

# On Knowledge-Based Design Science Information System (DSIS) for Managing the Unconventional Digital Petroleum Ecosystems

Shastri L Nimmagadda<sup>1</sup>, Neel Mani<sup>2</sup> and Torsten Reiners<sup>1</sup>

<sup>1</sup>School of Management, Information Systems, CBS, Curtin University, Perth, WA, Australia

<sup>2</sup>CONSUS, School of Computing, University College Dublin, Ireland

Email: `shastri.nimmagadda@curtin.edu.au`

**Abstract.** The unconventional digital petroleum ecosystems are associated with fractured reservoirs that are usually unpredictable, but can produce for longer periods depending on size of petroleum systems and basins. Currently, conventional reservoirs do produce oil & gas even without integrated workflows and solutions. The heterogeneity and multidimensionality of data sources at times can make the data documentation and integration complicated affecting the exploration and field development. We examine the conventional database technologies and their failures in organizing the data of unconventional digital ecosystems. Big Data driven intelligent information system solutions are needed for addressing the issues of complex data systems of unconventional digital ecosystems. Geographically distributed petroleum systems and their associated reservoirs too demand such integrated and innovative digital ecosystem solutions. We propose an innovative design science information system (DSIS), an integrated digital framework solution to explorers, dealing with unconventional fractured reservoirs. The integrated Big Data analytics solutions are effective in interpreting unconventional digital petroleum ecosystems that are impacted by shale prospect businesses worldwide.

**Keywords:** Digital ecosystem, domain ontologies, big-data, data integration, data warehousing and mining, data fusion, unconventional data sources, evaluation properties

## 1 Introduction

Worldwide sedimentary basins [5] comprise of numerous conventional and unconventional shale deposits. Each of these basins has the capability of generating and producing numerous unconventional oil and gas fields [6, 8] from multiple petroleum systems and basins. Each field has multiple oil and gas producing wells and each drilled-well has multiple unconventional reservoir pay zones, with each pay zone having different

fluids - either oil and or gas. In a basin scale, this is a typical data hierarchy with existence of super-type and subtype dimensions that attribute to an *oilplay* [1], a scalable and atomic property of a *prospect*. The source rocks are unconventional reservoirs. In a digital oil field scenario, the data sources interpreted as entities and or dimensions in multiple domains are logically organized through ontological descriptions to integrate in a data warehouse environment [12, 13]. The data in multiple unconventional oil and gas fields are organized in a way their connectivity can be established from multiple databases, using semantic tools and technologies. In the present research, issues associated with existing methodologies, significance and motivation of innovative approaches and how they address the challenging issues are discussed.

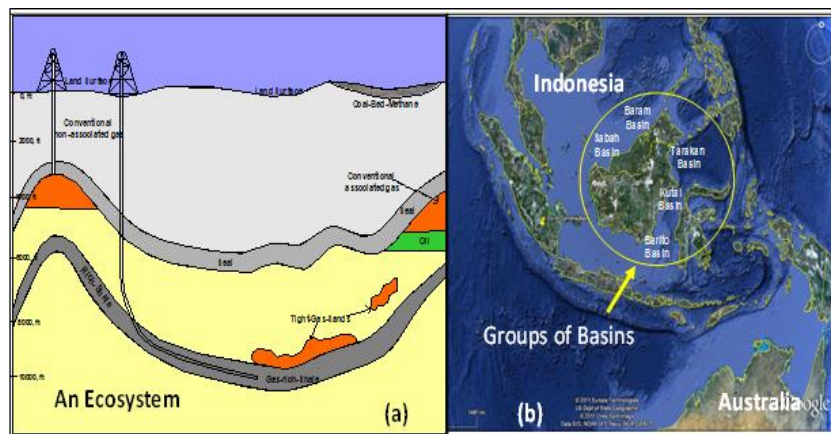
## **2 Previous Studies**

There has been an explosive growth in the volumes of published petroleum data, as a result of the rapid exploration and development activities worldwide [5, 13]. Organization of such data by industry's professional organizations, national, private oil and gas companies including major service companies is a huge challenge. This has motivated us to organize and manage a wealth of data [2, 3, 13] accumulated in multiple sources linked to different ecosystems [13]. Several elements and processes involved in building a digital ecosystem are described in [11]. The feasibility and applicability [13] of the proposed technologies are analyzed for exploring the connectivity and the boundaries of system elements and processes. Use of ontologies, implementing them in a data warehouse approach is discussed comprehensively in [13, 14]. Use of sets and set theory in different applications are given in [7, 14]. How the principles of set theory can adapt to Big Data features are described in [12].

## **3 Issues and Challenges**

Around the world, the unconventional gas resources are widespread, but with several exceptions that they have not received careful attention by natural gas operators, especially with lack of logistics for exploiting unconventional petroleum systems. In addition, there is a chronic shortage of expertise in specific technologies needed to develop the unconventional resources successfully. As a result, only limited development of technology has taken place outside of North America. An interest is however growing and during the last decade, efforts are being made to develop unconventional gas reservoirs worldwide. Integrating or assembling information from several databases to solve problems and discovering new knowledge are other major challenges in geo-informatics. Exchanging volumes and variety of exploration data into valuable geological knowledge is the challenge of knowledge discovery process. Data mining and interpretation of interesting patterns hidden in petabyte size of seismic and other exploration data are critical goals of geo-informatics. The goals cover identification of useful reservoir bearing structures from multiple petroleum ecosystems [11, 13] and their connectivity that are of commercial interest to explorers and investors.

In addition, other major issues are poorly managed upstream integration with multi-faceted Big Data dimensions. Besides, heterogeneity and multidimensionality of conventional data structures (relational, hierarchical and networking data types, [12, 13]) including complex spatial-temporal data [9] sources pose serious challenges. Similar issues [14] exist in exploration and production of unconventional Big Data sources. Massive storage devices are needed for volumes of business/technical data dimensions (instances to the petabyte size for each unconventional oil and gas field, an entity in a major producing basin) that describe oil and gas fields: for each field – number of *surveys* and number of *drilled-wells*; in each well, number of *horizons* (geological formations) and in each survey: number of survey lines, with hierarchical and relational ontology descriptions [14]. Handling numerous data dimensions (entities and or objects) to the order of thousands of attributes, mapping and modeling their dimension and fact tables are tedious process. Data integration (sometimes among 50 to 80 geographically distributed operational centers with an Indonesian case study [17]) of multi-disciplinary data dimensions is a serious business issue in large oil and gas companies.



**Fig. 1.** a: Representation of (a) unconventional resources within a single petroleum ecosystem; (b) Google imagery view showing Southeast Asian Sedimentary basins and their inherent connectivity

As illustrated in Fig. 1, multiple sedimentary basins, associated petroleum systems and sets of oil and gas fields existing in a marine environment all linked together, constitute an ecosystem. In conventional petroleum ecosystem environment, the oil and gas can produce without much effort on exploration and field development because of types of reservoirs, and their qualities. In unconventional digital ecosystems, the reservoirs are unconventional and their petroleum sources are typically shales with network of fractures [2] and cannot produce on their own unless the shale rocks (petroleum-source) are stimulated. In the ecosystems view point, we are of the opinion that both conventional and unconventional systems share information and knowledge, all encapsulated in large size digital ecosystem [13]. This phenomenon opens new opportunities to develop smarter and knowledge base technologies in upstream petroleum industries.

## 4 Motivation and Significance

The resources needed in running large size integrated upstream projects are abundant. The heterogeneous and multidimensional Big Data sources that surround the upstream oil and gas companies motivate us to develop new data integration methodologies. Knowledge-based and smarter multidimensional data constructs and models are proposed for designing and developing an integrated design science information system (DSIS) framework [9, 16]. Besides, the conceptualization and contextualization attributes of various entities and dimensions that support the innovative methodology have further led to develop a multidimensional digital ecosystem-based inventory or repository. Additionally, the fine grained multidimensional data structuring with domain ontology descriptions makes the DSIS approach more smart and flexible in an environment, where varieties of business rules apply to models and when the constraints rapidly change [14] either in geology or upstream business. The visualization and interpretation are other significant artefacts of the Big Data facilitating the use, reuse, test the interoperability and effectiveness of the data models for sustainable DSIS including an effective data analytics [12, 14].

## 5 Methodology

Designing and developing an integrated digital framework have been the focus of current research in integrating intelligent digital ecosystem scenarios [7, 13, 14]. In worldwide unconventional shale-gas basins multiple petroleum systems (information systems) exist with variety of elements and processes and volumes of their fact instances. We examine these unstructured data sources for modelling unconventional digital petroleum ecosystems. The conventional petroleum resources inherently embed with unconventional shale gas, shale oil and coal bed methane (CBM) resources, all existing within a single digital ecosystem scenario. A robust methodology [3, 12] is vital for managing the digital ecosystem scenarios under the control of Big Data sources and their sizes. Design Science Information System (DSIS) strategy is an alternative approach, addressing the data modelling and integration challenges with heterogeneous and unstructured data sources. Big Data-driven DSIS is simulation of an ecosystem framework [9, 12] in which several research activities are articulated including implementation of research outcomes that are validated by various evaluation properties.

Unconventional digital ecosystems hold large volumes and varieties of data in multiple domains. For designing, developing and implementing the digital ecosystems in Big Data scale, high-quality and reliable data sources are necessitated. For the purpose of bringing multidimensional heterogeneous data into the integration process [13], several data instances are gathered from unconventional ecosystems. Hundreds of dimension- and -fact tables including attributes and their corresponding data instances are considered in the modeling purposes. These attributes are either hierarchically or relationally structured or both used for modeling the data instances in a metadata [14]. Several *point*, *line*, *spatial* (areal) dimensions are ontologically described and integrated

in such a way data structures can establish connectivity between seismic- and well-domain datasets. Geographically metadata are mappable for fractured network models.

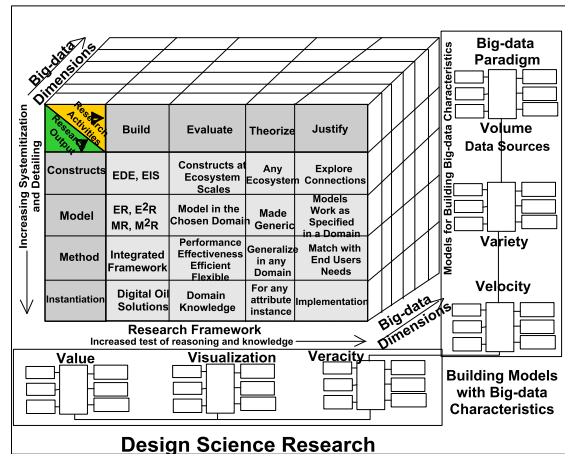


Fig. 2. Big data supportive DSIS Framework

For building data structures, the domain ontologies are described, supporting the DSIS warehouse repository. It can connect hundreds of logical multidimensional data relationships from digital ecosystems (Fig. 2). We reiterate that the concept of a digital ecosystem in the DSIS context is to bring multiple data dimensions together and connect with manifold multidimensional schemas. We incorporate the characteristics of Big Data logically in DSIS constructs and models, integrating several multidimensional warehouse repositories into a meta-metadata. Domain ontologies search for connections between disconnected reservoirs through interconnected multidimensional data relationships and their associated schemas. Various algorithms are described in [14] demonstrating the artifacts including their integration process. Other artifacts in the form of ontology models that go into the DSIS are described in the following sections.

### 5.1 Time-Domain Ontologies

The multidimensional data models are in both time- and depth-domains, with ontological descriptions for every time- and - depth ranges. Information on lithology existing in different intervals of drilled wells and their time occurrences in the seismic data are incorporated in the multidimensional modelling process. The seismic exploration data are represented in different space and time dimensions and separate ontologies are written for multiple seismic vertical time ranges: 0-500ms, 500-1000ms, 1000-1500ms, 1500-2000ms and 2000-2500ms. Each vertical time range has different geologically interpretable knowledge, such as horizons, structures, reservoirs, source and seal rocks, and other various petroleum system's processes. Data instances from a time-domain cross section are documented as shown in Fig. 3. This cross section covers a complex reservoir, interpreted from integrated seismic and drilled-well metadata.

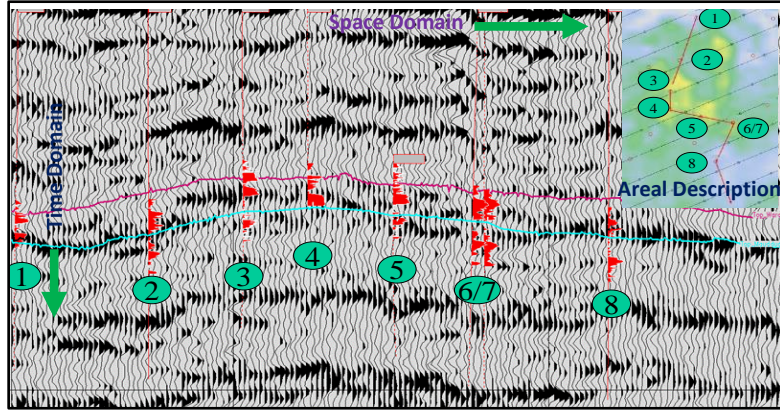


Fig. 3. Seismic *peak* and *trough* dimensions and their connectivity (time-domain ontologies)

## 5.2 Depth-Domain Ontologies

Similar knowledge based ontologies are described in depth-domain, in which different ranges 0-500m, 500-1000m, 1000-1500m and 1500-2000m are labeled (Fig. 4). Each range has different knowledge levels; for example, CBM occurs in the range of 0-500m, conventional non-associated gas from 500-1000m, conventional associated gas from 1000-1500m and tight gas sands beyond 2000m. Hydrocarbons associated within shale reservoir can be deeper (more than 3000m), where in, they experience matured temperatures for hydrocarbon generation. Extraction and interpretation of knowledge for deeper reserves of shale gas requires skill and effort and it is more expensive to exploit.

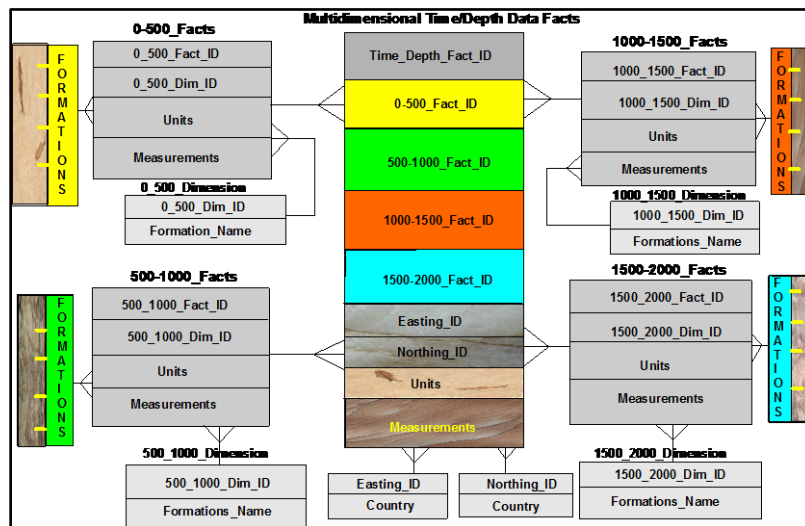


Fig. 4. Depth domain ontologies

### 5.3 Horizon-Based Ontologies

The seismic data possess multiple horizons with *peak* and *trough* dimensions in modeling either seismic times or computed depth structures. We believe a horizon (either seismic or geological) should never be viewed as a single entity or dimension nor isolated from multiple horizons at any stages of processing and or interpretation. Multiple horizons are ontologically interlinked or interconnected in an ecosystem scenario with conceptualized “velocity” attribute. Horizons in a sedimentary basin scale form a collective ecosystem, in which all the horizons interact and communicate (during several geological periods) among themselves through shared properties such as velocity, density including porosity or permeability of reservoirs. If there is any change in a single horizon, there is a corresponding change in other horizons (reasons could be structural connection and reservoir extensions in multiple horizons in 2D/3D seismic datasets) and their properties. In an analogy, multiple horizons may have been associated with multiple geological ages. A simulated fractured cross-section is shown in Fig. 5a. Each horizon is characterized by attribute dimensions of *structure*, *reservoir*, *seal* and *source* elements as illustrated in a dimensional model in Fig. 5b.

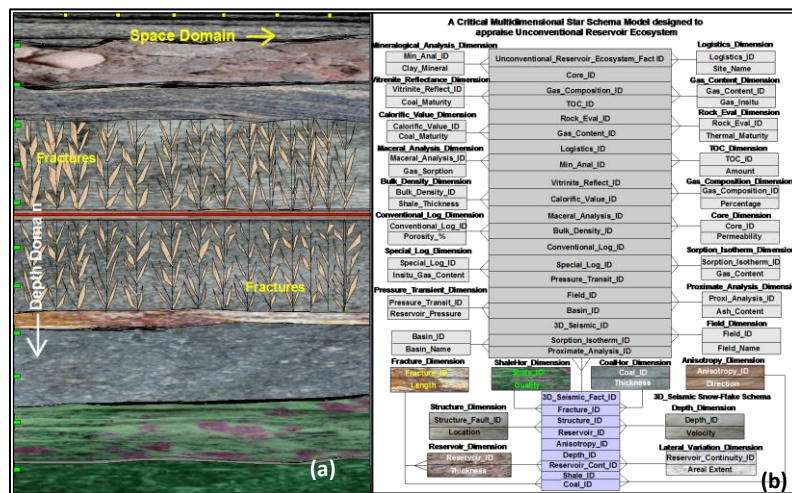


Fig. 5. (a) Fractured geological section (b) multidimensional data schema

One of the key multidimensional models representing the *shale-gas* exploration and development is presented in Fig. 5. Several interconnected dimensions which represent one-to-one, one-to-many and many-to-many data relationships in *exploration* domain, are logically (ontologically described) connected to their corresponding fact tables within a warehouse environment. For connecting elements of multiple petroleum systems, corresponding ontologies are articulated as described in the following sections.

**Source Ontologies.** The *source* rock is the main element in the ontology description. A rock rich in organic matter, if heated sufficiently can generate oil or gas. Typical source rocks, usually shales or limestones, contain about 1% organic matter and at least

0.5% total organic carbon (TOC), although a rich source rock might have as much as 10% organic matter. Preservation of organic matter without degradation is critical to creating a good source rock, and necessary for a complete petroleum system. Under right conditions, the source rocks can also be reservoir rocks, as in the case of shale gas exploration & production (E & P). Among several dimensions involved within source ontology, ontology description of maturity of source rock is a characteristic property of any petroleum ecosystem that can ascertain the potentiality of unconventional reservoirs.

**Structure Ontologies.** The *structure* and or entrapment of hydrocarbons [11] are other key elements of petroleum ecosystem. All the necessary inputs needed to explore the trapping mechanism are articulated from ontology descriptions through structural (geological) connectivity and corresponding data connectivity in the warehouse repository.

**Seal Ontologies.** The *seal* rock is another characteristic property of an entire petroleum ecosystem habitat. Unless an appropriate seal rock is interpreted on a regional scale, oil and gas field trapping mechanism cannot be well understood in its totality within an ecosystem setting. Accordingly the ontologies associated with seals described in [11] are integral part of the ecosystem to make connectivity between systems through seals.

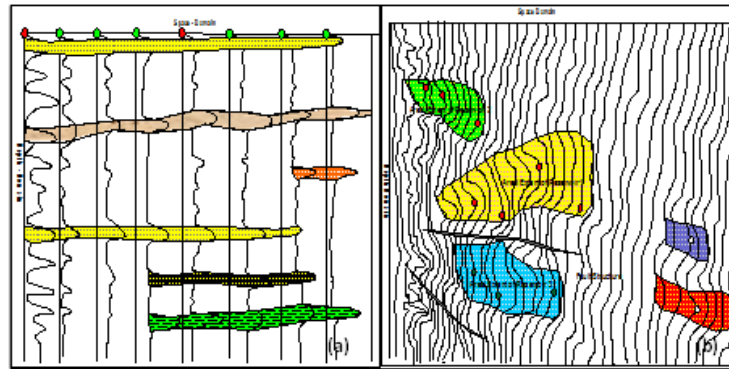
**Reservoir Ontologies.** A reservoir is one of the principal components of a petroleum ecosystem. The reservoir potentiality demonstrates large accumulations of hydrocarbons in a basin. Our primary focus is modeling the reservoir connections using an integrated ontology framework that supports the multidimensional warehouse modeling and mining [14] in Big Data scale.

## 6 Evaluation and Implementation of DSIS Artefacts

Like elements of an ecosystem, the existing processes also affect the potentiality of the E & P. The elements and processes are together explored for their connections to a new conceptualized element, termed as {chains}. “Chains” within an ecosystem are more conceptualized and they are based on cognitive interaction between elements and processes. A Chain can either be between individual elements and or processes of petroleum systems from groups of basins. The seismic data that make up the chains consist of numerous *peak* and *trough* dimensions [14, 15]; they are used for connecting horizons and their associated elements structure, reservoir, source and seal of a petroleum ecosystem. The data associated with elements and processes are digitally modeled, as sets using set theory; cardinalities of elements are thus described in several permutations and combinations [14]. Then each element is narrated with an ontological description to formulate the connectivity. Description of ontologies for each individual element is, though a primary focus, establishing the reservoir connections and integrating the peak and trough dimensions (of seismic data) within warehoused metadata are other foci as shown in Fig. 6. The map view drawn from the meta-metadata corroborating the



reservoir connections and exploring drillable prospects demonstrates the model validity in the upstream business contexts.



**Fig. 6.** Exploring reservoir connections in space and depth dimensions on well-log data (a) one dimensional map view and (b) 2D areal map view

## 7 Knowledge Management and Research Deliverables

The digital framework that characterizes the data connectivity between multiple domains and systems is a new innovative solution, integrating digital schemas in a unified multidimensional repository. The constructs, models, and methods are final deliverables of the current research application. Models and metadata cubes [7, 14] deduced from integrated framework (Fig. 2) facilitate us an effective and cognitive visualization and interpretation of digital petroleum ecosystems. The digital petroleum ecosystem framework ensures an interpretable metadata with predictive knowledge of unconventional petroleum systems as demonstrated in Fig. 6. Artefacts used in the DSIS are evaluated using various utility properties such as "data quality, data governance, interoperability, usefulness, ease of use and effectiveness", establishing the efficacy of DSIS and managing the unconventional digital petroleum ecosystems [14]. Overall, the DSIS has a vital role in making huge impacts in smarter integrated interpretation upstream projects, especially during *prospect identification* and *risk evaluation* stages.

## 8 Conclusions and Future Work

Big data characteristics may have inherited from volumes of unconventional petroleum data sources, meant for exploring connections between *shale gas*, *shale oil*, *coal bed methane* (CBM) and *tight oil* sands, all existing within a single sedimentary basin, an unconventional ecosystem. Unconventional data sources are heterogeneous and multi-dimensional. For data integration, ontology based data warehousing and mining is used. Besides, visualization and interpretation artifacts are part of knowledge-based smart digital ecosystem solutions. Ontologies written for multiple dimensions facilitate the connectivity among unconventional petroleum ecosystems. The Big Data technologies

are invaluable tools for investigating, examining and interpreting the sweet-spots, besides smart delivery of fine-grained metadata cubes of several unconventional reservoir ecosystems that validate the connections among multiple reservoir systems.

## References

1. Beaumont, E. A. and Foster, N. H.: Exploring for Oil & Gas Traps, *AAPG Treatise of Petroleum Geology*, Publications of Millennium Edition, Memoir 78, 2<sup>nd</sup> Edition, UK (1999)
2. Brown, D.: Looking deeper into fracture impacts, *AAPG Explorer*, March Archive, USA (2013)
3. Carvajal, G., Maucec, M. Cullick, S.: *Intelligent Digital Oil and Gas Fields*, Gulf Professional Publishing, 1<sup>st</sup> Edition, Elsevier, 374p, (2017)
4. Cleary, L, Freed, B, and Elke, P.: “Big Data Analytics Guide”, Published by SAP, 2012, CA 94607, USA (2012)
5. Li, G.: *World Atlas of Oil and Gas Basins*, John Wiley and Sons Ltd, 496p, USA, (2011)
6. Castaneda, G. OJ. Nimmagadda, S.L., Cardona, M. Lobo, A. and Darke, K.: On Integrated Quantitative Interpretative Workflows for interpreting structural and combinational traps for risk minimizing the exploratory and field development plans, presented and published in the *Bolivarian Geophysical Symposium* proceedings, held in Cartagena, Colombia (2012)
7. Coronel, C. and Morris, S.: *Database Systems: Design, Implementation, & Management*, Cengage Learning US, Edition 12, 784 p, (2016)
8. Durham, S.L.: An unconventional idea, open to interpretation, *AAPG Explorer*, March Series, USA (2013)
9. Indulska, M., and Recker, J.C.: Design Science in IS Research: A Literature Analysis. In GREGOR, SHIRELY and HO SUSANNA Eds. *Proceedings 4<sup>th</sup> Biennial ANU Workshop on Information Systems Foundations*, Canberra, Australia (2008)
10. Khatri, V. and Ram, S.: Augmenting a conceptual model with geo-spatiotemporal annotations, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16(11), pp. 1324-1338 (2004)
11. Magoom, L. B., and Dow, W. G.: *The Petroleum System from Source to Trap*, AAPG/Datapages, Digital Reprint of AAPG Memoir 60 (2009)
12. Nimmagadda, S.L. and Rudra, A.: Big Data Information Systems for Managing Embedded Digital Ecosystems (EDE), a book chapter in a book entitled “*Big Data and Learning Analytics in Higher Education: Current Theory and Practice*”, Springer International, DOI: 10.1007/978-3-319-06520-5, ISBN: 978-3-319-06519-9, The Netherlands (2016)
13. Nimmagadda, S. L. and Dreher, H.V.: On new emerging concepts of Petroleum Digital Ecosystem (PDE), *Journal WIREs Data Mining Knowledge Discovery*, 2: 457–475 doi: 10.1002/widm.1070 (2012)
14. Nimmagadda, S.L.: *Data Warehousing for Mining of Heterogeneous and Multidimensional Data Sources*, Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG, p. 1-657, Germany (2015)
15. Parasnis, D. S.: *Principles of Applied Geophysics*, Chapman & Hall, USA (1997)
16. Vaishnavi, V. and Kuechler, W. Jr.: *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, NY: Auerbach Publications, Boca Raton, FL, Taylor & Francis Group (2007)
17. Wight, A.W.R. Hare, L.H. and Reynolds, J.R.: *A Sedimentary Basin, NE Kalimantan, Indonesia: a century of exploration and future potential*, Geological Society of Malaysia, Circum – Pacific Council for Energy and Mineral Resources (1992)