Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 1871–1880

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

# Big Data Guided Design Science Information System (DSIS) Development for Sustainability Management and Accounting

Shastri L Nimmagadda[a*], Torsten Reiners[a] and Gary Burke[b]

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

*Curtin University Sustainability Policy Institute, Humanities Department, Curtin University, Perth, WA, Australia[b]*

**Abstract**

S*ustainability* is a dynamic, complex and composite data relationship among geographically distributed human and environment ecosystems. The ecosystems may have strong interactions among their elements and processes, but with dynamic implicit boundaries. Multi-scalable and multidimensional ecosystems have significance based on a commonality of basic structural units and domains. We intend to develop a holistic information system for managing different ecosystems within a sustainability framework/context, using an empirical qualitative and quantitative interpretation and analysis of the measured observations. Design Science Research (DSR) approach is aimed at developing an information system using the volumes of unstructured Big Data observations. Collaborating multiple domains, interpreting and evaluating the commonality, uncovering the connectivity among multiple systems are key aspects of the study. The Design Science Information System (DSIS), evolved from DSR approach is used in solving the ecosystem issues associated with multiple domains, in which the *sustainability* challenges manifest. In this context, we propose a human-environment-economic ecosystem (HEES) framework consisting of human, environment and economic elements and processes. In broad terms, human, environment and economic domains are conceptualized as different players/agents that operate within a range of *sustainability* scenarios. This approach recognizes the existing constraints of the systems as well as the emerging knowledge of the boundaries of ecosystems and their connectivity. The connectivity and interaction among the systems are analyzed by data mining, visualization and interpretation artefacts within a sustainability policy framework.



Peer-review under responsibility of KES International

*Keywords:* Design Science; Digital Ecosystem; Sustainability; Multidimensional Artefacts; Data Interpretation;

* Corresponding author. Tel.: *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.2017.08.233

## 1. Introduction

Within a sustainability scenario, as long as the symbiotic conditions exist, the human ecosystems survive. Human activities exist in multidimensional contexts and domains; in particular within environmental ecosystems. Human survival and well-being depend either directly or indirectly in ways in which natural assets and resources are nurtured, protected, used and managed in our biophysical and sociocultural environment. There is growing concern about the emergent social, environmental and economic risks surrounding the existing progress, attained through a variety of commonly intertwined modernization processes[2]. Sustainability describes the surroundings in which the human and environment entities work together with the creative and active symbiosis, permitting and satisfying the cultural, social and economic necessities of present and future generations, including the legacies of past wastefulness and exploitation of natural resources. We propose a structure HEES, in which human activities are conceptualized as an ecosystem, comprising of cultural, social, economic and political systems. Each participant of an ecosystem has potential to contribute to or detract from the sustainability of the system of which they are subset[10] as demonstrated in Fig. 1. The dataflow is shown as arrows( ↔ ) among various ecosystems. The human, environment and economic domains/ecosystems are emphasized as entities from which the connectivity can be established through an integrated framework, simulating amenable artefacts.

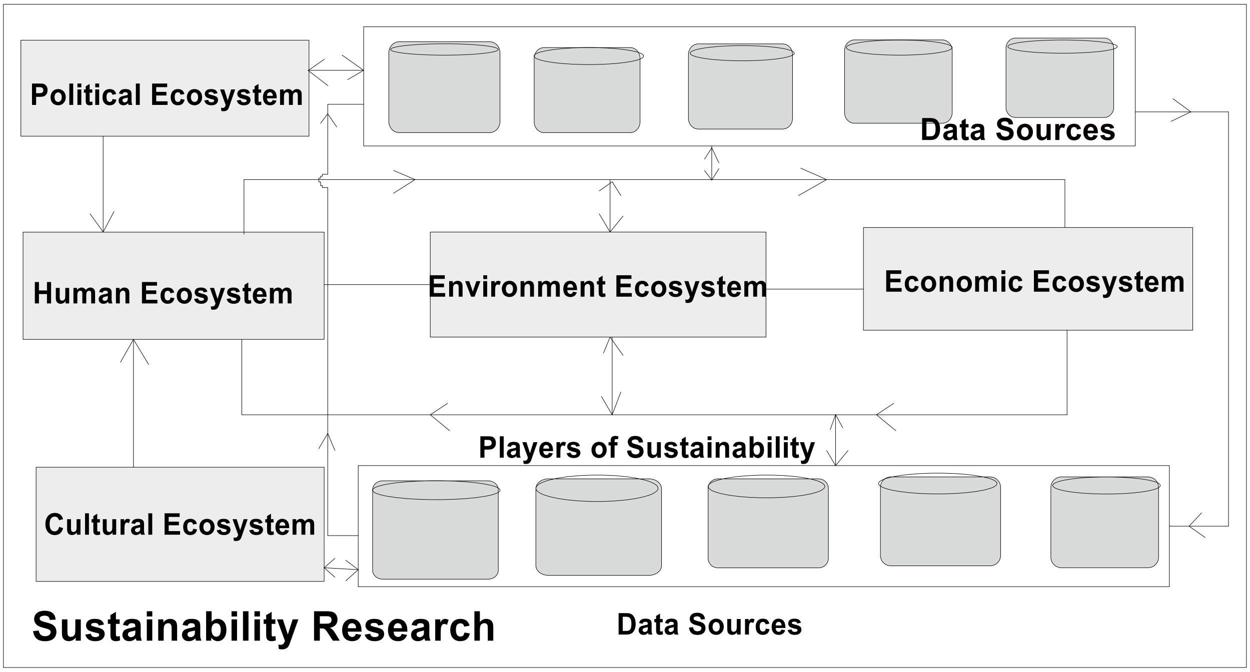


Fig. 1: *Sustainability* and its players in the HEES structure[10]

For the purpose of integrating domain ontologies with multifaceted dimensions, the data warehouse approach is adopted. However, various challenges encountered in data integration tasks are analyzed for envisaging multiple domains in ecosystems' research that embraces the conceptualization and contextualization values. For broader geographically varying data sources, multidimensionality and heterogeneity contexts need comprehensive and more generic data solutions. The data may be from a variety of sources, such a laboratory or a field project and specific types, from decisive repositories, such as a particular demography or environment category. The conceptualization and contextualization with more specialized ontologies or semantics complicate the methodological integration in a single warehouse repository. The data designs and access methods evolve and often change over time. The data source representing a system may have an intensive and fortuitous impacts on other systems that may cause the integration to fail on new formats. As our understanding of interconnectedness among elements and processes of the ecosystems develops, with increasing complexity of data structure, even with subtle changes in variables may have big impacts on data models. These challenges and issues have made us thinking an alternate[13] methodological framework that can handle data analytics and new knowledge management in the ecosystems' arena. The data analysts and scientists need more flexible robust data modelling methodologies that can facilitate retrieve fine-grain data, information and knowledge. Such inventories can facilitate the supply of resources in demanding and sustainable businesses.

## 2. Issues and challenges of the existing research

In the first instance, significant challenges of aligning diverse ways of concepts, across disciplines are addressed. A critical issue arises because of the semantics related to data sources of diverse ecologies – and the conceptualizations from which they derive are hard to define[6, 7, 8], because they are heterogeneous, multidimensional and dynamic, difficult to define categorically and comprehensively. Nevertheless, attributes and properties that describe different aspects of the concepts and data are interpreted and included in the data modelling. The mere challenge is the interpretation of HEES, an encapsulated ecosystem development, with the shared classification of vocabularies and terminology, dependent on semantic perspectives; they may well be understood hypothetically, but they do not have a precise meaning across all the dimensions and dynamics in which the ecosystems exist and function. The community has considerable knowledge of semantics comprehensively and explicitly maintaining the needs of the research. The HEES structure is composed of several large domains with numerous attribute dimensions, making the structure more complex and unstable, but complemental to many focused research domains and communities. If researchers wish to retrieve data from a single domain, it is not a major issue, but in the context of HEES, research communities involve exploring solutions from multiple entities and domains. Researchers need access to integrated metadata of HEES in diverse domains and visualizations, resolving complex contemporary issues. The vocabularies, terminologies and their semantics evolve over time and geography. They require agreeable different semantics even across dissimilar communities, making the HEES work flexible and perform well even when observations from a particular community use terminology and data that are not unambiguously recognized. The associated domains of HEES have the capacity to hold and handle characteristic structure/axiom information. To this extent, the ecosystems of HEES provide sequential information and annotation that can allow a combination of definitions and evolving terminologies in multiple domains[10].

## 3. Data integration

An ecosystem is deemed 'useful composited entity with spatial-temporal dimensional habitat'. The juxtaposition of coexistent habitats (systems) in various scales in nested hierarchies may result to overlaps and discontinuities. The key challenges are data integration, analysis and sharing of information from different ecosystems. The strategic feature of the research is organizing, structuring[8], and integrating the existing data in multiple realms of HEES, filling the gaps of conventional methods. To date, an orderly analysis of the data sources using dedicated technologies is lacking. Careful examination of *sustainability* issues, how they manifest and impact in associated ecosystems in their hierarchies is research that could be of enormous assistance to scientists and policymakers. A methodology is required that can elucidate the data patterns that might be found in chaotic and an amorphous volumes of huge data stores. As shown in Fig. 2, there is an opportunity for cross-domain research in bringing and connecting governmental and non-governmental organizations through domain ontologies. Analysis of data patterns, correlations and trends amongst data views and delivery of ecosystem services within a *sustainability framework* through rigorous analysis of facts is one of the key objectives of the research. The foremost focus is on how the connectivity among systems can be identified through analysis of data trends and patterns. Visualization can be used to clearly demonstrate the latent or unknown knowledge from these patterns that add depth and meaning to existing knowledge and interpretations.

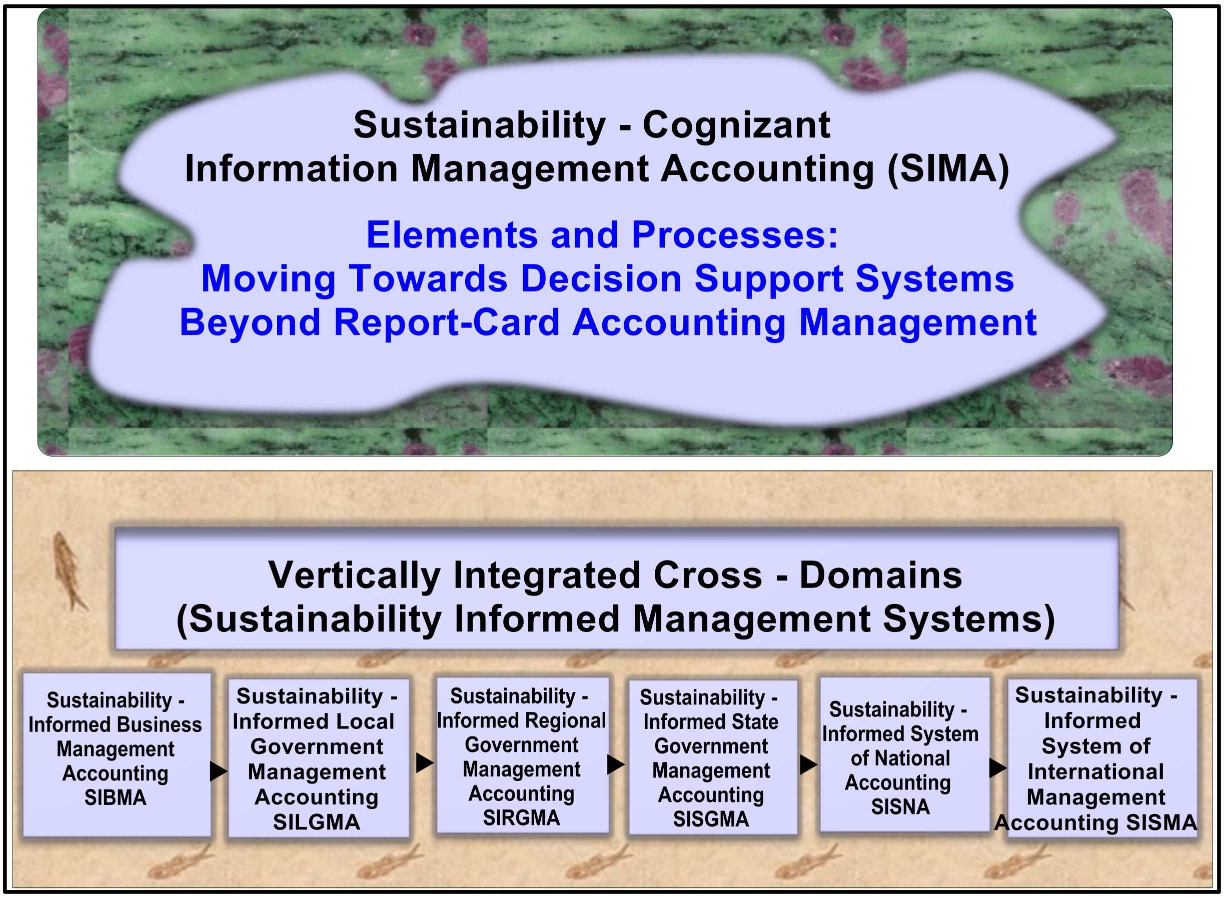


Fig. 2: Cross domain research elements and processes and their connectivity[1]

In a digital ecosystem perspective, the simulation[4, 5, 9] of an integrated framework is described. The major challenge is undertaking thousands of existing data entities, dimensions and their associated attributes, including their conceptualized and or contextualized dimensions. The constructs and models with flexible and varying business rules and constraints in a sustainability-informed framework can provide an innovative perception to sustainable ecosystems research. The challenge is demonstrating the models that built based on conceptualized dimensions of practical and applicable sustainability actions can deliver new knowledge to facilitate the needed research outcomes.

The process flowchart drawn in Fig. 2, demonstrates the multidimensionality and heterogeneity of data sources associated with different ecosystems. Understanding the data integration processes, involved therein requires connecting Big Data sources. The Big data characteristics – conceptualized as dimension attributes – are incorporated in data structuring processes in multiple domains. We aim at developing a robust, holistic and flexible DSIS comprising of data- various artefacts that can facilitate an effective implementation in the DSS.

## 4. Complexity of sustainable ecosystems

The sustainability[1] is the capacity to endure, regenerate and flourish through time. The governance and culture are core values of the sustainability, how they can be interrelated with ecological systems in a way the data science and strategic knowledge ease the decision support systems. In the information system perspective, it is a composited dimension, signifying that sustainability-related impediments cannot be addressed adequately in a single demography, geography or an environment in which, the human ecosystem survives with economic gains. The sustainability lies with the fact appreciating the governance, cultural system and how they interrelate with environmental systems so that they are better designed and managed to produce science and knowledge for strategy making purposes. In the information systems perspective, the sustainability is viewed as a merged attribute dimension thereby signifying that the sustainability-related complications cannot be addressed adequately in a single perspective. Whether it is one demography or one culture, the environment in which the human ecosystem takes on other coexistent ecosystems, create economic gains or values. Although the concept of sustainability is evolving and dynamic, the ambiguity in its perception and description, as well as the relationship, the associativity among coexistent systems within the purview of ecosystems and their limits are clearly interpreted in the data science, knowledge and engineering perspective. The Big Data technologies are exploited to strengthen the concept of *sustainability* with data warehousing and mining accentuations and articulations. The HESS attributes such as the environmental acuteness, economic viability, socio-technical transition and corporate social obligations are interpreted in the integrated articulations. Various components and requirements of the framework simulations are analyzed, how sustainable they are in the multiple ecosystems' contexts, in a way the Big Data sources are better documented and managed. Besides, an understanding of ecosystems functions and their connectivity is emerging with dynamics of ecology and life sciences. There are large amount and variety of heterogeneous data to prove these phenomena. Managing huge volumes of data sources and extracting new hidden knowledge from these ecosystems is a major challenge that, once resolved, a huge benefit to the decision makers across the globe.

With the advent of Big Data concepts and tools, researching the digital ecosystems has taken a new direction to uncover the vast amount of new information and knowledge. The digital ecosystems are simulated with various artefacts and integrated into a robust and holistic methodological framework. In this process, various dimensions, attributes and their instances are incorporated with Big Data events and their characteristics. Various activities accounted from sustainability-informed management, and accounting system is linked to the Big Data events to support the decision making. As an example, in the methodological framework, the data modelling, data integration, data mining and visualization artefacts are used to generate metadata. Various fine-grained data views are extracted for interpreting new knowledge of systems' elements and processes and add values to ecosystems' projects and services. The fine-grained data structures, along with their impacts and benchmarks can sustainably comprehend the fabric of the connectivity between social-cultural, environmental and economic ecosystems as described in Fig. 3.

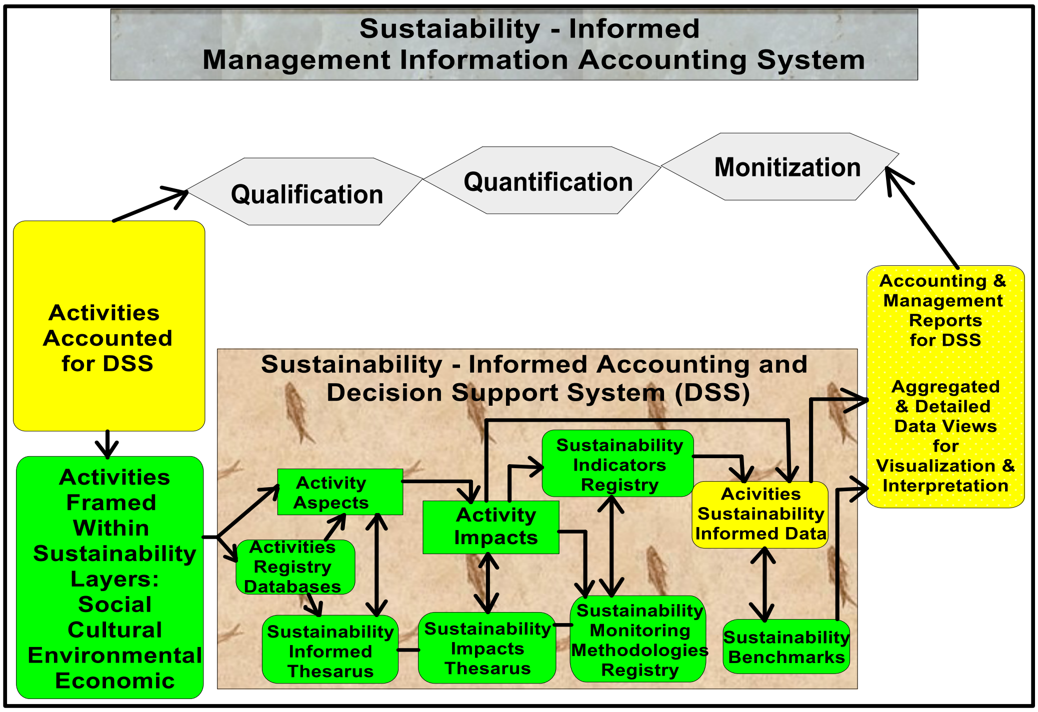


Fig. 3: Demonstrating the connectivity and evaluating it through various layers[1]

We further describe the phenomena of ecosystems in the context of associated multiple domains and various components involved in the depiction of their digital ecosystems. We analyze several components and requirements of the digital ecosystems and technologies (DEST) framework simulations that lead to the development of DSIS, how sustainable they are in managing the Big Data sources so that endurance and combined impact of the multiple ecosystems are aptly assessed. The dimensions of the framework are described in Fig. 4, making connectivity through a star schema. Each of the dimension contributing to the integrated framework is described in the sections (a) – (f).

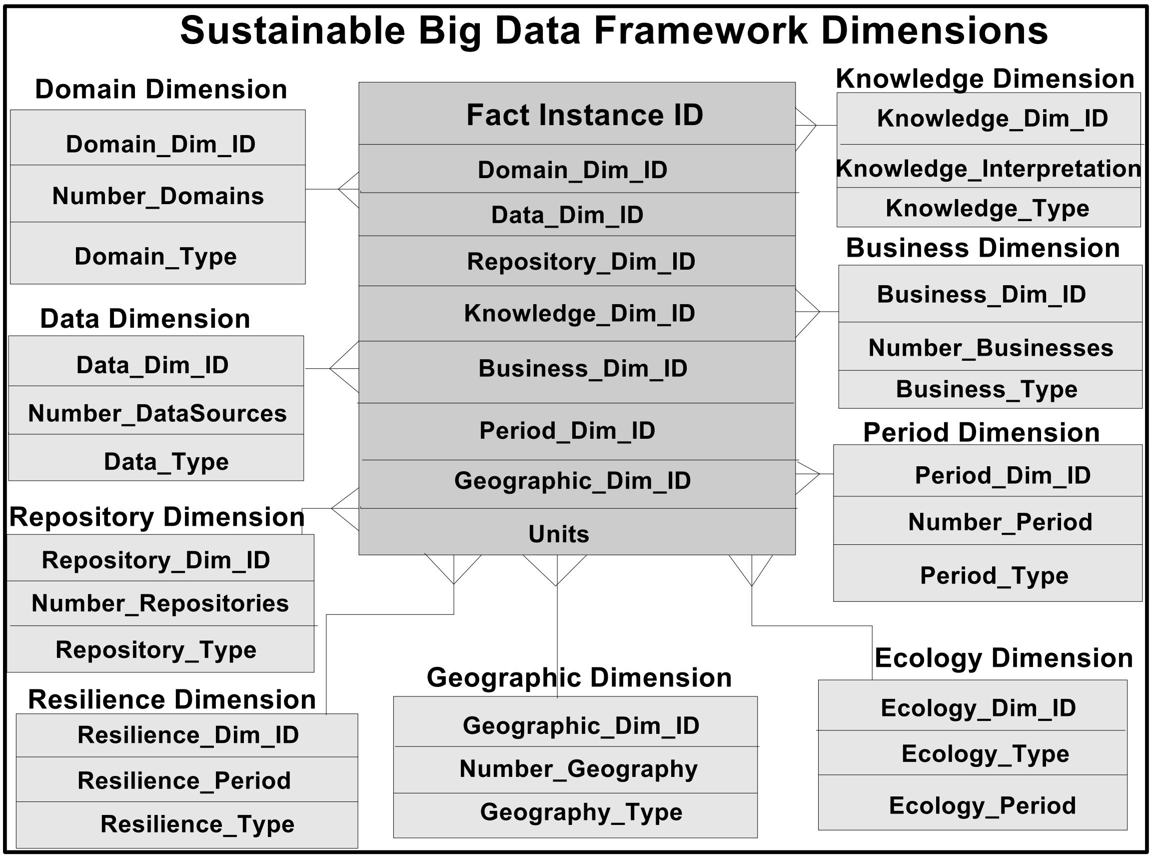


Fig. 4: Star schema model representing the Big Data dimensions

From an informatics perspective, the key aspects of sustainability analysis may be regarded as a function of the domain (context), data integrity and salience, repository accessibility and security, knowledge cogency and adequacy, business viability and ecological function and resilience. In the IS perspective: Sustainability = *function* (domain, data model, warehousing & mining, visualization & Interpretation, building new knowledge for businesses) = *function* (domain context, data integrity and salience, repository accessibility and security, knowledge cogency and adequacy, business viability, ecological function and resilience); these components are briefly described in the following sections.

*(a) Domain context*

In the computing world, the domain addresses and controls the space or area in which common attribute dimensions are shared. For describing the domain sustainability in the current context, we explore the other domains, closely associated or connected to the HEES entity. For example, the geomorphic and environmental systems are closely related to petroleum ecosystems[9, 11] from which common element and process attribute dimensions of the domain spaces can be interpreted. Besides, we realize the human, economic and technological domains are closely linked to the petroleum domain to share the common space and economic gains.

*(b) Data integrity and salience*

Data integrity and salience depends on representative domains linked to the current data sources and interconnected for an effective data modelling process[1, 2, 3] that accords with sustainability principles and perspectives. Heterogeneity and multidimensionality also affect the data modelling method. Use, reuse, efficient, effectiveness, completeness of the data structures and how they respond to different domains in a unified metadata structure are described. Data analysts examine how the repositories respond to the number and size of the data sources including storage and processing capabilities. Data integrity and salience also relies on the dynamics of knowledge creation and how the integrity of data is maintained/validated within each domain entity—for example, when the tolerance of business and production entities may have to be achieved.

*(c) Repository accessibility and security*

Warehouse repositories rely on descriptive domains and data models, including how logically and physically they are integrated into the data warehouse environment. Often, complex issues have aspects and limits that are ambiguous: that is, they can legitimately be interpreted in different ways, depending on the epistemological framework[12, 13]. However, the impacts of ambiguity can be reduced when attributes and properties are assigned to improve data qualities and domain knowledge. This can affect the interpretation of variables that impact on sustainability data in warehouse repositories. We demonstrate multidimensional data warehouse repositories[9] in the context of closely linked ecosystems, illustrating how the Big Data scaled metadata structures sustainably collaborate with the cognitive knowledge and business data analytics artefacts of the integrated framework.

*(d) Knowledge cogency and adequacy*

The cogency and adequacy of knowledge depend on the domains, data sources and repositories described within the purview of a digital ecosystem. When considering the HEES structure, the knowledge relies on the facts of indulgence from multiple domains and data models that are relevant to sustainability attributes in business and geographic dimensions. The knowledge interpreted in the form of areal extents of ecosystems and their boundaries have immense scope and outlook in the contexts of bridging business viability and ecosystem sustainability. The knowledge of the connectivity and interaction of ecosystems facilitate better management and production of data science in decision support systems.

*(e) Business viability*

Analysis of business viability occurs within closely associated domains of multiple digital ecosystems, data sources and sizes; the focus is on the integration of levels of sustainability-informed knowledge in the business process. The prior knowledge of domain and data may help in assessing the business opportunities. The knowledge obtained from closely related human, environment and economic domains (HEES) and associated data models can sustain for longer periods, larger geographic and demographic dimensions, so that businesses can survive for longer periods based on viable, sustainable parameters.

*(f) Ecological function and resilience*

The concept of resilient ecosystem[1, 11] describes functions that can be used to improve environmental monitoring and management. The concept of ecological resilience recognizes multiple conditions and the ability for systems to resist and maintain the functions as described in (a) – (e) (above). Ecosystems are emerging, evolving and dynamic with time. Accordingly, from an information systems' informatics perspective, the sustainability framework accords with the goal of more resilient ecosystems that can create synergies within the HEES entity. This entity has multidimensional attribute dimensions. The degree at which an ecosystem function can accommodate or recover rapidly from environmental perturbations or human-induced disturbance, is analyzed in a way that function is maintainable/sustainable to leverage and enhance social and economic activities. The multifaceted data attributes are made good use of analyzing the ecological functions and their resilience.

Besides, analytical tools for elucidating the connectivity among embedded digital ecosystems (HEES), various business intelligence (BI) and data analytics tools are proposed. A rigorous approach for evaluation of DSIS methodology is fundamental to the sustainability informed management and accounting research. In this approach, analysis of multifaceted data and their connectivity in multiple domains and sources respect the ontological cogency of the original data. This integration across domains can significantly risk minimize the ambiguity of information needed in the integrated interpretation of knowledge among varied ecosystems. The proposed technologies provide immense scope for future analysis and policy development across multiple domains and ecosystems. There is a need to examine the Big Data on diverse constructs and models to test the phenomena and hypothesis. For appropriately accessing, extracting and investigating the data views and their knowledge, the scope and opportunity of agents and interfaces in the integrated frameworks are explored. The robust methodologies are expected to facilitate the framing, analysis and strategic development of strategies and policies for ameliorating *sustainability* issues, creating benchmarks and analyzing impacts so as to provide multidimensional sustainability-informed analysis at micro, meso and macro levels worldwide[1]. The registry or inventory of the assets of existing ecosystems, their attributes and resources (in a sustainability scenario) is analyzed to forecast, and provide quality data and information to the managers and policy makers of the ecosystem providers. The HEES is a simulated version of DSIS demonstration, a digital replication of the sustainability informed management and accounting registry, which is not only for reporting but the decision support systems' (DSS) applications. Providing economically and environmentally sustainable ecosystem services are the core objectives and contributions of the current research. Logistics and supply chain and disaster management are other potential areas of future research for enhancing the practicality and capacities for implementing new sustainability-informed policies and strategies.

## 6. Significance

The government agencies, academicians, domain experts and ecosystem researchers including data analysts are the beneficiaries of the ecosystem services. The multidimensional data resources in multiple domains and their knowledge have impacts on examining the research problem solutions more critically. For knowledge base systems and sustainability research solutions, more structured, fine-grained data and quality information are demanding. The intricacy, heterogeneity and multidimensionality complicate the description of the composited HEES entity, and for designing and developing the integrated articulations, a new direction is needed. Implementation and validation of the data models in the HEES context are our continuing research efforts.

## 7. Methodologies

We have undertaken and followed up the study through secondary data observations with qualitative and quantitative interpretation efforts. Various artefacts are created using Big Data tools and technologies, comprehending the DSIS development. The DSIS has varied opportunities[13] in the context of resolving ecological issues. An integrated framework[9], driven by DSIS-Big Data approach, as illustrated in Fig. 4, is intended to be generalized for different research application areas. The domain ontologies in the context of HEES structure are in the form of the logically connected constructs and models, accommodating in a data warehouse environment and enabling interpret coexistent ecosystems in a sustainability-informed repository or registry. Data mining schemes, visualization and fusion tools are adopted to analyze the metadata. Exploring the connectivity between systems that affected by socio-cultural, economic instabilities and imbalances are interpreted cognitively through data view trends with periodic and geographic dimensions. Investment opportunity decisions to be made in the ecosystems' projects and services are influenced by global economic trends. The native title requirements, cultural heritage protection, human ecology and environment protection, even mitigating the political instabilities are other specific issues resolved with economic constraints.

### *7.1 Big data assembly*

The secondary data volumes and varieties though are available online and or offline, but at present, all are in haphazard manner[10]. For the purpose of Big Data assembly, large size data sources are identified and documented for simulating a conceptual framework as demonstrated in DSIS in Fig. 1. Several research activities and expected outcomes are put in a cuboid structure with a focus on artefacts' design and development of the DSIS. Big Data features and components[10, 11] (as described in Fig. 5) are incorporated within DSIS, following up the framework with DSR guidelines and evaluation properties[11]. Both the longitudinal and lateral design features are incorporated in the Big Data assembly connecting to its characteristics as explained in Fig.5. The fine-grain data views expected for qualitative and quantitative interpretations may be effective when the cross-sectional designs are incorporated with both the longitudinal and lateral dimensions to arrive at significant *ecosystems* and *sustainability* research perspective. The volumes and varieties acquired and integrated, are empirically analysed through the proposed integrated framework. Various data mining schemes are used for evaluating the multifaceted data views and other types of ecosystems' data.

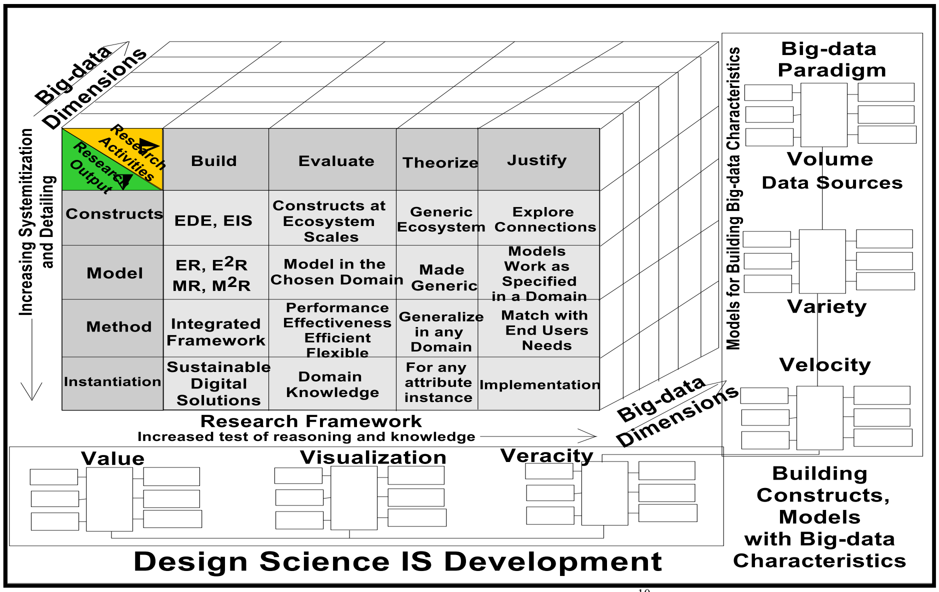


Fig. 5: A conceptual research framework[10]

The real challenge is an implementation of the new research methodology in several application domains. The statistical data correlations, graphic visualisation and fusion, neural computing and other data mining schemes help evaluate and facilitate the cuboid metadata interpreted from simulated integrated framework efficiently.

*7.2 Method of evaluating the constructs, models and their properties*

The use, reuse and effectiveness of constructs, models, methods and knowledge[4] are typical in evaluating utility properties[13]. The constructs, models and methods that make up the DSIS are analysed using evaluation utility properties and other technology rules[13]. Dimensions, the strengths of their attribute instances and facts of HEES are sensitive to the connectivity and boundaries of multiple ecosystems, positively constructive to the survival of sustainability phenomena. *Data quality, use, ease of use, interoperability and effectiveness* of data models are some of the evaluation properties of the constructs and models, analyzed robustly to make the DSIS more holistic and applicable in wider domain applications.

## 8. Research deliverables, impacts of research solutions and economic benefits

We describe the impacts and economic benefits of the research solutions in various industry and service sectors:

- Constructs, models and methods for DSIS development with HEES focus.

- Envisaging the concepts of sustainability, establishing the limits and boundaries of ecosystems that may have impacts on delivering ecosystems' services provided by government organizations.
- To enable use and reuse of structured data and domain knowledge in multiple domains.
- Digital ecosystem approach can unveil new knowledge from embedded ecosystems.
- Accounting/inventory system of ecosystems' resources.
- The economic benefit of an inventory of the renewable resources within a single repository including better management of the ecosystems services.

The government organizations, environmentalists, researchers and academicians, DB experts, data analysts are beneficiaries of the current commercial design science research.

## 9. Conclusions and recommendations

We focus on simulating and developing sustainable digital ecosystems and services to deliver to various producing/manufacturing and service companies worldwide. In this context, we use the Big Data guided integrated methodologies that can handle the unstructured heterogeneous and multidimensional data sources and their integration, enhancing the value addedness of the large scale HEES project. Sustainability is analysed in response to the increased awareness of Big Data characteristics that control the integrated articulations of the framework. The artefacts designed in the integrated framework prove to be sustainable and implementable in the large scale HEES projects.

## 10. Future scope and outlook

Our ongoing research is on implementing the workflows and models, presented in various ecosystem perspectives of the HEES. We also explore new research avenues of DSIS design, development and implementation in the ecosystem scenarios. The qualitative and quantitative analysis of the data views obtained from metadata cubes are our future research motivations, besides evaluating and validating the data models of the HEES.

## References


1. Burke GR. (2013). Making Viability Sustainable, PhD Thesis, Department of Humanities, Curtin University, Perth, WA, Australia.
2. Coronel, C., Morris, S., and Rob, P. (2011), Database Systems, Design, Implementation and Management, Course Technology, Cengage Learning, USA.
3. DOVERS SR., NORTON T, HANDMER JW. (2001). *Ecology, uncertainty and policy : managing ecosystems for sustainability,* New York, Pearson Education.
4. Jarrar M., Meersman R. (2002). Scalability and Knowledge Reusability in Ontology Modelling. Proceedings of the International conference on Infrastructure for e-Business, e-Education, e-Science, and e-Medicine (SSGRR'2002s).
5. Kemp R, Martens P. (2007). Sustainable development: how to manage something that is subjective and never can be achieved? *Sustainability: Science, Practice, & Policy,* 3**,** 5-14.
6. Meersman R. (2000). Can Ontology Theory Learn from Database Semantics? Proceedings of the Dagstuhl Seminar 0121 'Semantics on the Web.
7. Meersman R. (2001). New Frontiers in Modelling Technology: The Promise of Ontologies. Proceedings of the SISO ESM Conference on Simulation.
8. Meersman RA. (2004). Foundations, implementations and applications of web semantics, parts 1, 2, 3, lectures at School of Information Systems, CBS, Curtin University, Australia.
9. Nimmagadda SL. (2015). Data Warehousing for Mining of Heterogeneous and Multidimensional Data Sources, Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG, p. 1-657, Germany.
10. Nimmagadda SL, Burke G, Reiners T. (2016). On Big Data Guided Embedded Digital Ecosystems (EDE) and their Knowledge Management, Proceedings of IEEE Industry Informatics, Futuroscope, France.
11. Oliver TH, Isaac NJB, August TA, Woodcock BA, Roy DB, Bullock JM. (2015). Declining resilience of ecosystem functions under biodiversity loss, Nature Communications, 6, Article Number: 10122, doi: 10.1038/ncomms10122.
12. Venable J, Pries-Heje J, Baskerville R. (2016). FEDS: A framework for evaluation in design science research, European Journal of Information Systems, 1-13, doi:10.1057/ejis.2014.36.
13. Wand Y. (2000). An ontological analysis of the relationship construct in conceptual modelling, ACM Transactions on Database Systems, Vol. 24 (4), pp. 494-528.