Sedimentary Basin - a Petroleum Digital Ecosystem

Shastri L Nimmagadda
PTS, Schlumberger, Moscow, Russia
(snimmagadda@slb.com)

Heinz Dreher
Curtin University, WA, Perth, Australia
(heinz.dreher@cbs.curtin.edu.au)

Andi Noventianto
MEDCO Energy, Jakarta, Indonesia

Aswin Mustoffa
MEDCO Energy, Jakarta, Indonesia

Parapaty Halley
MEDCO Energy, Jakarta, Indonesia

Abstract**—** Existence of petroleum system and its elements is often narrated for each oil and gas field and for each petroleum-bearing sedimentary basin. Sedimentary basin is an emerging digital ecosystem within a generic petroleum system. On the broader scale, total petroleum system (TPS, where geology has no boundaries) can virtually be envisaged as a digital petroleum ecosystem. The significance of this concept lies in connecting onshore and offshore petroleum subsystems, in a basin (e.g., Indonesian Onshore-Offshore basin, a 19,300 mi^2 [50,000 –km^2] transition zone), through elements and process attributes and by modeling sub-basins thru contextualization and specification concepts.

In this paper, we present an ontology based data warehousing and mining technology in which, conceptualization and contextualization of multiple data dimensions are modeled within a sedimentary basin (e.g., elements and processes of all petroleum systems, existing within basin). In addition to their integration within a data warehouse environment, data mining and visualization of interpretable data views that emerge from petroleum digital ecosystems are made feasible. Multidimensional data warehousing and mining models facilitate an effective interpretation of petroleum ecosystems, minimizing the ambiguities involved during knowledge mapping of geological *structure* and *reservoir* qualities - not only for exploration and field development planning, but also for reserve estimations.

***Index Words* — ontology, petroleum systems, data warehousing, mining, sedimentary basins data**

I. Introduction

Petroleum system is an information system, in which several elements and processes are described. These elements and processes are described as dimensions [5] from an ecosystem point of view. We, the authors, believe these dimensions constantly interact and communicate, sharing each of their properties [4]. In the context of a broader notion of sedimentary basin, integration of these entities or dimensions fit with the notion of data warehousing, with a true representation of metadata and integration of multiple petroleum systems. In our study, an attempt has been made to acquire multiple petroleum systems from different sedimentary basins [1] and integrate them, using the concept of data warehousing. Logical structuring of multiple data dimensions is a perquisite for an effective data integration that facilitates the data mining process including visualization and interpretation. Several data mining views drawn from the data warehouse are used for interpreting domain ontologies, such as conceptualized and contextualized relationships among dimensions *structure*, *reservoir*, *source* and *seal* elements in the a sedimentary basin [1] and [14]. Unless the phenomenon of integration of elements and processes attributes of petroleum systems of sub-basins is understood, narration of petroleum ecosystem (both at "field" and "basin" scales) and its potentiality, including the characterization of "interconnectivity" among ecosystems, cannot be well explained.

A. Petroleum Ecosystems

An ecosystem is a system whose members benefit from each other's participation via symbiotic relationships [3], [7] and positive sum relationships. In the context of sedimentary basin research, modeling, petroleum systems analysis and narrating several petroleum systems, in a Total Petroleum System (TPS) scenario, are inevitable. Each oil and gas field in each petroleum system is a complex community, but its environment is functioning as an ecological unit. More realistically, a single repository may have hundreds of databases and associated data attributes that describe multiple sedimentary basins. Each basin may have one or more number of petroleum systems.

B. Ontology as a specification mechanism, describing petroleum ecosystem

A sedimentary basin is formally represented as knowledge-domain ecosystem, based on a *conceptualization* [3]: the *structure*, *reservoir*, *source, seal* and other *maturity* and *migration path*way process dimensions that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization may be an abstract, simplified view of the basin that we wish to represent for some purpose. Every knowledge-based system or knowledge-level agent is committed to have contextualization and conceptualization, either explicitly or implicitly.

978-1-4799-0752-6/13/$31.00 ©2013 IEEE

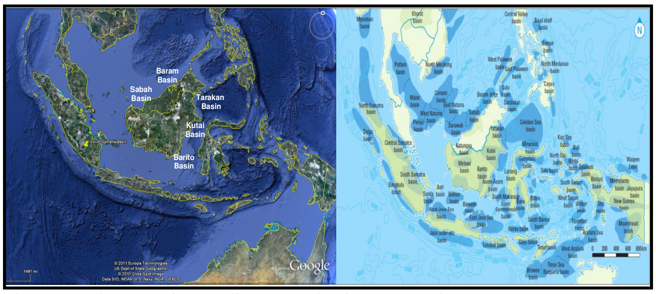

Fig.1: Southeast Asian Sedimentary basins ([1] and [14])

In the context of Southeast Asia (Fig.1), Indonesia possesses several investigating basins and petroleum systems. A sedimentary basin, which is located in the northeast of Kalimantan, its associated petroleum systems, interpreted to have multiple hydrocarbon plays and prospects, have complex known relationships including many that are unknown and contain hidden data relationships. Multiple dimensions or entities are interpreted among these petroleum systems and each dimension has multiple data attributes and strengths. Ontology description is meant for domain knowledge and domain knowledge is represented in a declarative formalism. In broad sense of a *digital basin*, representable sets of dimensions, digital fields and digital petroleum system are narrated. Ontology of a petroleum system is developed, which is represented by systematic existence, logic and intelligent design of data structures. Relationships are built among *structure*, *reservoir*, *source rock*, *seal rock* elements and processes such as, *timing* of formation of elements and *migratory* pathways with several logical conceptualizations and contextualization. The connectivity and intelligent interaction, communication among these element-dimensions, are well described via ontological commitments. In case of sedimentary basin, several oil and gas fields discovered by multiple seismic vintages have multiple dimensions such as 2D and 3D seismic datasets including drilled-data (more characteristically one dimensional). Each dimension, either in exploration or production datasets, commits to ontology if its observable actions are consistent with the definitions and designs of the ontology including imposition of business rules or constraints. The idea of ontological commitment is based on the knowledge-level perspective. The knowledge level is a level of description of the knowledge of a dimension that is independent of symbol-level representation, used internally by the dimensions, existing inherently. Knowledge is attributed to dimensions by observing their actions, a dimension ‘knows’ something if it acts as if it had information and is acting rationally to achieve its goals. The actions of dimensions, including knowledge-based servers and knowledge based-systems, can be seen through a ‘tell and ask’ functional interface ([3], [6] and [7]), where a client interacts with a dimension by making logical assertions (tell), and posing queries (ask).

II. ISSUES AND PROBLEM STATEMENT

A petroleum system is an information system, similar to any other information systems, described in different domains of different applications. It is significant to demonstrate this fact through concept of “integration of digital ecosystems at sedimentary basin scale”. Exploration of a sedimentary basin [1], [14] is better described by means of an interpretation of stratigraphic sequence and structural styles of sedimentary rocks. Several relationships exist among elements of a sedimentary basin, which are known (relating the geological *structure*-with *reservoir*) and unknown or they may be hidden among several petroleum system elements and processes. Establishment of a concept to any implementation is an important issue, in assessing the feasibility and economic value of a petroleum system. Another issue is description of multiple dimensions and their associated cardinalities that make up petroleum ecosystems that exist in a sedimentary basin.

Petroleum system requires timely convergence of certain geologic elements, processes and events essential to the formation and entrapment of the hydrocarbon deposits. As stated earlier, a petroleum basin (or province) is a geologic entity containing at least one or more proven petroleum systems. The concept of petroleum system in prospect-contextualization domains, in widely distributed sedimentary volumes may contain pools, describing:

- Reservoirs of similar or dissimilar productive geologic sequence (or seismic sequences)
- Similar or dissimilar chemical compositions
- Similar or dissimilar trap types
- Similar or dissimilar sources charging the reservoirs along similar or dissimilar migratory pathways
- A common and regional seal for reservoirs

Each petroleum system is genetically classified [3, 8] in terms of several processing factors, such as *charge*, *migration*, *drainage* and *entrapment style*. Each charge factor again, is critically categorized in terms of *super-charged*, *normally-charged* and *undercharged*. Under migration process category, hydrocarbons are either vertically and or laterally charged. Under entrapment category, hydrocarbons holding reservoirs could be high or low impedance. In the case of a sedimentary basin, there could be multiple depositional centres, intra-formational seals, multiple trap-types and migratory pathways, covering the entire sedimentary basin. Petroleum system elements and processes in a sedimentary basin are interrelated and are continuously networking with each other in geological ages. As a matter fact, authors conceptualize the relationships among data structures for effective data integration process. An effective way of interpreting economically viable petroleum system of a sedimentary basin is to integrate the tectonic framework, sequence stratigraphy, geologic history, thermal history along with sedimentary basin analysis and modelling. A sedimentary basin analysis comprises investigation of a sediment distribution in patterns, depositions at different geological ages including formation and deformations of structures. Basin modelling reveals how different elements/processes, or in combination, form part of integrated hypothesis.

In case of sedimentary basin, to better describe petroleum ontology, hierarchical, relational and networking relationships are constructed among petroleum system elements and levels of investigations. Elements and processes of petroleum systems are either hierarchically and or relationally linked among attributes and characteristics of these attributes.

III. METHODOLOGIES

Tools and concepts used for designing and developing data-warehouse for bio-informatics [11], [13] support our ideas of data integration process of system elements. All the elements and processes of a sedimentary basin are visualized via multi-dimensional metadata, representing several attributes and their characteristics. The data warehouse approach brings together petroleum systems' data from different oil and gas fields from sedimentary basins' different depositional and geological regimes. In Petroleum Digital Ecosystem scenario, data from geological, geophysical and geochemical domains are integrated in a data warehouses environment. The data warehouse approach [13], [2], [11] is used to benchmark and track the effectiveness of petroleum system productivity over space and time dimensions. It also allows processed (knowledge based) data shared among professionals and geographically distributed worldwide. The need to integrate petroleum systems data from multiple systems and sources is well known [4], [6]. It is important that data warehouse designers define the scope, depth, comparability and accuracy of data entering the warehouse. The scope of data refers petroleum systems data, sedimentary basins data, and geological, geophysical and geo-chemical data from multiple periods (time-dimension), and geographic locations (space-dimension). Depth of data refers the level of details obtained. To be comparable, data from multiple dimensions and different sites should adopt the same classifications, as much as possible. No matter how differently data are collected across sites, they are significantly altered for integration before moving into the data warehouse environment. To reduce the burden of alteration, it is important for petroleum systems analysts and geo-modellers to use compatible software systems to acquire and send data to the repositories. It is also important to standardize the data collection processes. Accuracy of data is desired for all types of data in any given situation and this is a fundamental requirement for reliable use of data. A sedimentary basin is typical example; in which varied data dimensions are embedded, such as *seismic*, *drilled-well*, *petro-physical and production datasets*.

Several data models are deduced that represent ontology, narrating relationships among petroleum system elements. Data structures or schemas are modelled in different star-schemas. In these data schemas, there are multiple dimensions narrated and interpreted conceptually with multiple relationships among several dimensions (Fig. 4) and physically with fact data tables. Ontology models are constructed for shallow marine and deltaic petroleum systems, in which several dimensions are described, connecting the factual data with one-to-one, one-to-many and many-to-many data relationships of dimensions. These models are represented in star-schemas, as described in the following sections.

IV. DATA MODELING APPROACHES

In a large-scale sedimentary basin, the data schemas are simple and flexible enough to modify as per the geological situations and users requirements, despite complex geological situations. If the data schemas are too large, complexity is a significant barrier to widespread adoption of the warehouse technology, because users will find the schema so difficult to understand that they will be unable to write queries and application programs. Schemas are evolved to grow and support new data types. Limiting the scalability to more data sources has definite advantage, but in petroleum exploration industry, data structures are typically very large in size and multidimensional.

More logical data structuring is to define single common tables for information that is common to many warehouse data types, to decrease the schema complexity. For example, for many basins and petroleum systems, there may be common *reservoirs*, *structure* styles, *source* and *seal* types [8]. Each can be implemented through a single common table. To this extent, an associative dimension or entity needs to be defined among common *structure, reservoir, source, seal* and depositional systems. Dimensions, representing spaces, such as *unique ID* and its *type* are required, for constructing the schema. Their associations are uniquely identified in the warehouse schema. Different dimension identifiers may have the same Fact ID. Thus a simpler approach is used in which all data warehouse dimensions share a single space of dimension identifiers within a warehouse instance. The identifiers are known as warehouse identifiers and are integers that are assigned at database load time. The schemas designed for a warehouse must have a support with concurrent presence, accessibility and addressability of multiple datasets, multiple versions of a given database [5], [10], [11], within a petroleum systems warehouse instance.

The warehouse schema should facilitate coercion of different sources of the same type of data into a common semantic framework. If there is no common information of attribute or data property among data dimensions, then different databases exist side by side within a warehouse environment, without having merged. Once common data property information exists, databases get merged. Star-schema models derived in Fig. 2 may be merged with logical key attributes among the participating dimensions in the structuring process.

A. Data Loader Requirements

Because basins and petroleum systems are in large size (to the order of 19,300 mi^2 [50,000 $-km^2$] or terabytes in storage) and their associated data are often large. Poorly defined syntax and frequent load failures could result crashing of loaders. For this reason, database loaders should be able to recover gracefully from errors encountered during parsing of their input files. The loaders are designed to keep loading even in the presence of an error. If partial data has been inserted into the warehouse, a load error maintained on the related dimension is updated to indicate that an error occurred while parsing the dimension and that the warehouse entry for the dimension may therefore be incomplete or contain errors.

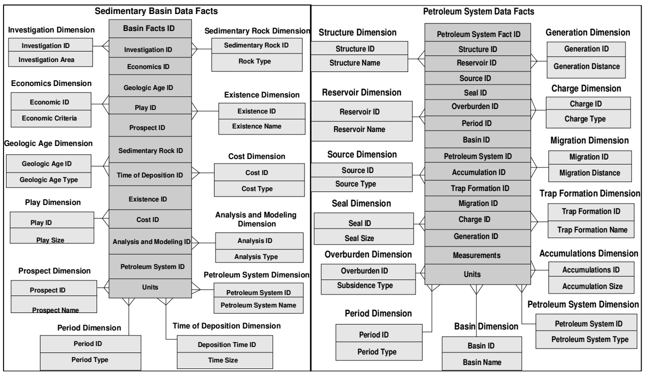

Fig. 2: Star-schema models representing relationships and their connectivity among various attributes of petroleum and depositional systems of a sedimentary basin

B. A Sedimentary Basin -Petroleum Systems Warehouse Design

Warehouse schema is designed by first analysing and testing the schemas of each database or super-type dimension to be integrated [10], [11], as well as the schemas of other databases that use the same data type and semantics. The development of warehouse is guided by several principle database schemas for sub-sets, each having similar data type or property.

Since databases typically conceptualize elements of petroleum systems in different ways, any kind of cross-database operation faces the problem of semantic heterogeneity, whereby information partitioned in different fields use different definitions (such as different units of measure). One possible approach to supporting petroleum system databases within warehouse would be to create different schema definitions for each of the conceptualization (domain ontology) of petroleum systems used by the source databases. In case of a sedimentary basin, sub-basins are designed in different domains of petroleum systems while respecting the semantic heterogeneity. Differences among multiple sources and databases are sorted out without complicating the resulting queries to integrated schema of petroleum systems' warehouse. As narrated in Fig. 3, an integrated framework is designed, explaining data structuring and integration process.

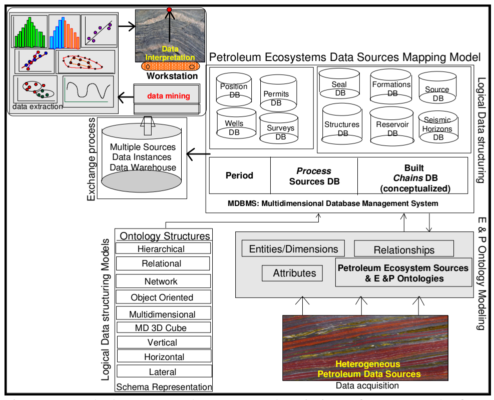

Fig. 3: An integrated data warehouse and mining framework for narrating a sedimentary petroleum ecosystem

A typical hierarchical structuring may be useful in deducing relationships not only connecting different data dimensions within a petroleum system, but also among different petroleum systems. By and large, this warehouse may typically contain a different petroleum system for every petroleum system database that it loads. Warehouse schemas are often large and hence queries made to user databases are not explicit. In such cases, users may write separate sub-query for each sub-schema.

In our approach, instead, a single set of schema definition is used and covers a given data type, even if that data type is ontologically conceptualized differently in different databases. For example, a single set of schema definitions is created to span all attributes for petroleum systems that possess similar *structure*, *reservoir* and *source* characteristics of producing horizons. Another example may be a single set of schema definitions for petroleum systems that have similar geo-chemical compositions of hydrocarbons. The database loaders are responsible for translating from the ontology conceptualization (specifications) used in each database family to the domain conceptualization used by the warehouse. This approach eliminates the semantic heterogeneity of multiple databases, allowing users to query all petroleum systems databases using the same schema. This approach ensures encoding of the dataset from which each data dimension within the warehouse is derived. Since entries from any petroleum system databases are loaded into the same set of tables (multi-dimensional schema definitions), it is critical for user queries (data views for interpretation) to be able to distinguish different petroleum bearing basins in Southeast Asia, Middle East, or basins in the Western Hemisphere. Thus, queries are made to extract data views for all warehoused fluvial, shallow marine and or deltaic depositional system databases that were loaded.

V. RESULTS AND DISCUSSIONS

Secondary data published by Indonesian Petroleum Association ([3], [1], [14], IPA) are used for testing the data models in the current study. Petroleum multidimensional star schemas are integrated in a single metadata structure of a sedimentary petroleum system warehouse. This type of multidimensional data warehouse respects and maintains the integrity of individual datasets, thus preserving information and validating the source of warehouse entries; so that users know and determine exactly, which depositional systems (knowledge-based) are parts of a particular set of petroleum systems. Different data views are computed for different sub-basins of a broad sedimentary basin (TPS, as the case may be with Southeast Asian basins) in similar and dissimilar depositional environments of similar and dissimilar petroleum systems. Ontology based multidimensional approach removes the semantic heterogeneity among petroleum systems and basins, removing redundancy among datasets including conflicts, if any arise during the description of the meaning of conceptualized attributes or dimensions. This approach also allows detecting the non-redundant petroleum system databases, for which different algorithms are written for mining purposes. Non-redundant petroleum system database may be created as a separate schema and database within a warehouse, which can apply to a specific mining algorithmic application or a specific user query. This could satisfy the users, who want to study large size non-redundant petroleum systems databases. As shown in Fig. 4, such cubes are generated that represent petroleum system databases, describing pool of sedimentary basins.

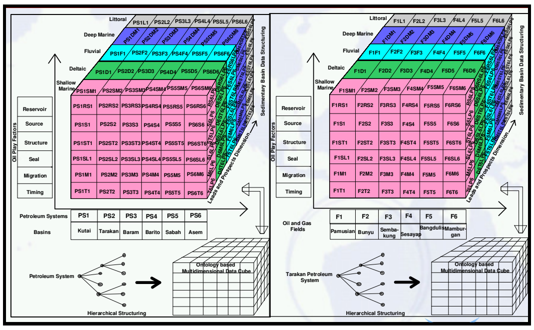

Fig. 4: Multidimensional representation of data cubes for multiple oil and gas fields and sedimentary basins of Indonesia

A Schema Implementation and Data Mining Approaches

Petroleum system warehouses support several petroleum ecosystem informatics data types, each of which is implemented as one or more tables in the schemas (dimensional and factual data tables are not shown in this paper, for reasons privacy and protecting the intellectual property). Every dimension in the warehouse has multiple rows in the entry table that defines the metadata such as the time, is inserted in the warehouse, and its time of last update. Every warehouse entity is also associated with the dataset and database from which domain it was derived. Different data views are drawn from the metadata of a warehouse for interpreting trends and correlations [9], [12] for understanding the makeup of the connectivity and interaction among relevant data attributes of multiple dimensions. Using *Grapher* and *Surfer* [15] solutions, data views are plotted in the form of bubble and surface map views, as shown in Figs. 5 - 6. Data relationships plotted between period and other conceptualized (logical) data attributes are deduced, with interesting connectivity among the plotted data properties.

IPA published data, are used for implementing the current proposed methodologies. Structure, reservoir (in geological domain) and production data attributes among different petroleum systems and among different depositional regimes, provide valuable data on correlations and trends among periodically and geographically distributed petroleum systems. As visualized in Figs. 5-6, several data views drawn from warehouses are interpreted.

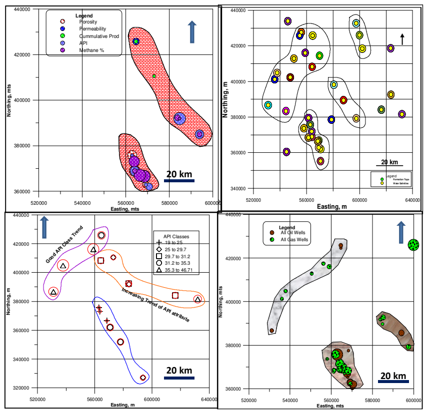

Fig. 5: Data view plots, showing multidimensional data attributes lobes interpreted exploration & production data trends

In a 2D bubble plot, as shown in Fig. 5, several multidimensional data views are plotted to understand varying instances of multidimensional attributes and their strengths. In the bubble plot, the diameter of each bubble can vary in size, providing a way to represent an additional dimension of the data. For example, as demonstrated in Fig. 5 scatter plots of drilled-depths with salinity of water formations at different stratigraphic intervals and also geographic locations (inside basin boundaries) suggest correlations and trends among formations within an ecosystem scenario. Similar, is the case with reservoir and production data trends, trending NW – SE direction.

As stated earlier, for a sedimentary basin, several multidimensional data attributes and their strengths given in [1], [14] are ontologically described for designing and developing Petroleum Digital Ecosystem (PDE). As an example, in the current paper, PDE of a sedimentary basin is demonstrated. Similar PDE may be designed for other sedimentary basins in the Indonesian basins. These studies can further be extended in Southeast Asia (Fig.1). It is significant to observe, multiple dimensions extracted from an integrated framework (Fig. 3), which narrate definite correlation-trends, as it is interpretable by a knowledgeable explorer. Here, size of each bubble has significance in terms of its characteristics of attributes (and direction of characteristic property) described in a sedimentary basin as shown in Fig. 6.

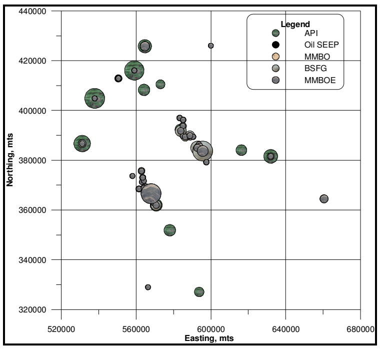

Fig. 6: Plot view of multidimensional data attributes, showing coherency of alignments among attributes

In Southeast Asia, as a Total Petroleum System (TPS) scenario, there are multiple sedimentary basins, each basin has multiple oil and gas fields, each field has multiple oil/gas drilled wells, each well has multiple oil/gas net-pays. Each well has encountered different geological formations and each formation has multiple reservoir characteristics. Hierarchically, in any geological domain, all these multidimensional data are ecologically (inherent) interconnected and being interacted (evolution) continuously in multiple geological ages.

VI. CONCLUSION AND SCOPE

1. Southeast Asia has many countries with hundreds of islands and their associated sedimentary basins. Data warehouse provides a single information repository for massive collection of data for Petroleum Systems Analysis and Sedimentary Basin Modelling from multiple Indonesian productive basins (for example). The current studies organize and deliver data to different users, in particular those who explore and mine the data, perform analyses, interpret different geological events and prepare data models in a sedimentary basin context. Data views and their interpretations facilitate the managers and explorers, who plan and manage drillable exploratory and development targets in different basins.
2. Multiple dimensional data structuring developed in a sedimentary basin is more flexible and these models are reusable in different knowledge domains (ontologies) in other basins of Southeast Asia.
3. Data warehousing approach is more suitable, when considering large size sedimentary basins that comprise of multiple petroleum systems in border countries of Malaysia and Indonesia, including Papua New Guinea and Australia in the north-eastern part of Indonesia, in a Total Petroleum System (TPS) scenario, where multiple oil and gas fields exist.
4. Several data properties deduced from different depositional systems, such as fluvial, delta-plain, delta-front and pro-delta systems, suggest interesting correlations among petroleum systems existing in the sedimentary basin.
5. Data attributes and properties drawn among several dimensions show significant correlations, trends and relationships among multiple data attributes such as, *structure*, *reservoir*, *production* and other geological, geophysical and geochemical data dimensions and attributes considered for analysis.

Studies are in progress, building models for different petroleum systems and their data-views for interpretation in different knowledge domains.

REFERENCES

[1] Courteney, S. Cockcroft, R. Lorentz, R. Miller, R. Ott, H.L. Prijosoesilo, A.R, Suhendan, A.W.R, Wight and Biman, S.K, 1991. Indonesia – Oil & Gas Field Atlas Volume V: Kalimantan, IPA 032, p1-21.

[2] Hoffer, J.A, Presscot, M.B and McFadden, F.R, Modern Database Management, 7th Edition, Prentice Hal l, 2005

[3] Magoom B. L and Dow, D.W (1994) The Petroleum System – from Source to Trap, AAPG Memoir 60, p. 1-625

[4] Nimmagadda, S.L. and Dreher, H. (2006) Ontology-base data warehousing and mining approaches in petroleum industries: in Negro, H.O, Cisaro, S.G. and Xodo, D. (Eds.*), Data Mining with Ontologies: Implementation, Findings and Framework*, a book published in 2007 by Idea Group Inc. http://www.exa.unicen.edu.au/dmontolo/

[5] Nimmagadda, S.L, and Dreher, H. (2006a) Mapping and modelling of Oil and Gas Relational Data Objects for Warehouse Development and Efficient Data Mining, a paper presented and published in the *proceedings of the 4th International Conference of IEEE Industry Informatics*, held in Singapore, August

[6] Nimmagadda, S.L. and Dreher, H., Rudra, A (2005a) Ontology of Western Australian petroleum exploration data for effective data warehouse design and data mining, *Proceedings of 3rd international IEEE conference on Industrial Informatics*, held in Perth, Australia

[7] 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 Indonesian Petroleum Association (IPA), held in Jakarta, Indonesia.

[8] Parasnis, D.S, Principles of Applied Geophysics, Chapman & Hall, 1997

[9] Pujari, A.K. Data Mining Techniques, 2002, University Press, 2002, p. 7-66

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

[11] Shastri L Nimmagadda and Heinz Dreher, (2011) 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.

[12] Shawkat Ali, A. B. M. and Wasimi, S. A. (2007) Data Mining: Methods and Techniques, p. 196-219 and p. 25-267.

[13] Thomas, J.L, Yannick, P. Valerie, W., Gupta, P. Stringer-Calvert, D.WJ, Tenenbaum, J.D and Karp, P.D Bio-warehouse: a bioinformatics database warehouse toolkit; BMC Bioinformatics 2006, 7:170, p.1-14; http:///www.biomedcentral.com/1471-2105/7/170

[14] Wight, A.W.R, Hare, L.H, and Reynolds, J.R, A Sedimentary Basin: (1992) NE Kalimantan, Indonesia: a century of exploration and future potential, Geological Society of Malaysia, Circum – Pacific Council for Energy and Mineral Resources.

[15] *http://www.goldensoftware.com*