Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 1891–1900

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

# Development of a Total Environment Data Science Approach in a Big Data Scale

Shastri L Nimmagadda*, Torsten Reiners and Amit Rudra

*School of Information Systems, Curtin Business School, Curtin University, Perth, WA, Australia*

**Abstract**

We use the Big Data paradigm, as a driving mechanism of an integrated research framework. As a case study, we consider analysing various ecological systems and their connectivity in the framework. An unknown coexistence among different species and lack of knowledge on their sustainability motivate us for undertaking the current research. For describing the recycling systems in nature, an articulated design science research (DSR) framework is necessary for which we have constructed data models for composite lithosphere-atmosphere-biosphere-hydrosphere ecosystem (LABHE). The unstructured big-size environmental data sources and their anomalies existing in nature are taken advantage of, to describe various constructs, compute models and validate them by DSR guidelines. For this purpose, the domain ontology artefacts are drawn and integrated into a warehouse approach to compute an environmental metadata and interpret it in different knowledge domains. The data models and the proposed integrated framework facilitate the environment explorers for planning and management of environmental resources worldwide. The Big Data associated LABHE bring out new knowledge and its interpretation in a variety of environmental data map and plot views. The constructs, models and methodologies used in the current domain application are research deliverables for the total environment researchers and explorers.



*Keywords:* Big data; Design Science; Information Systems; Environment Ecosystem;

## 1. Introduction

At present, the conventional data management procedures have issues in documenting and organizing the heterogeneous data sources[14] of ecological systems. In recent years, the Big Data paradigm has taken a new direction for handling and managing complex data types.

* Corresponding author. *E-mail address:* shastri.nimmagadda@curtin.edu.au

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.237

The environmental data sources that are multifaceted in nature are voluminous, complex and dynamic (LABHE). They are characterized by volumes, variety, velocity, visualization, veracity and value dimensions[17]. The data anomalies[16] of multiple ecosystems exhibit varying ranges, scales, types, varieties[9] with periodic and geographic dimensions and their attribute anomalies. Also, the total environment embraces with human elements, description of cultivated landscapes and built environments. The Big Data characteristics motivate us in proposing an environment research framework, with new construct and model articulations that are validated by design science guidelines[8, 18]. The framework explores the links between living and non-living elements that are associated with atmosphere, lithosphere, biosphere and oceano-sphere ecosystems. It is a new drive in managing data relationships and their connections among ecosystems' elements and processes.

Interrelationships, scales and changes that occur and attribute to our surroundings of existence constitute an environment[9]. We illuminate the environment as a set or group of entities, objects and dimensions of these entities. The environment that represents our surroundings is interpreted to have linkages or interrelationships with connectivity, with conceptualized attributes of different scales and magnitudes. It is also termed as nature, which is a super-dimension, described with common effects of atmosphere, lithosphere, biosphere and oceano-sphere named as sub-type dimensions. Identifying relationships among these entities and dimensions (LABHE) is challenging task. Neither entities nor objects and dimensions exist or act in their environments in complete isolation. Each is affected by adjacent entities, objects or dimensions and each of them is influenced by the other. The concepts inspire us to acquire new facts and their varying instances that have different scales ranging from the microscopic scale of a molecule to a macro scale, like the whole Earth that has impact and influence on the ecosystems' relationships. The changes are dynamic in the environment. If the changes are stable, the environment is said to be stable, but if the change surpasses the environment in which the change is accommodated, a disruption is caused.

We use design science approach and guidelines[10, 11] for articulating the data constructs and models of ecosystems, in which multiple dimensions attribute to the characteristics of the Big Data and integrate to build a metadata. This describes data comprising of various recycling systems. The framework described in[8] motivates us examining the existing artefacts, exploring new artefacts contemplating the digital ecosystems' integration and connectivity. In the current study, issues and challenges of the environment management, objectives of the current study, methodologies, results and discussions of a case study are discussed in support of the implementation of the articulations in the environment domain.

## 2. Problem statement and issues

The connectivity among the elements of the environment-ecology systems is either poorly understood or mismanaged[3]. Managing the volumes and varieties of data and integrating their linked elements and processes is a major challenge, where the semantic, schematic and system inconsistencies[13, 14] exist within the data sources. Besides, the heterogeneity and multidimensionality of the data are other asserting concerns. Lack of knowledge on the ecosystems and their boundaries have held up the major research in this direction. We describe the problem in a broader context, such as LABHE, their inherent relationships, and how they impact in totality the environment. Air and water pollution, the disappearance of forest lands, soil erosion, and deterioration of urban areas, population growth, resources depletion, energy crisis and nuclear power are issues, knowingly cannot be disregarded. All these dimensions and facts instances are considered in the schema modelling.

While characterizing the ecological systems, the Big Data features may exhibit ambiguities[17] in their current representations. The connectivity between the elements of multidimensional environment-ecological systems and their participation in balancing the environment are loosely interpreted[9]. Air and water pollution, the disappearance of forest lands, soil erosion, $CO_2$ emissions, deterioration of urban areas and population growth are causative to depletion of existing resources, affecting the sustainability and endangering of the species. Documentation and organization of these data sources associated with various elements of lithosphere, atmosphere, biosphere and hydrosphere are challenging, and domain expertise is needed to explore the connections among these systems. We take advantage of DSR framework to integrate and interconnect the elements and their dimensions as described in the following sections.

## 3. Design science information system (DSIS) and research objectives

Having understood the problem statement, we frame the research objectives and research methodologies:

1. Big Data relevance in the design and development of a "total environment" data science.
2. Description of domain ontologies and Big Data articulations in the LABHE.
3. Incorporating domain ontologies and their integration within an integrated framework, envisaging and deriving the total environmental metadata through a design science research framework.
4. Interpretation of ecosystems' boundaries, which are otherwise ambiguously perceived.
5. Implementation of warehoused metadata from the LABHE and establishing the connectivity among ecosystems of the total environment.

## 4. Motivation of Big Data in the DSR framework

The data analytics that needed the data capture, curation, sharing, transfer, and visualization including information privacy procedures, all accommodating in a single framework is a huge challenge, unlike the traditional database management applications. The Big Data architecture and characteristics[4, 17] describe massive volumes of structured and unstructured data of the LABHE, which are so largely difficult to process and manage by traditional database applications. Identifying volumes of structured and unstructured data sources associated with lithosphere, atmosphere, biosphere and hydrosphere entities and their dimensions is challenging and an enormous task. The total environment embeds with multiple ecosystems and the data sources associated with these ecosystems are characteristically heterogeneous and multidimensional. Having accumulated multidimensional data sources in the diverse embedded ecosystems is a motivation; documenting and organizing them within a robust warehouse data structure are critical using Big Data paradigm. We investigate the role of Big Data characteristics in modelling various constructs and methods that accommodate the DSR research framework. The guidelines[6] that validate the DSR framework (Fig.1) and evaluate in the ecosystems' arena are described. Various evaluation properties[18] are interpreted for validating the DSR approach in the current domain application. Implementation of the articulations in uncovering the new knowledge on the environment, visualization and interpretation are the focus of current research and motivation.

## 5. Research methodologies

We are of the view that environmental issues are relevant to Lithosphere, Biosphere, Atmosphere and Hydrosphere, which are modelled in a domain modelling. As described in the research framework8, several research activities are undertaken. Selection of a domain, schema, data integration, data view imaging and interpretation[14, 16] are different artefacts of the integrated research framework. As shown in Fig. 2, Big Data characteristics are additional dimensions considered in articulating the data models, equipping the DSR framework. Connectivity is explored among a sequence of events occurring during construct design and domain ontologies integration. Design science articulates methodological tools that natural scientists use[9] in building models. Research methodologists suggest valid ways for collecting and analysing the facts that support (or negate) a posted theory[2, 8], [10]. They are human-created artefacts that have value, address this task. Dimensions of Big Data characteristics are mapped through various artefacts of ecosystems, as demonstrated in the following sections. For this purpose, dimension and numerical fact instances are acquired and documented in various database structures to build a metadata. The instances are from large volumes of environmental data sources. The following research activities[10, 17, 18] are used for investigating and mapping the environmental data sources.

### *5.1 Modelling and research activities*

*Artefact construction:* Artefacts[4] are constructed to accomplish a particular task. Is the artefact going to work? Building an artefact validates the viability. The artefacts then become the purpose of incorporating them into the research framework. The constructs, models, methods and instantiations[10] are research outcomes expected besides motivating the research framework. Each component of the framework is a technology, built and evaluated scientifically in the integration process (see Fig.2).

*Artefact assessment*: Artefacts are evaluated to determine if any progress is made. How best an artefact is evaluated, and an evaluated artefact can work? Evoke the progress attained when a tool or concept swaps with a more efficient artefact. Assessment involves making up of quantifiable measures and the dimensions of artefacts according to those metrics[10]. The quantifiable measurement is an accomplishment to evaluate the functioning of an artefact. In our case, the models deduced to logically represent the schemas, surrounding the problem situation, are evaluated.

*Artefact theorization*: For a chosen artefact, the performance is evaluated, signifying to determine why and how the artefact functioned or did not perform within its ecosystem. The IT artefacts need the support and basic foundation of natural sciences. We hypothesize and substantiate the theories behind the artefacts. Besides, the conceptualizations explain the features of the artefacts and their collaboration with the ecosystem that affected the discerned performance. Both conceptualization and contextualization need to accept the natural laws governing the artefacts as well as the environment in which they get operated. Theorizing the internal working performance of an artefact with which the environment responds, depends on the connectivity between the artefact and environmental data science.

*Artefact justification:* Once the artefacts are hypothesized, the semantic-based constructs are envisaged more with an end user's ability to articulate the databases than do the relational or multidimensional data models. The description of the generality of the constructs and models is vindicated. The reasoning is gathered to test the concepts behind constructs and models. In the current context, the environmental data sources from multiple domains are probed to examine such justifiable constructs and models.

The research outcomes are validated[13, 6] using the design science guidelines:

*Design, as an artefact:* Artefacts, in the present context, are innovations that narrate data structures, used in domain ontologies and their integration in the current context. Heterogeneous and multidimensional data sources identified from ecological systems are made good use of designing the data structures. The ER, OO and dimensional models are various constructs considered for different ecosystems' scenarios, with several assumptions and limitations. These artefacts have relevance to the current research problem and domain application.

*Problem significance:* An intended artefact ensures the application and relevance, demarcating amid the methodology and integrated framework, a research solution. Poorly organized total environment data sources is a major issue. Interpretation of connectivity among various ecological systems is challenging with a more generic problem relevance. Another issue is methodologies that are withstanding to alterations among ecosystems and their existence, accessing the domain knowledge among multi-users' environment and ecosystems services. So, artefacts are aimed at and focused on research problems.

*Design evaluation:* The artefacts and their alliance need to go through the integration process. The design evaluation criteria refer to the successful integration of domain ontologies or data structures. An added criterion is that knowledge-based domain ontologies are integrated into a warehouse environment. Based on a variety of business rules, ecosystem scenarios, and other constraints, the artefacts are evaluated. The functionality, completeness, consistency, accuracy, performance, reliability, usability are utility properties[10, 23] of the IT design evaluations, obeying the inclusiveness of the ecosystem and its existence. All the dimensions and their fact instances involved in the multidimensional data structuring process are evaluated concerning their consistency, accuracy, reliability and usability in the framework.

*Research contribution:* The research highlights the ongoing artefact design implementation and the scope in other domain application areas. Digital ecosystems[3], as total environment solutions ensure the connectivity among the chosen artefacts. Domain ontologies ensure the viability in different knowledge-based research problem solutions. The creative development and the evaluation of artefacts support such research scopes.

*Research Rigour:* Rigour focusses on the methodological view and its direction. The design and evaluation of constructs and models need rigour based DSR guidelines. In particular, on the construction activity, the precision is assessed on its applicability and generalizability of the artefact in a balanced way. How best an artefact performs and fits into the framework, not theorizing or proving anything about why and how the artefact works and in other ecosystem sceneries. It is vital to cognize why an artefact performs or does not perform, enabling to look for alternate artefacts and exploiting the former and avoid the latter[10, 11]. It may also be domain specific in which we put a rigour on application design research.

*Design as a pursuit:* An effective problem solution ascertains the constructs, models and their design process. The problem solution is viewed as developing the existing designs to reach the desired results, sustaining the rules of the ecosystem and its associated data science[18] of the total environment. Generalization and demonstration of an appropriate mean of the end users' laws and requirements are vital components of DSR approach.

*Communication of research:* Both the data and design science design ensure and offer solutions to technology and management personnel. IT specialists need enough specifics to enable describe, use and reuse of the artefacts and implement them within either an ecosystem scenario or an appropriate organizational context or problem-solving situation. The researchers take advantage of the values presented by the artefacts, enabling the ecologists to make new facts for further evaluation. It is important addressing such researchers, apprehending the practices based on which the artefact was created and assessed. It ascertains testing and validating the research outcomes, further building a new knowledge-based solution for total environment data science driven DSIS. The highlight is the significance of the problem, the uniqueness and effectiveness of the research solution, appreciated for the artefact. The guidelines motivate us to frame and redesign the research objectives and approaches in several knowledge realms. The constructs, models and methods are sets of artefacts ensuring the semantics and schematics of the data in multiple domain applications[7, 17].

### *5.2 Ontological constructs and logical schema modelling*

The chemical elements such as carbon, nitrogen, sulphur are recycled continuously, as is water. Plants and animals are also recycled in a way, in as much as they can perpetuate their species through their reproduction. Unfortunately, these cycles are disrupted by human-made activities, which lead to a variety of environmental problems[9, 15]. Similar is the case with fossil fuels[1]. The distraction of the carbon cycle, for example, may be associated with global warming [9, 12] and the sulphur cycle may be related to acid rain. Pressure on the hydrologic cycle causes a shortage of water with serious water pollution. Ontological descriptions are drawn for a natural cycle system using chemical elements such as carbon, nitrogen, sulphur attributes, in the form of conceptual and logical schemas, as shown in Figs. 3a and 3b.

### *5.3 Environmental dimensions affecting the human ecosystem*

*Toxins and Pollutants:* The chemicals found in the environment can enter the body and cause metabolic change. A toxin is a chemical produced by a living organism that has a harmful effect on other living beings. A pollutant, on the other hand, is a by-product of human activity that can harm living beings. Xenobiotic or Toxins and pollutants collectively affect the environment and also living human being. A foreign chemical substance termed as xenobiotic is found within an organism, not naturally produced or expected within the organism. At times, they may present with high concentrations. The human body does not produce the antibiotics themselves as xenobiotic processes, nor part of the normal diet. The elements, processes, chains of connections and their sequential events among chemicals existing in the environment are used in conceptualizing constructs and building multidimensional models. The facts and attribute dimensions interfere natural barriers, controlling the xenobiotic events.

*Ingestion* dimension (Ing Dim): *Name of the tract, period of occurrence* and *absorption effect* are attributed dimensions of the table. The demographic and geographic data of all ingestions are tabulated. Paints containing lead and or mercury if consumed. Chemicals and types of chemicals used in industrial applications and domestic use, their

attribute dimensions and instances are documented. If xenobiotic is directly absorbed in the mouth that can happen through solvents and enter through body circulation and organs, such as the liver. Ingestions through stomach or intestine are a better chance of being metabolized by the liver into a harmless form. These dimensions and facts and their instances are used to make connectivity with other dimensions, such as inhalation, dermal and injection as described here.
*Inhalation Dimension:* (Inh Dim): *Name of Gas*, *Period of Occurrence*, *Human Organ Affect* are attributed dimensions of the table. The classified information of all gases with laboratory experiments are tabulated. The data integration process of a warehouse creates internal identifiers of data that characterize inhalations. The specimen test data dimensions and fact tables are incorporated. *The name of gas, a period of occurrence* and *human organ affect*, are primary keys as dimensions interconnecting patient dimensions and facts tables in the data warehouse environment. In a pragmatic approach, numerous primary key dimensions of associated attributes and fact tables of a healthcare system are linked inhalation dimension tables. The semantics of "inhaled gas name", units of "period of occurrence" and "human organ" affected by inhaled gas are also included. Queries made for data views determine the norm of different primary key dimensions and their instances that associated with the inhaled gas attribute dimensions. Which records are physically kept or rejected in the databases of a warehouse, in consultation with the warehouse designers and the ecosystem needs and services? Inhalation affects the respiratory tract. Aerosols, gases and vapours are examples of this dimension. Asbestos is swept away from the lungs by the cilia in the trachea or manage to reach the bronchioles and alveoli, where they are engulfed by microphages and carried to the lymphatic system. Similar to inhalation dimension, other dimensions and facts are documented.

*Dermal Dimension:* Similar to inhalation, a data structure is designed using dermal dimension and its associated facts. Toxins found in poison ivy can destroy the surface layers of skin. All associated facts and instances are documented for constructing the data model with dermal dimensions and facts.

*Injection Dimension:* Injection done through the skin is another controlling factor for xenobiotic. Stings and bites occur by insects, snakes, spiders, and other animals and plants affect the cell membranes. Foreign chemicals that get dissolved in water and pass through pores in the membrane via diffusion, dimensions and facts are identified and documented for modelling and warehousing purposes.

Other dimensions and factual data influencing the effect on any xenobiotic are described in the following sections.

- *Type Dimension (of Compound):* Xenobiotic come in different forms such as solid, liquid and gases; some come in different chemical compounds such as liquid mercury or mercury salts. Different types have different effects.
- *Size Dimension (of Dose):* This dimension is associated with a measured unit, which is appropriate to the type of compound or suspected concentrations (such as ppm, ppb, or g/L). Gases are measured in mg per cubic meter of air.
- *Method of Entry into the Body:* This dimension is an affect how quickly a xenobiotic is stored or metabolized.
- *Duration Dimension (of Exposure):* Long term exposure leads to an increase of damage, which can be detectable. Acute exposure may cause rapid fatality or sickness.
- *Synergistic Factors:* Synergism that refers multiple factors affecting the body organs, can create a risk. For example, workers working in asbestos industries, have smoke habits have more risk of lung cancer, compared with non-smokers.
- *Host Factors:* Age, health factors, hereditary issues, fitness, immunities are other dimensions affecting the xenobiotic.

Table 1: Pollutant dimensions, connectivity and sources

| Pollutant Data Dimensions and Facts | Connectivity to super type dimensions | Data Sources |
|---|---|---|
| ***Ozone Dimension:*** A pale gas, photochemical smog of the Earth's surface. protects from harmful rays of the Sun in the upper atmosphere | Lithosphere Atmosphere<br>Smog is combination of gases and particulate matter | Data on automobiles, plants, landfills, industrialised solvents, gas stations, farms and lawns |
| ***Carbon Monoxide Dimension***: Exhaust of motor engines emanate unscented and colourless gas, where there is incomplete fossil fuel combustion. | Biosphere<br>Lithosphere Atmosphere | Data about automobile and transport industries; high concentrations, confinement areas including heavy traffic. |
| ***Nitrogen Dioxide Dimension:*** The data related with a light brownish gas at lesser concentrations; unpleasant, urban haze in higher concentrations. | NO and $NO_2$ lead to pulmonary damage.<br>Biosphere<br>Lithosphere Atmosphere | Data about automobile and chemical Industries |
| ***Particulate Matter Dimension:*** a solid or liquid condensations from smoke, dust, fly ash and condensing vapours, held in the air for longer periods | Biosphere<br>Lithosphere Atmosphere<br>Hydrosphere | Manufacturing/producing industry sources, diffuse sources, natural sources, transport data sources |
| ***Sulphur Dioxide Dimension:*** Colourless gas, odourless at low concentrations, but pungent at high concentrations | Lithosphere Atmosphere | Industrial and petro-chemical utilities, smelters and paper mills |
| ***Lead Dimension:*** Lead and lead compounds harm human health through either ingestion of lead polluted soil, dust, paint or direct inhalation | Lithosphere Atmosphere | Automobile exhaust gases, in the soil as pest control element and paint and dye industries |
| ***Toxic Air Pollutant Dimension:*** Arsenic, asbestos, and benzenes | Lithosphere Atmosphere | Chemical plants, industrial processes, motor vehicle emissions, fuels and building materials |
| ***Stratospheric Ozone Deplete Dimension:*** Rise of chlorofluoro carbons, halons, carbon tetrachloride, and methyl chloroform to atmosphere may destroy the ozone layer | Lithosphere Atmosphere | Sources may be from the reaction of sunlight with chemicals found in automobile exhaust and gasoline vapours, household utilities |
| ***Greenhouse Gases Dimension:*** Gases accumulated in the atmosphere are causative to induced global climate change. The gases include carbon dioxide, methane, and nitrous oxide. | Lithosphere<br>Atmosphere Hydrosphere | The combustion for energy use and transportation. Methane from landfills, oil & gas seeps, cud-chewing livestock, coal mines and rice paddies. |

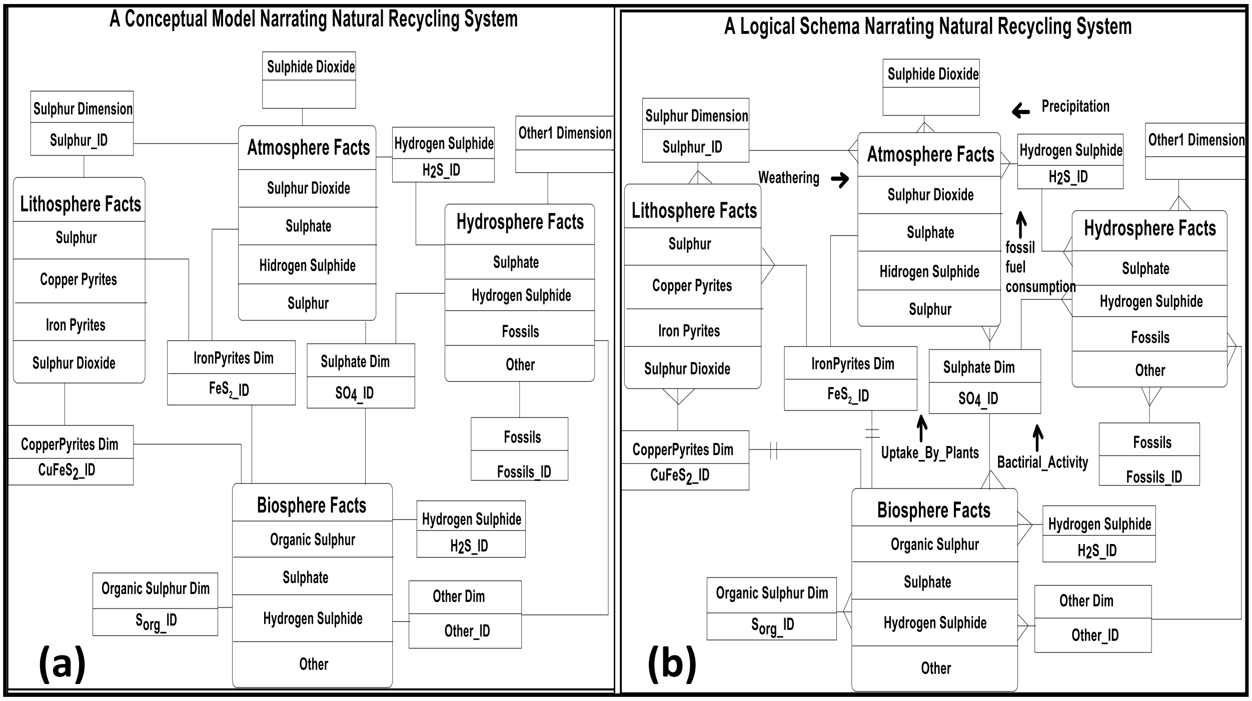
Table 2: Pollutant, dosage and human organ-affect dimensions and facts

| Pollutant Dimension | Dosage Dimension, ppm | Human organ-affect dimension |
|---|---|---|
| Ozone, $O_3$ | 0.1 and 0.08; 0.04 is tolerable | Skin Cancer, pulmonary tissues |
| CO, Carbon Monoxide | More than 100 | Blood vessels, haemoglobin |
| $NO_2$, Nitrogen Dioxide | 1.6 – 5 per hour;<br>25 – 100 per hour;<br>>100 per hour | Increased airway resistance<br>Acute, reversible bronchitis and pneumonitis<br>Death from pulmonary oedema |
| Particulate Matter, | 2.5 micrometre, 10 micrometre | Asthma, breathing issues, lung |
| Sulphur Dioxide | 8-10ppm per hour; >20ppm per hour | Throat and lungs |
| Lead | | Gastrointestinal tract, heart, kidneys and liver |
| Toxic Air: Arsenic | 0.01<br>0.04<br>0.28<br>0.46 | Brain<br>Blood<br>Nails<br>Hair |
| Stratospheric Ozone Deplete | <10ppmv | Skin cancers, Cataracts, Immune System, |
| Greenhouse Gases | $CO_2$: 392.6ppm; $CH_4$: 1874ppb; $N_2O$: 324ppb; Tropospheric Ozone: 34ppb | Warming of Earth, extreme weather, climate changes affecting human survival on planet |

The constructs are conceptualized models using various environmental dimensions and facts (LABHE context), with which relationships among attributed dimensions are stated specifying the connections. Models use the connections logically to represent a physical schema. The integrated framework[14] uses the processes and tools (for example Big Data characteristics) that guide and drive the workflow to apply and implement in domain applications, solving the situational and contextual problem of ecological systems with a view of "unravelling embedded inherent connectivity" among ecosystems. Similar to the construct and logical schema, the following dimensions and their attributes are used in building the conceptual models.

1. Toxin schema (attributes: metabolism, Foreign chemical substance, Living organism type, concentration, period of occurrence)
2. Pollutant schema (attributes: chemical pollutant, water pollutant, air pollutant, period of occurrence, radiation)
3. Ingestion schema (attributes: name of tract, period of occurrence, absorption effect)
4. Inhalation schema (attributes: name of gas, period of occurrence, organ affect, specimen, period of exposure)
5. Dermal schema (attributes: skin related, radiation, period of occurrence)
6. Injection schema (attributes: stings, bites, period of occurrence)
7. Type schema (attributes: solid, liquid, gas, pollutant)
8. Size schema (attributes: measurable unit, concentration, compound)
9. Entry schema (attributes: inject, doping, metabolism, period of exposure/entry)
10. Duration schema (attributes: period, period of exposure)
11. Synergy schema (attributes: health hazard, habitual hazard, work hazards, concentration)
12. Host schema (attributes: age, hereditary, fitness, immunity, units)

In an ecosystem scenario, the environment in which an object is found, consists of many associated objects or elements that surround it. In the environment ecosystem development scenario, data relationships' are significant criteria in binding ecological dimensions of the existing attributes of the multidimensional data structures[16]. The conceptual and logical schemas based on which the information system built and described as in Figs. 1a and 1b, illustrate the data relationships and their connectivity. Several data sources and their anomalies are documented for schema modelling. Various dimensions and their associated fact instances identified for conceptual and logical star schema modelling purposes are documented in various database codes. Our objective is to connect the *lithosphere*, *atmosphere* and *hydrosphere* entities in a natural recycling system, LABHE. As described in a methodological framework[10], various research activities are undertaken. The artefacts[14, 16] used in multiple domain applications are




different components of the integrated framework to reveal new knowledge from the environmental digital ecosystems (LABHE) and their associated data science.

Fig. 1: (a) A conceptual model and its description of natural recycling system (b): A logical schema drawn for the natural cycle system

As a follow-up of implementing the guidelines[18] of DSIS research, the integrated framework is made generic for manifold domain applications. The use and reuse of the DSIS are accomplished, based on the improvements in the design, implementation and analysis of practices, technical capabilities and products. The current IS research issues of the total environment data sciences (LABHE perspective) are analysed keeping in view the stock of the artefact design procedures and their evaluations[10, 5, 15]. The constructs and models represent the real world situations of the total environment data science scenarios along with the design issues and related solution spaces. The problem solutions and standardised models support the current domain applications, further exploring the influences between problems and solutions, taking advantages of both data and design sciences and their fluctuations in the real world. As illustrated in Fig. 2, the metadata solutions are provided showing the efficacy of the artefact design.

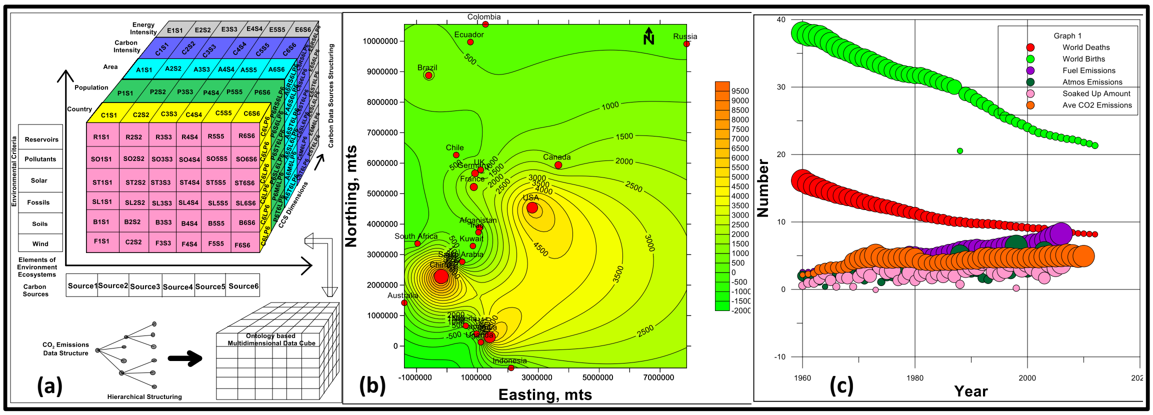



Fig. 2: Computed environmental metadata (a) cuboid data structure (b) map view (c) plot view

## 6. Results and discussions

The connectivity between ecological systems establishes the domain knowledge, providing the efficacy of the constructs, models and methods built into the framework and its implementation in the environment application domain. The map and plot views are drawn from cuboid artefacts for interpretation and new knowledge discovery, as described in Fig.2 provide valuable information and knowledge on ecological systems.

## 7. Conclusions and future scope

The research framework described in the current research has a scope of extending to problem solutions of other domains of the green information systems including the total environment system, LABHE. The domain ontologies written for complex ecosystems, such as LABHE are integrated into the framework. The models and integrated framework described in the current domain application facilitate the environment explorers with systematized solutions. The studies provide a scope of analysing the extents of ecological systems' limits and boundaries. The Big data paradigm, driving the articulated framework facilitates unravelling of new knowledge and implementing its connectivity among complex digital ecological systems of the LABHE.

## References

1. Al-Fares AA, Bouman M, Jeans P. A new look at the middle to lower Cretaceous stratigraphy, Offshore Kuwait, *GeoArabia*, Vol. 3(4), 1998, p. 543-560.
2. Al-Awadhi JM, Al-Shuaibi, AA. Dust fall-out in Kuwait City: Deposition and Characterization, Science of the Total Environment, Elsevier Publishers, UK, 2013.
3. Damiani E. Key note address on 'Digital Ecosystems: the next Generation of Service Oriented Internet", IEEE-DEST, Phitsanulok, Thailand, Feb, 2008.
4. Gilberg R. A Schema methodology for Large Entity-Relationship Diagrams, proceedings of the 4th International Conference on Entity Relationship Approach, Chicago, Illinois, ISBN O-13186-0645-2. October, 1985, pp. 320–327.
5. Gornik D. Data modeling for data warehouses, *IBM rational software white paper*, TP 161 05/02, Rational E-development Company, USA (2002).
6. Hevner AR, March ST, Park J, Ram S. Design science in information systems research, *MIS Quarterly*, Vol. 28 (1), 2004, pp.75-105, Society for Information Management and the Management Information Systems Research Center, MN, USA, ISSN: 0276 – 7783.
7. Hadzic M, Chang E. Ontology-based support for human disease study, published in the *38th Hawaii international conference on system sciences*, Hawaii, USA, 2005.
8. Indulska M, Recker JC. Design Science in IS Research: A Literature Analysis. In Gregor, Shirely and Ho, Susanna, Eds. Proceedings 4th Biennial ANU Workshop on Information Systems Foundations, Canberra, Australia, 2008.
9. Kemp DD. Exploring Environmental Issues – An Integrated Approach, Routledge Taylor & Francis Group, New York, USA, 2004.
10. March ST, Smith GF. Design and natural science research on information technology, Decision Support Systems 15, 1995, 251-266, Elsevier, UK.
11. Myers MD, Venable JR. "A Set of Ethical Principles for Design Science Research in Information Systems", Information & Management, Special issue on IS Ethics: Past, Present and Future, Vol. 51, No. 6 (September 2014), pp. 801-809, available online at http://dx.doi.org/10.1016/j.im.2014.01.002.
12. Newman P, Jennings I. Cities as sustainable ecosystems, Principles and Practices, Island Press, USA, 2008.
13. Neuman WL. Social research methods, qualitative and quantitative approaches, 4th Edition, Allyn and Bacon Publishers, USA, 2000.
14. Nimmagadda SL, Dreher HV. "On new emerging concepts of Petroleum Digital Ecosystem (PDE)", Journal *WIREs Data Mining Knowledge Discovery,* 2012, 2: 457–475 doi: 10.1002/widm.1070.
15. Ozkarahan E. "Database Management, Concepts, Design and Practice", Prentice Hall Publications, USA, 1990.
16. Shastri L Nimmagadda, Heinz Dreher. Data warehousing and mining technologies for adaptability in turbulent resources business environments, *Int. J. Business Intelligence and Data Mining,* Vol. 6, No. 2, 2011, p 113-153.
17. Schermann M, Hemsen H, Buchmüller C, Bitter T, Krcmar H, Markl V, Hoeren T. Big Data, An Interdisciplinary Opportunity for Information Systems Research, DOI 10.1007/s12599-014-0345-1, Springer Fachmedien Wiesbaden, 2014.
18. Venable JR, Pries-Heje J, Baskerville R. FEDS. A Framework for Evaluation in Design Science Research, European Journal of Information Systems (2016), 1–13 © 2014 Operational Research Society Ltd., doi:10.1057/ejis.2014.36. -26.