eres, includes functionalities for loading and mapping of data from external sources, transformation functionality for pre-processing of data and the retrieval of user-configured datasets for data analyses. It also provides customized security configuration for data sharing and distribution internally in an organization and provides a dashboard for information of the datasets available in the system.

It is envisioned that the system will serve as the initial step for the publication of metadata to the Internet of environmental research datasets stored in the database with full consideration of the protection of intellectual property rights.

2. Related Works

Various data management systems for environmental research data and other scientific topics have been implemented. For example, PANGAEA (Data Publisher for Earth and Environmental Science), a web-based information system for environmental sciences used for processing, long-term storage and publication of geo-referenced data related to earth's science files, has been in existence for more than 20 years now [7]. In this system, analytical data are stored consistently together with related metainformation which follow international metadata standards. Information retrieval is either through simple engine search or the use of built-in data mining tools. It also provides data mining tools necessary for their understanding and evaluation. By 2001, the mean number of daily database queries on the PANGEA was about 60 retrievals for analytical data.

For the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) program, which has been collecting and processing into Network Common Data Form (NetCDF) format, Gaustad et al. have developed and deployed the ARM Data Integrator (ADI) [5]. ADI is a framework which automates data retrieval, preparation, and creation. It goes further than the default capabilities of data management systems by providing an environment for a structured algorithm development that expedites the development of algorithms and improve their standardizations. This is the major contribution of their work which supports three programming languages: C, Python and IDL.

Scientists at the Institute for Polar Ecology, describes their study and initial implementation of a data management system as a working repository for terrestrial biological data [4]. The motivation behind this work is for the publication of these repositories for the security and long-term availability of environmental data especially for unrepeatable sampling events. Their primary objective is to gather a combined storage of data in order to compare their contents and make additions through qualitative descriptions of soil properties which could be added and re-evaluated at any time.

The paper of Lawrence and Kruger describes a scalable architecture for managing large, real-time, scientific datasets [6]. The foundation of the architecture is a real-time data warehouse that archives metadata for querying and provides data distribution transparency. The contributions of their work include a metadata generation and retrieval system that allows researcher-defined metadata statistics to be calculated; a transparent data distribution system that uses the data warehouse to dynamically distribute the data across multiple machines; and an architecture implementation archiving weather radar data that is more usable than the system deployed by the National Climatic Data Center (NCDC) which is the United States government organization mandate. The files are distributed over web servers and retrieval is done by using its HTTP's Uniform Resource Identifier (URIs). To demonstrate the feasibility of the architecture adopted, the system was used to archive weather data generated by the NEXRAD system.

Bainbridge also introduced in his paper architecture of a simple data management system for various types of environmental observational data that implicitly allows for and promotes data integration and interoperability [9]. He cited that there is a demand for multidisciplinary datasets to be freely available through service based interfaces linked into decision support and modeling systems.

Although not an environmental data, it is also noteworthy to describe the work of Zhou et al. on the development of a data warehouse platform for the management, processing and analysis of a large structured electronic medical record (SEMR) data, Clinical Data Warehouse (CDW) [8]. The CDW incorporates the SEMR data for medical knowledge discovery and the Traditional Chinese Medicine (TCM) clinical decision support. TCM large scale clinical data is the knowledge source for TCM research.

The works cited above are descriptions of implementations of data management systems for scientific research and measurement data. They further explain the importance of having data management systems that gathers relevant data in one place to enable better decision making.

3. Work Done

The DW-Eres is a system designed for the use of environmental researchers. Figure 1 presents the process flow of research projects for the environment which aid in the identification of the system's initial requirements. Measurements or observations from scientific research projects that is usually stored in spreadsheets are the target data source of DW-ERes. These are extracted and loaded to DW-ERes and then restructured to create a dataset. The resulting dataset becomes the primary source for data analysis leading to knowledge discovery.

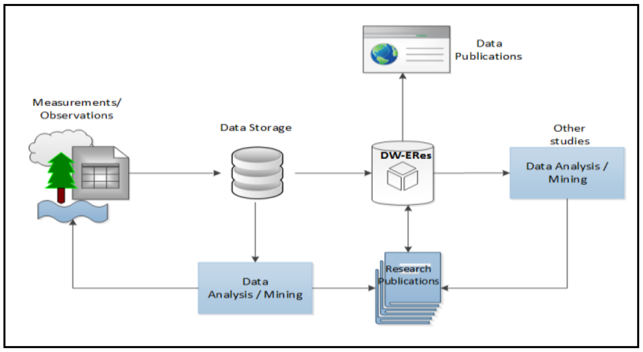


Fig. 1 Process Flow of Research Projects for the Environment

3.1 Data Model

The DW-Eres data model adopted the model of PANGAEA (https://wiki.pangaea.de/wiki/Data_model accessed on 2 July 2015) with some variations. The DW-ERes data model, running on a MySQL Relational Database Management System, is presented in Figure 2. Similar to PANGAEA's model, the main tables correspond to the major components of any environmental research undertaking. The database consists of two main tables, namely, Project (a project is created to organize and implement activities leading to the accomplishment of the project and oftentimes it is undertaken with other agencies as collaborators) and the Sampling Event (environmental measurements/monitoring activities at different locations and different environmental data are being measured using appropriate methodologies/techniques) and Dataset (the measurements from the field during sampling events and data analysis applied to them). Meta information pertinent to a project, sampling event and dataset are stored in terms of its relationship to a project and sampling event.

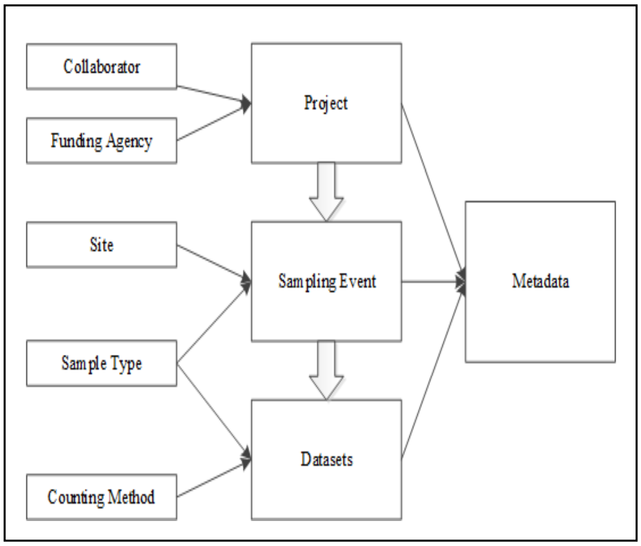


Fig. 2 DW-ERes Data Model

For every dataset uploaded into the DW-Eres, a corresponding set of metadata is attached. Metadata is captured through the following mandatory inputs:

1. Data sharing (YES/NO)
2. Type of environmental data (surface radioactivity, marine radioactivity, air pollutants, etc):
3. Title of research project
4. Brief description of project
5. Funding agency
6. Describe the various data in the dataset:
7. Provide keywords describing the dataset
8. Names of owners of the dataset
9. Provide other dataset from other work/project related to this dataset
10. Geographic area coverage (Longitude and latitude, if available)
11. Time period of the collection of dataset
12. Counting/measurement/sampling method
13. Describe the quality of the dataset in terms of missing data or any gaps found in the dataset
14. If cited by other authors, how should the source of the dataset be cited?
15. If published, title of publication and provide abstract
16. Author(s) of the metadata

In the future, the metainformation of the datasets will be made available to the public through the Internet. The system will add the functionality to allow potential users to request datasets of interest to them.

3.2 Architecture

The DW-ERes, is designed on a PHP CodeIgnitor platform, hosted in a Linux operating system with Apache as its web server and MySQL as its Relational Database Management System. The architecture of the system, as presented in Figure 3, provides an environment for the data loading of environmental measurements, input of metadata information, data cleaning, and extraction/retrieval of datasets for data analysis/mining tasks external to the system.

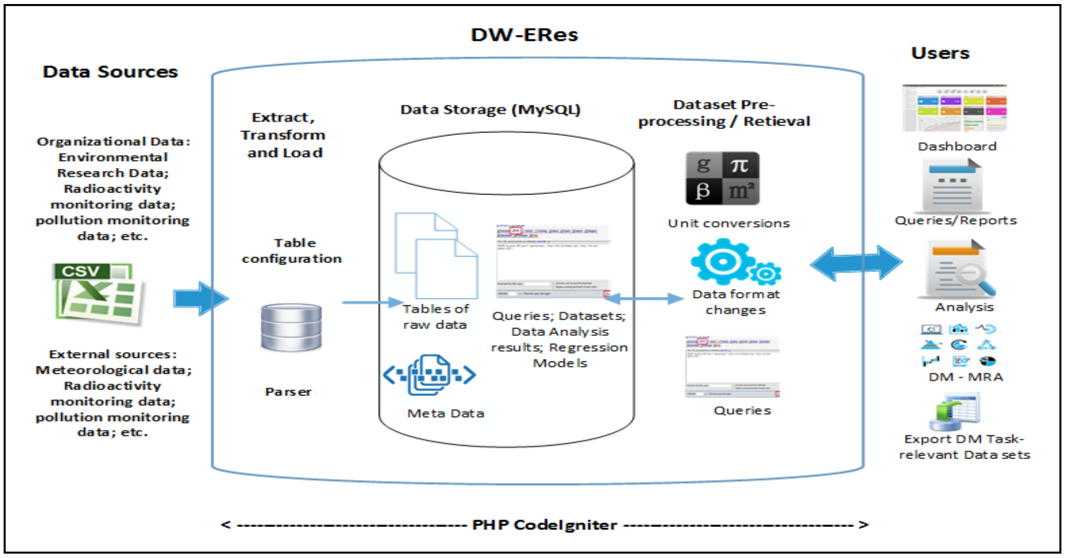
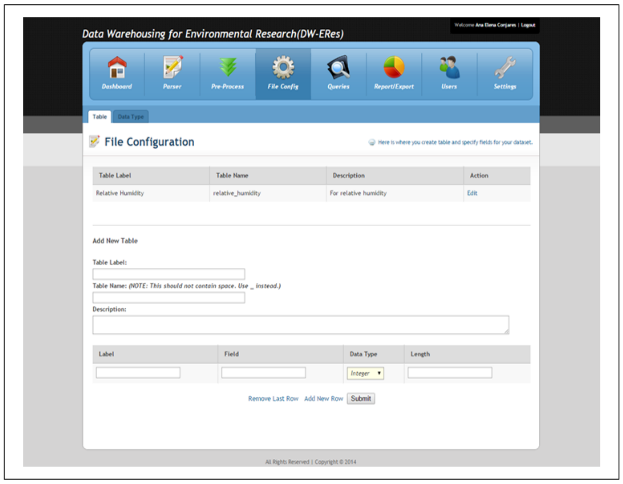

DW-ERes
Data Sources
Organizational Data: Environmental Research Data; Radioactivity monitoring data; pollution monitoring data; etc.
CSV
External sources: Meteorological data; Radioactivity monitoring data; pollution monitoring data; etc.
Extract, Transform and Load
Table configuration
Parser
Data Storage (MySQL)
Tables of raw data
Queries; Datasets; Data Analysis results; Regression Models
Meta Data
Dataset Pre-processing / Retieval
Unit conversions
Data format changes
Queries
Users
Dashboard
Queries/Reports
Analysis
DM - MRA
Export DM Task-relevant Data sets
< ------------------------------- PHP CodeIgniter ------------------------------- >

Fig. 3 Architecture of the Data Management System for Environmental Research

3.3 File Configuration Interface

The File Configuration Interface provides for the functionality to create tables corresponding to datasets that will be uploaded into the system. This is a powerful feature of the system as it provides for flexibility in defining new set of environmental data when it is not yet featured in the database. It also provides for editing of the parameters of an existing table, when necessary, but only those that have not been populated with data. Existing Tables in the database are displayed in the system with descriptions that provides information to researchers what tables of environmental data are already existing in the system. Modification of uploaded data is allowed for owners of the data. Figure 4 shows the screenshot of the table creation GUI.



Data Warehousing for Environmental Research(DW-ERes)

File Configuration

Here is where you create table and specify fields for your dataset.

Table Label	Table Name	Description	Action
Relative Humidity	relative_humidity	For relative humidity	Edit

Add New Table

Table Label:

Table Name: *(NOTE: This should not contain space. Use _ instead.)*

Description:

Label	Field	Data Type	Length
		Integer ▾	

Remove Last Row Add New Row Submit

All Rights Reserved | Copyright © 2014

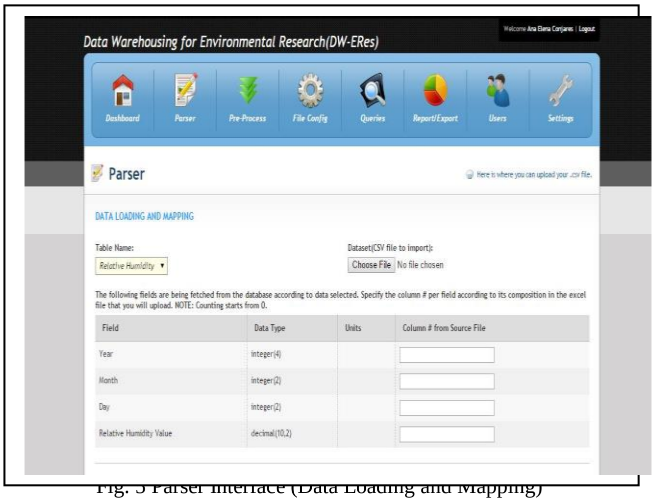
Fig. 4 Table/File Creation and Configuration Module

3.4 The DW-ERes Parser (Data Loading) Interface

This is the stage where DW-ERes acts as the central repository of all environmental data derived from scientific research projects. The parser subprocess includes the data loading and mapping interface as shown in Figure 5 through which datasets from the external data source are loaded into the DW-ERes easily by authorized user (researcher). The user configures the table according to the data source.

The parser module as shown in Figure 6 also includes the attachment of metadata as determined by the user. The metadata will be used as search strings for viewing generated datasets. The metadata will be used as search strings for viewing generated datasets.

https://doi.org/10.20943/01201605.171177


Data Warehousing for Environmental Research(DW-ERes)
Welcome Ana Elena Conjares | Logout
Dashboard
Parser
Pre-Process
File Config
Queries
Report/Export
Users
Settings

Parser

Here is where you can upload your .csv file.

DATA LOADING AND MAPPING

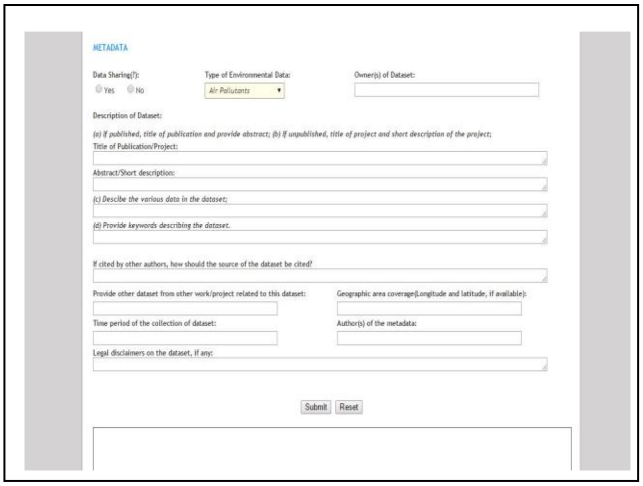
Table Name: Relative Humidity ▾

Dataset(CSV file to import): Choose File No file chosen

The following fields are being fetched from the database according to data selected. Specify the column # per field according to its composition in the excel file that you will upload. NOTE: Counting starts from 0.

Field	Data Type	Units	Column # from Source File
Year	integer(4)		
Month	integer(2)		
Day	integer(2)		
Relative Humidity Value	decimal(10,2)		

Fig. 5 Parser Interface (Data Loading and Mapping)



METADATA

Data Sharing(?): ○ Yes ○ No

Type of Environmental Data: Air Pollutants ▾

Owner(s) of Dataset:

Description of Dataset:

(a) If published, title of publication and provide abstract; (b) If unpublished, title of project and short description of the project;

Title of Publication/Project:

Abstract/Short description:

(c) Descibe the various data in the dataset;

(d) Provide keywords describing the dataset.

If cited by other authors, how should the source of the dataset be cited?

Provide other dataset from other work/project related to this dataset:

Geographic area coverage(Longitude and latitude, if available):

Time period of the collection of dataset:

Author(s) of the metadata:

Legal disclaimers on the dataset, if any:

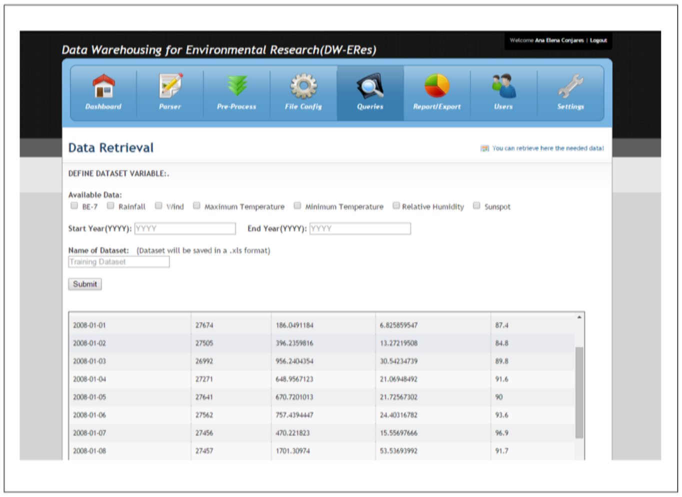
Submit Reset

Fig. 6 Parser Interface (Metadata)

3.5 The DW-ERes Dataset Retrieval Interface

Figure 7 presents the system's dataset retrieval interface. This module allows the user to create any dataset by defining the environmental data to be retrieved from the different tables in the DW-ERes, common reference is the sampling time of the different data.

Datasets are saved in comma separated values file formats or simply called as CSV files which is compatible with existing data analysis software such as the WEKA machine learning software. Once the dataset is created, it can now be accessed for use in exploration and data mining. For this initial study, environmental data for ^{7}Beryllium, rainfall, wind, maximum temperature, minimum temperature, relative humidity and sunspot have been uploaded into the DB to test the prototype developed.


Data Warehousing for Environmental Research(DW-ERes)
Welcome Ana Elena Conjares | Logout
Dashboard
Parser
Pre-Process
File Config
Queries
Report/Export
Users
Settings

Data Retrieval

You can retrieve here the needed data!

DEFINE DATASET VARIABLE:.

Available Data:
□ BE-7 □ Rainfall □ Wind □ Maximum Temperature □ Minimum Temperature □ Relative Humidity □ Sunspot

Start Year(YYYY): YYYY End Year(YYYY): YYYY

Name of Dataset: (Dataset will be saved in a .xls format)
Training Dataset

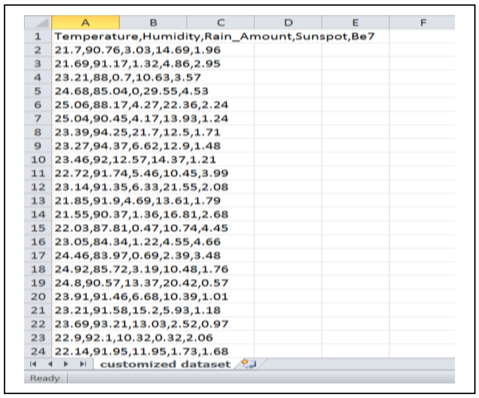
Submit

2008-01-01	27674	186.0491184	6.825859547	87.4
2008-01-02	27505	396.2359816	13.27219508	84.8
2008-01-03	26992	956.2404354	30.54234739	89.8
2008-01-04	27271	648.9567123	21.06948492	91.6
2008-01-05	27641	670.7201013	21.72567302	90
2008-01-06	27562	757.4394447	24.40316782	93.6
2008-01-07	27456	470.221823	15.55697666	96.9
2008-01-08	27457	1701.30974	53.53693992	91.7

Fig. 7 Dataset Retrieval Module

3. Simulation Results

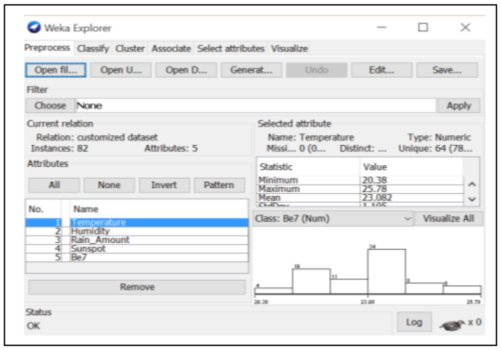
Figure 8 presents the snapshot of the customized dataset retrieved from DW-ERes for simulation purposes. The dataset retrieved contained 82 instances which were retrieved in .05 seconds. It was observed that the retrieval of dataset was fast enough considering the number of significant instances involved. It is significant to take note of the system's speed as Guerra and Andrews stated that data warehouses are built in high speed data entry design and high speed retrieval [10]. This dataset is generated in .csv format that could easily be accessed by any data analysis software specifically on regression.



```
1  Temperature,Humidity,Rain_Amount,Sunspot,Be7
2  21.7,90.76,3.03,14.69,1.96
3  21.69,91.17,1.32,4.86,2.95
4  23.21,88,0.7,10.63,3.57
5  24.68,85.04,0,29.55,4.53
6  25.06,88.17,4.27,22.36,2.24
7  25.04,90.45,4.17,13.93,1.24
8  23.39,94.25,21.7,12.5,1.71
9  23.27,94.37,6.62,12.9,1.48
10 23.46,92,12.57,14.37,1.21
11 22.72,91.74,5.46,10.45,3.99
12 23.14,91.35,6.33,21.55,2.08
13 21.85,91.9,4.69,13.61,1.79
14 21.55,90.37,1.36,16.81,2.68
15 22.03,87.81,0.47,10.74,4.45
16 23.05,84.34,1.22,4.55,4.66
17 24.46,83.97,0.69,2.39,3.48
18 24.92,85.72,3.19,10.48,1.76
19 24.8,90.57,13.37,20.42,0.57
20 23.91,91.46,6.68,10.39,1.01
21 23.21,91.58,15.2,5.93,1.18
22 23.69,93.21,13.03,2.52,0.97
23 22.9,92.1,10.32,0.32,2.06
24 22.14,91.95,11.95,1.73,1.68
```

Figure 8. Customized Dataset

The customized dataset was then tested by loading it into Weka, an open-source software for solving real-world data mining problems. Figure 9 shows that the dataset was successfully imported into Weka which implies that the resulting customized dataset of DW-ERes is well structured.



Filter: None

Current relation
Relation: customized dataset
Instances: 82 Attributes: 5

Selected attribute
Name: Temperature Type: Numeric
Missi... 0 (0... Distinct: ... Unique: 64 (78...

Statistic	Value
Minimum	20.38
Maximum	25.78
Mean	23.082
StdDev	1.105

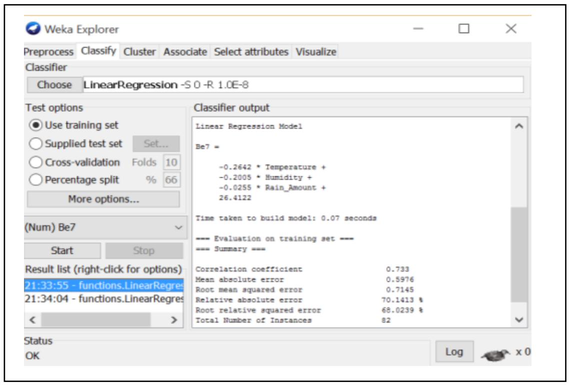
No.	Name
1	Temperature
2	Humidity
3	Rain_Amount
4	Sunspot
5	Be7

Class: Be7 (Num)

Status
OK

Figure 9. Dataset Importation to Weka

Figure 10 shows the regression model successfully generated by Weka after the application of linear regression algorithm.



Classifier: LinearRegression -S 0 -R 1.0E-8

Test options
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Num) Be7

Result list (right-click for options)
21:33:55 - functions.LinearRegres
21:34:04 - functions.LinearRegres

Classifier output

```
Linear Regression Model

Be7 =

     -0.2642 * Temperature +
     -0.2005 * Humidity +
     -0.0255 * Rain_Amount +
     26.4122

Time taken to build model: 0.07 seconds

=== Evaluation on training set ===
=== Summary ===

Correlation coefficient                  0.733
Mean absolute error                      0.5976
Root mean squared error                  0.7145
Relative absolute error                 70.1413 %
Root relative squared error             68.0239 %
Total Number of Instances               82
```

Status
OK

Figure 10. Generation of Regression Model using Weka

4. Conclusions and Future Works

As an initial implementation of a data warehousing tool, a data management platform for environmental data called Data Warehouse for Environmental Research (DW-ERes) has been designed. PHP CodeIgnitor hosted in a Linux operating system with Apache as its web server and MySQL as the database server was utilized successfully. Compared to other data management systems and existing methods, this approach considered the loading of data from external sources and attachment of metadata, and data transformation to generate customized datasets that are compatible with data analysis software such as the WEKA machine learning software. DW-ERes was found to be a useful storage space as it has a stable data model for data consolidation from external sources and has high speed dataset retrieval that will support reporting, analysis, and knowledge discovery.

Future work involves extending the capability of the system to act as data analysis and predictive modeling platform which integrates a data mining task, specifically regression, to assist in the dissemination of new knowledge and decision making in this field. Automated selection of significant variables is also an additional concern of the authors by integrating a feature selection algorithm into the system.

References

[1] Y-F. Li, G. Kennedy, F. Ngoran, P. Wu, J. Hunter, “An ontology-centric architecture for extensible scientific data management systems,” in Future Generation Computer Systems, vol. 29, pp. 641-653, 2013.

[2] D. Fleischer, M. Bolter, R. Moller, “The implementation of initial data populations of environmental data and creation of a primary working database,” in Polar Science, vol. 6, pp. 97-103, 2012.

[3] W.H. Inmon, Md A Gettinger and K. Krishnan, Appendix B - Building the Healthcare Information Factory: Healthcare Information Factory: Implementing Textual Analysis. Data Warehousing in the Age of Big Data, 2013, pp. 289-332.

[4] D. S. Sayogo, T. A. Pardo, “Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data,” in Government Information Quarterly, vol. 30, pp. 519–531, 2013.

[5] K. Gaustad, T. Shippert, B. Ermold, S. Beus, J. Daily, A. Borsholm, K. Fox, “A scientific data processing framework for time series NetCDF data,” Enivronmental Modelling & Software, vol. 60, pp. 241-249, 2014.

[6] R. Lawrence, A. Kruger. “An architecture for real-time warehousing of scientific data,” [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.1447&rep=rep1&type=pdf

[7] M. Diepenbrock et al., “PANGEA – an information system for environmental sciences,” in Computers and Geosciences, vol 28, pp. 1201-1210, 2002.

[8] Xuezhong Zhou, et al., ‘Development of traditional Chinese medicine clinical data warehouse for medical discovery and decision support,” Artificial Intelligence in Medicine, vol. 48, pp. 139-152, 2010.

[9] S. Bainbridge, “A services based architecture for acheiving interoperability of environmental observational data” in Proceedings of the Environmental Information Management Conference 2011, pp. 15-20.

[10] J. Guerra, D. Andrews, “Why You Need a Data Warehouse”, [Online]. Available:http://datalyticstechnologies.com/wp-content/uploads/2014/01/2013-03-Why-You-Need-a-Data-Warehouse.pdf

About Authors:

Ana Elena L. Conjares finished her BS in Mathematics in 1982, Cum Laude, from the University of Nueva Caceres in Naga City, Philippines. She obtained her Master of Science in Computer Science degree from the Ateneo de Manila University, Loyola Heights in Quezon City in 2000 as a government scholar of the Department of Science and Technology (DOST) Human Resource Development Program. She is now currently pursuing her degree in Doctor of Information Technology at the Technological Institute of the Philippines, Quezon City campus.

Ms Conjares works at the Philippine Nuclear Research Institute for the past three decades and currently heads the Technology Diffusion Division of the Institute. She is also designated as Chief Information Officer and Information Systems Strategic Planner. She had twice for one year each worked at the Comprehensive Nuclear Test-Ban Treaty Organization based in Vienna, Austria as project management team for their Science and Technology 2009 and 2011 events, respectively. She is also recognized as an expert of the International Atomic Energy Agency’s (IAEA) Regulatory Authority Information System which is deployed to Member States, having performed expert missions in Laos PDR, Myanmar and Brunei Darussalam as well as hosted and supervised on-the-job trainee from Member State in the Philippines.

Bobby D. Gerardo finished his BS in Electrical Engineering in 1994, with High Distinction from Western Institute of Technology at Iloilo, Philippines. He took his Master of Arts in Education Major in Mathematics from University of the Philippines in Diliman Quezon City in 2000 being the grantee of DOST-SEI scholarship for Math and Science Faculty. He pursued his Ph.D. in Information and Telecommunications with major in distributed systems at Kunsan National University, Korea in 2007 being the grantee of Korean Scholarship for Brain Korea (BK-21) project.

He is currently the Vice President for Administration and Finance of West Visayas State University, Iloilo City, Philippines. His dissertation: Discovering Driving Patterns using Rule-based intelligent Data Mining Agent (RiDAMA) in Distributed Insurance Telematic Systems. He has published 54 research papers in national and international journals and conferences. He is a referee to international conferences and journal publications such as in IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Knowledge and Data Engineering. He is interested in the following research fields: distributed systems, telematics systems, CORBA, data mining, web services, ubiquitos computing and mobile communications.

Dr. Gerardo is a recipient of CHED Republica Award in Natural Science Category (ICT field) in 2010. His paper entitled SMS-based Automatic Billing System of Household Power Consumption based on Active Experts Messaging was awarded Best Paper on December 2011in Jeju, Korea. Another Best Paper award for his paper Intelligent Decision Support using Rule-based Agent for Distributed Telematics Systems"
Asia Pacific International Conference on Information Science and Technology on December 18, 2008. An Excellent Paper award was given for his paper "Principal Component Analysis Mechanism for Association Rule Mining", Korean Society of Internet Information's (KSII) 2004 Autumn Conference on November 5, 2004. He was given a University Researcher Award by West Visayas State University in 2005.

Ruji P. Medina is Dean of the Graduate Programs and concurrent Chair of the Environmental and Sanitary Engineering Program of the Technological Institute of the Philippines in Quezon City. He holds a Ph.D. in Environmental Engineering from the University of the Philippines with sandwich program at the University of Houston, Texas where he worked on the synthesis of nanocomposite materials. He finished his MS in Environmental Engineering from the Mapúa Institute of Technology, graduating *Summa Cum Laude*. He obtained his Bachelor's degree in Chemical Engineering from the University of the Philippines in Diliman, Quezon City. His research interests include urban mining, electronic wastes, and nanomaterials, He counts among his expertise environmental modeling and mathematical modeling using multivariate analysis.