Big Data guided Digital Petroleum Ecosystems for Visual Analytics and Knowledge Management

Shastri L Nimmagadda
School of Management
Perth, WA, Australia
Shastri.nimmagadda@curtin.edu.au

Andrew Ochan
Petroleum Authority of Uganda
Entebbe, Uganda

Neel Mani
Amity University
Noida, UP, India

Torsten Reiners
School of Management
Perth, WA, Australia

SUMMARY

The North West Shelf (NWS) interpreted as a Total Petroleum System (TPS), is Super Westralian Basin with active onshore and offshore basins through which shelf, - slope and deep-oceanic geological events are construed. In addition to their data associativity, TPS emerges with geographic connectivity through phenomena of digital petroleum ecosystem. The super basin has a multitude of sub-basins, each basin is associated with several petroleum systems and each system comprised of multiple oil and gas fields with either known or unknown areal extents. Such hierarchical ontologies make connections between attribute relationships of diverse petroleum systems. Besides, NWS has a scope of storing volumes of instances in a data-warehousing environment to analyse and motivate to create new business opportunities. Furthermore, the big exploration data, characterized as heterogeneous and multidimensional, can complicate the data integration process, precluding interpretation of data views, drawn from TPS metadata in new knowledge domains. The research objective is to develop an integrated framework that can unify the exploration and other interrelated multidisciplinary data into a holistic TPS metadata for visualization and valued interpretation. Petroleum digital ecosystem is prototyped as a digital oil field solution, with multitude of big data tools. Big data associated with elements and processes of petroleum systems are examined using prototype solutions. With conceptual framework of Digital Petroleum Ecosystems and Technologies (DPEST), we manage the interconnectivity between diverse petroleum systems and their linked basins. The ontology-based data warehousing and mining articulations ascertain the collaboration through data artefacts, the coexistence between different petroleum systems and their linked oil and gas fields that benefit the explorers. The connectivity between systems further facilitates us with presentable exploration data views, improvising visualization and interpretation. The metadata with meta-knowledge in diverse knowledge domains of digital petroleum ecosystems ensures the quality of untapped reservoirs and their associativity between Westralian basins.

Key words: Northwest Shelf; Digital Petroleum Ecosystems; Total Petroleum System

INTRODUCTION

We examine existing issues relevant to the exploitation of structural and strati-structural plays under different geological settings (Gilbert et al. 2004). The risks of exploratory drilling in under-explored areas are given in Longley et al. (2001). Data integration and connectivity between petroleum systems are key challenges for which new applications are needed. Aside from brief discussions on data structuring of the resources data, there is limited literature on ontology applications in oil and gas industries, especially the implementation of Digital Petroleum Ecosystem (DPE) in the oil and gas industry. The purpose of the research is to investigate the scope of data integration and its intricacy involved in the big data-guided digital petroleum ecosystems (Canstaneda et al. 2012). The sedimentary basin or group of basins in the NWS contexts are simulated as a digital petroleum ecosystem (Nimmagadda and Dreher, 2012). Data modelling and exploration metadata descriptions of Total Petroleum System (TPS) are decisive in the DPE perspectives. Heterogeneity and multidimensionality, including the unstructured nature of data sources, are other challenges, can preclude the data integration process.

Additionally, the research in ecosystem data mapping and modelling needs a new direction. The authors interpret the DPE community as a function of the domain (context), data integrity, repository accessibility and security, knowledge cogency and adequacy, including business viability attributes. In the DPE viewpoint, sustainability is a composite attribute and function of the strength of artefacts (domain, data model, warehousing and mining, visualization and interpretation, new knowledge and business value). Exploring the connections between common attributes (including conceptualised and contextualised attributes) in the DPE domains has many benefits, though it is challenging. Data sources pertained to the elements and processes of petroleum systems, are available in the open- and shared-sources, but need organization for the integration process. Organizing, structuring, and integrating the existing data sources in multiple domains of multiple petroleum ecosystems are key features of the research as envisaged in conventional approaches (Nimmagadda et al. 2018). The systematic investigation and management of data sources using integrated and interpretative workstation tools have motivated us to examine the pitfalls of the interpretation and the integration process in multiple petroleum exploration industries. Besides, an opportunity exists for cross-domain research in connecting diverse domains through domain ontologies in the petroleum data modelling (Rudra and Nimmagadda, 2005). Interpretation of data patterns, correlations and trends perceived from exploration metadata has led us to do more rigorous-research, particularly on big data feasibility through a sustainable integrated framework.

In addition, in big data contexts, the data visualization (Nimmagadda and Rudra, 2016 and Nimmagadda and Rudra, 2017) has significance in envisaging the exploration data and their visual and graphical depiction for meaningful geological interpretation. Quality information delivery is aimed at through effective graphical means. In such contexts, the authors explore the concept of data fusion, a kind of visualization, with which interpretable exploration data are made meaningful for geological interpretations. In other words, from different graphic visuals, the knowledge can be extractable to interpret exploration metadata. The designers in Post et al. (2002) and

Nimmagadda and Rudra (2017) focus on visuals and their functionality, ensuring new insights of information, especially in sparse and complex data areas, in a more intuitive way. Without losing the clarity and perception, we attain the knowledge through visual representations and graphic displays of geological and geophysical (G & G) entities, including knowledge of exploration and production (E & P) dimensions. The knowledge-based ontology structures provide additional support on building integrated framework architectures (Nimmagadda and Rudra, 2016). More details on various visualization techniques to interpret multiple datasets are given in Lugmayr et al. (2017). Different data-association mining rules are designed to articulate the data structures and work as per business rules. An application of data visualization technique existing in oil and gas exploration is given in Nimmagadda et al. (2012). The visualization of petroleum prospects is described in Australian sedimentary basin contexts with 4-D seismic technologies (Gilbert et al. 2004). We present the big data paradigm, its feasibility and applicability as an emerging think-tank of the DPE in this context. The volume and variety characteristics play a dominant role; in particular, the data views from warehouse metadata are more explicit through visualization and interpretation artefacts.

METHOD AND RESULTS

Big Data Systems: The oil and gas industry has not fully exploited the use of big data systems, their applicability and feasibility (Agarwal and Dhar, 2014 and Chen et al. 2012). For managing the size and magnitude of data (including geographical and periodical extents) and their sources, an integrated system is necessary to leverage big-size heterogeneous and multidimensional oil and gas data sources. For example, hundreds of square kilometres of onshore and offshore data sources are widely spread in the North West Shelf (NWS) that comprises hundreds of trillion bytes of data. An integrated framework is needed to make fine-grained data structures for effective data mining, visualization, and interpretation and manage the heterogeneity and multidimensionality data. Big data systems emerge to offer new knowledge representations in multiple domains and explore multiple connections among large size data sources of NWS. We develop "Ecosystem and embedded ecosystem" ideas in sedimentary basins contexts in areas where no geological boundaries exist (reference). We explore the connectivity between sedimentary basins through various artefacts and their collaborations (Figure 1).
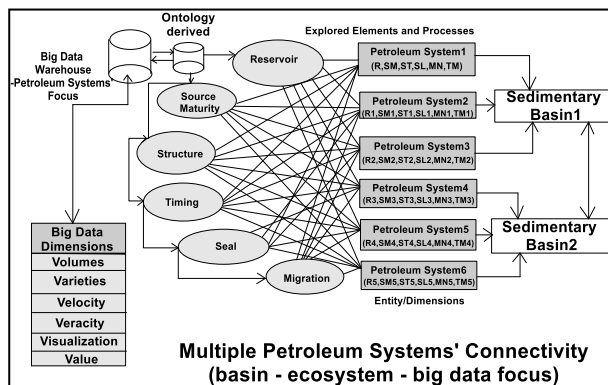

Figure 1. Connecting multiple petroleum ecosystems

In addition, big data systems rely on evaluable data models and implement their interpretative analysis of domain knowledge in upstream business environments. One of the primary goals of examining big data analytics is to discover exploration data

patterns and their intelligence for interpretation. It is generally an accepted notion that unstructured data, mostly located in text files, account for at least 80% of an organization's data. If left unmanaged, the sheer volume of unstructured data generated each year within a company can be costly in storage.

As illustrated in Figure 2, anomalies may arise while incorporating and collaborating with big data characteristics and their modelling, including their data analytics in upstream businesses. Big data analytics is often associated with cloud computing because the analysis of large data sets in real-time requires a framework like *MapReduce* to distribute the work among thousands of computers. Big data analytics is crucial to help interpreters in operating companies and make better technical and business decisions. The users analyze huge volumes of technical and commercial data that may have left untapped by conventional business intelligence (BI). The data modelling methodologies that support the big data tools and technologies are given in Figure 2.
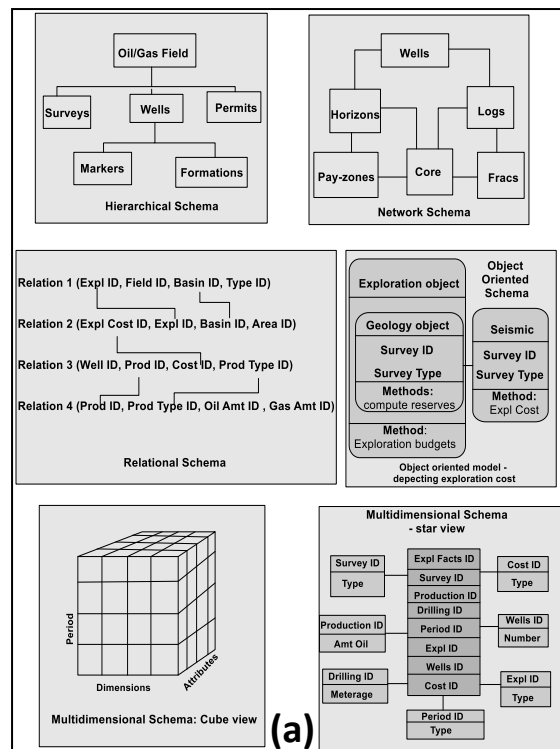

Figure 2. Various data artefacts used in the study

Big data analytics (Cleary et al. 2012) is done with software tools commonly used in advanced analytics disciplines such as predictive analytics and data mining. However, the unstructured data sources used for big data analytics may not match with traditional data warehouses. Potential pitfalls can give rise to organization's big data analytics initiatives. They include a lack of internal analytics skills and the high cost of hiring experienced analytics professionals, with added challenges of integrating Hadoop systems and data warehouses. However, vendors are starting to offer software connectors between those technologies and integrated platforms (Post et al. 2002).

Digital Ecosystems: It is a system with a capacity to manage information and build relationships between entities, dimensions and or object classifications of elements and processes of petroleum systems. The scope of developing the visualization models is explored from digital ecosystems of petroleum bearing basins. Our focus is on conventional, unconventional and matured oil and gas fields of these basins,

exploring the scope of interpretation and analysis by multidimensional metadata, including their mining with visualization tools. Schemas drawn for E & P in PDE connect multiple attributes relevant to seismic surveys and drilled-well domains (Figure 3) to arrive at an integrated metadata structure.
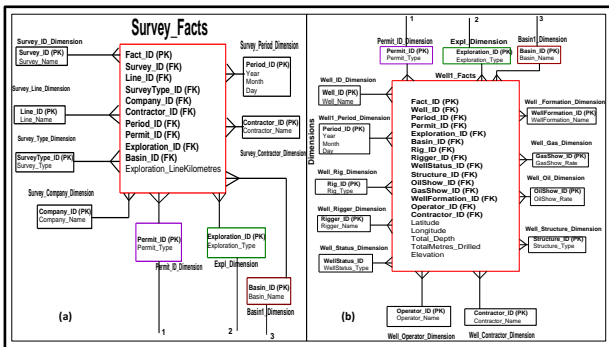

Figure 3. Couple of schemas drawn to connect multidimensional data attributes in the PDE
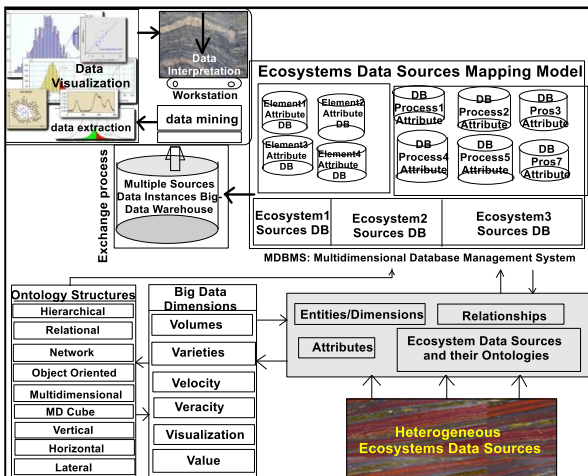

Figure 4. Design of an integrated framework

Australian sedimentary basins (NWS): There is an immense scope of integrating and analysing many sedimentary basins in onshore and offshore areas proposed methodologies in the NWS. Sedimentary basins possess heterogeneous and multidimensional data sources described in various digital formats. Integrated framework and workflows (Figure 4) explore the use of exploration and production datasets for risk minimizing the oil and gas business in Australia. Super Westralian basin, a total petroleum system (TPS), such as North West Shelf (NWS), possesses the shelf, slope and deep geological events, which appear to have connectivity through phenomena of a digital ecosystem (Longley et al. 2001). Besides, this super basin has many sub-basins; each basin is associated with multiple petroleum systems and each system with unknown or limited areal extents. Each petroleum system contains various oil and gas fields, with the hierarchical structuring of data dimensions and their associated attributes. North West Shelf in Western Australia has a scope of unstructured heterogeneous and multidimensional data (Nimmagadda et al. 2012); we explore the scope of analysing their dimensions, attributes and instances in a warehousing environment with data mining and visualization opportunities exploiting untapped reservoirs of these basins.

The major challenge is the interpretation of connectivity between petroleum systems to analyse data patterns between attribute instances of similar and or dissimilar domains.

However, visualization can explicitly envisage the hidden knowledge from data patterns to add depth of information in the form of new interpretable domain knowledge. Big data collection covers both longitudinal (temporal) and lateral (geographic) design aspects. The data visualisation is one of the packages of innovative artefacts of Big Data paradigm, aiming:

1. Presenting business scenarios in unified metadata views.
2. Slicing and dicing of warehoused metadata cubes in spatial-temporal attribute dimensions.
3. Interpretation of visual data and map trends in new knowledge domains.
4. Adding values to digital business ecosystem products and services

We interpret the extracted data-views for evaluating the effectiveness of the integrated framework and data models designed in different knowledge domains and contexts. As presented in Clearly et al. (2012), cognitive Big Data is a new data interpretative research tool with several business contexts. The rationale behind this approach is to bring together and integrate different contexts. Data analysis is crucial, in addition to testing the validity of data models, data warehousing and mining and the effectiveness of visualization. Qualitatively, the trends, patterns and correlations observed among various data events are interpreted in the knowledge enhancement domain (Nimmagadda et al. 2012). Extent and duration of usage of data models and integrated framework including implementation of contextual, short- and long-term research outcomes among latitude and longitude dimensions are other interpretation objectives. Data analysis and interpretation are meant for transforming the processed data into critical knowledge guaranteeing the research outcomes for descriptive analysis. The measure, consistency, and effectiveness of multidimensional and heterogeneous data organization, modelling, mapping, and data mining, including data visualisation effectiveness, are the other tests. Interpretation is either qualitative or quantitative, and the data patterns, trends and correlations are interpreted to discover new geological knowledge.

Mapping the domain knowledge: In the oil and gas industries, in particular exploration entities, "domain knowledge" is commonly used to interpret data of elements, processes and chains of petroleum systems and their ecosystems. Translating the existing knowledge with new knowledge using the new attribute dimensions interpreted in multiple domains is key. Several data attributes and instances of the geological structure, seismic data attributes are interpreted to identify new prospective areas. As a part of implementation in the present study, we propose several interpretation methodologies (Figure 5).
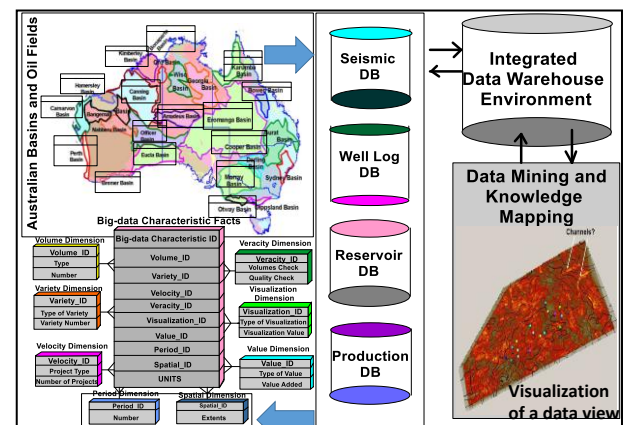

Figure 5. Big data implementation model

We use slicing and dicing data mining operations to extract data views, as shown in Figure 6.
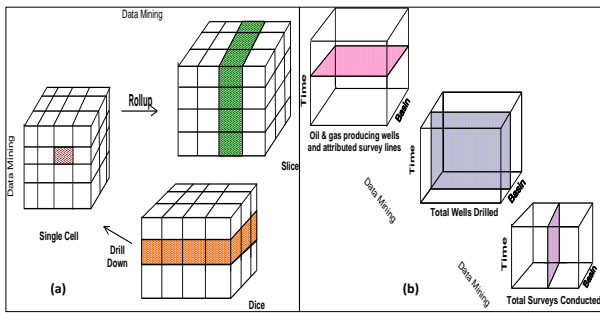

Figure 6. Slicing and dicing data mining operations

The Relative Advantage of PDE and Visual Analytics: To resolve data heterogeneity and multidimensionality in embedded systems, we propose an integrated framework within a conceptual sustainable digital ecosystem architecture. Multiple domains in coexistent systems are brought together to explore connections, interpret and evaluate diverse systems' commonality. Big Data characteristics – conceptualised as attribute dimensions are implemented with data marts (Figure 6) and improve the integration processes. To this extent, domain ontologies are built to iterate the structures and integrate them within a data warehouse environment (Figure 4). As shown in Figure 7, several multidimensional attributes are interpreted in cuboid data structures.
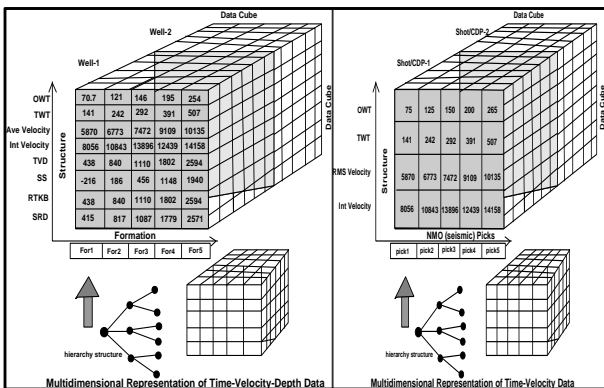

Figure 7. Multidimensional data structures in time-depth dimensions

We use several visual tools, providing an effective means of communication, with highly developed 2D and 3D pattern-recognition capabilities that allow processing and perceiving the pictorial digital data efficiently. By summarizing the data and highlighting visualization trends, we uncover unknown geological phenomena.

Several visualization techniques use volumes and varieties of Big Data, including spatial-temporal multidimensional datasets in the DPE contexts. We exploit the following visualization methods: One method of visualizing the results of data search is through scatter- plot display in a three-dimensional grid. A scatter plot is a visualization technique that shows each data point as a colour sphere or bubble. Scatter plots are referred to as bubbles in these visualizations. The size, shape and colour of the bubble each is used to represent a variable in the data. The three axes in the visualization ideally should be able to describe any dimension within the same data plot and should be able to randomly select and modify by the user. The user can choose to

explore more fully those data points using visual methods or perhaps click on the data and see a traditional numerical display in a spreadsheet. The data points can be rotated to observe clusters in different areas. By preparing and presenting the data graphically, the user can uncover properties of the data quickly, detecting any patterns or deviations from expected results.
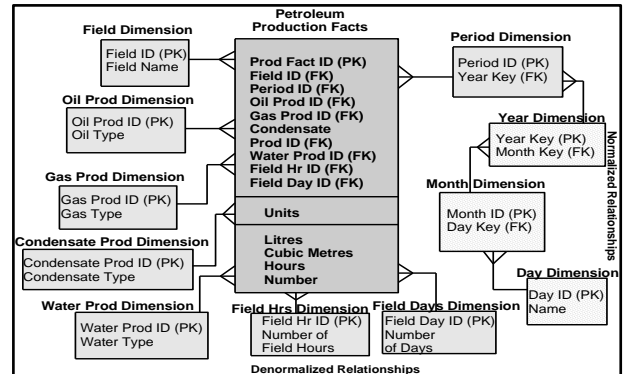

Figure 8. Snowflake data schema

The authors have added a snowflake schema with production data attributes to the warehouse architecture. Petroleum production-data dimensions and facts can represent in the snowflake star schemas, in which relationships between common attributes of each production entity are normalized as demonstrated in Figure 8. In the present context, to support attribute hierarchies, the dimension tables are normalized to create snowflake schemas. This consists of a single fact table and multiple dimension tables again. Like star schema, each tuple of the fact table consists of an attribute (foreign key) pointing to each of the dimension tables that provide its multidimensional coordinates. It also stores numerical values (non-dimensional attributes) for these coordinates. The advantage of this normalized schema is easy in its usage and maintenance; normalizing also saves storage space, though navigation of petroleum data across multiple tables may not be effective due to a large number of join operations.

Traditional data analysis tools are too rigid, and they have been domain-specific, constraining information flow between diverse domains and groups of ecosystem service providers with inactive reporting tools. We explore new visual analytics tools that enable us to explore multiple domains embedded in digital petroleum ecosystems in which hundreds of attributes and volumes of their instances are implanted. New insights are presented with ease and in real-time quickly to interpret interdependent and interconnected data relationships between petroleum systems. One of the slices presented in Figure 9 depicts the superimposition of structure, reservoir, production and months produced attributes. The structure has a good relationship with reservoir and production attributes. The data view is useful for making appraisals and financial decisions.
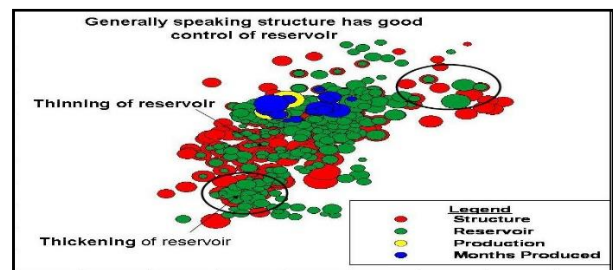

Figure 9. Exploration and production metadata view

Visual analytics can envisage the meta-knowledge from metadata structures to explore and exploit the metadata cubes and their data views for interpretation. The knowledge interpreted from metadata views can represent the information from multiple industry domains, usable and or reusable in the TPS contexts. The tools provide better collaboration among diverse domains (including domain experts), uncovering associations or data relationships from data systems that vary with geography, demography and even various business processes.

CONCLUSIONS

Ontology-based data modelling, data warehousing, mining, visualization and data interpretation, combined in a single canvas in an integrated framework, is advantageous in designing and developing digital ecosystems in Big Data contexts. Digital data of a petroleum system or groups of petroleum systems existing in a sedimentary basin, all simulated within an integrated warehouse scenario, is considered as a digital oil field solution. How the elements, processes, and chains (of an information system or a petroleum system) interacted and communicated with each other among multiple systems is another innovative concept in designing and developing the digital petroleum ecosystem. The digital petroleum ecosystem and digital oil field solutions are research outcomes that can change the economics of exploratory drilling campaigns. We conclude with the following inferences, based on the visualization and interpretation of various data views of the PDE and TPS metadata:

1. Heterogeneous and multidimensional data sources existing in government and private enterprises, national and multinational companies can successfully build multidimensional models.
2. Data visualization is a successful technology for examining the resources data hidden under great depths. Data mining is an iterative process, implying that a better visualisation attribute is perceived each time it refines the resultant data.
3. Interpretation models help managers of the resources industries in predicting the drillable exploratory locales.
4. The authors recommend using data warehousing and mining technologies together with visualization and interpretation artefacts, supported by Big Data. The map views are valuable tools for interpretation and implementation in the upstream business.
5. SQL and classical statistical mining are quite useful for mining and visualizing multidimensional data, including geographical and periodic dimensions.
6. Petroleum digital ecosystems (PDE) and petroleum information systems are digital oil field solutions in Big Data scales for various producing companies and multinational service companies.

REFERENCES

Agarwal, R. and Dhar, V. 2014. "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," Information Systems Research (25:3), pp. 443-448.

Castaneda G. O. J., Nimmagadda, S.L, Cardona Mora, A. P, Lobo, A, and Darke, K., 2012, On Integrated Quantitative Interpretative Workflows for interpreting structural and combinational traps for risk minimizing the exploratory and field development plans, Bolivarian Geophysical Symposium proceedings, held in Cartagena, Colombia.

Chen, H., Chiang, R. H. & Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact," MIS quarterly (36:4), pp. 1165-1188.

Cleary, L, Freed, B, and Elke, P. (2012), Big Data Analytics Guide: 2012, Published by SAP, CA 94607, USA.

Gilbert, R. Liu, Y. and Abriel, W. (2004), Reservoir modeling: integrating various data at appropriate scales, *The Leading Edge*, Vol. 23(8) (pp. 784-788), EAGE, The Netherlands.

Longley, I.M. Bradshaw, M.T. & Hebberger, J. (2001), Australian petroleum provinces of the 21st century, in Downey, M.W. Threet, J.C. & Morgan, W.A (2001) Petroleum provinces of the 21st century, AAPG Memoir, 74, 287-317.

Lugmayr, A. Stockleben, B. Scheib, C. M. Mailaparampil, M. Mesia, N. and Đanta, H. (2016), "A Comprehensive Survey on Big Data Research and It's Implications - What is really 'new' in Big Data? It's Cognitive Big Data," Proceedings of the 20th Pacific Asian Conference on Information Systems (PACIS 2016), S.-I.C. Shin-Yuan Hung Patrick Y.K. Chau TingPeng Liang, ed.

Nimmagadda, S.L, Dreher, H, Noventianto. A, Mustofa. A and Fiume. G., 2012, Enhancing the process of knowledge discovery from integrated geophysical databases using geo-ontologies, a paper presented and published in the proceedings of Indonesian Petroleum Association (IPA) conference, held in Jakarta, Indonesia.

Nimmagadda, S. L. and Dreher, H. (2012), "On new emerging concepts of Petroleum Digital Ecosystem (PDE)", Journal WIREs Data Mining Knowledge Discovery, 2012, 2: 457–475 doi: 10.1002/widm.1070.

Nimmagadda, S.L. and Rudra, A. (2016), Big Data Information Systems for Managing Embedded Digital Ecosystems (EDE), 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: 978-3-319-065199, The Netherlands.

Nimmagadda, S. L. and Rudra, A. (2017), Visualization as a Big Data Artefact for Knowledge Interpretation of Digital Petroleum Ecosystems. International SERIES on Information Systems and Management in Creative eMedia (CreMedia), [S.l.], n. 2016/2, p. 34-43, June 2017.

Post, H.F., Gregory, M. N. and Bonneu, G. (2002), Data Visualization: The State of the Art, Research Paper TU Delft, EUROGRAPHICS 2002, The Netherlands.

Rudra, A. and Nimmagadda, S.L. (2005), Roles of multidimensionality and granularity in data mining of warehoused Australian resources data, *Proceedings of the 38th Hawaii International Conference on Information System Sciences*, Hawaii, USA.