The IEEE International Conference on
Industrial Informatics (INDIN 2008)
DCC, Daejeon, Korea July 13-16, 2008

Petroleum Ontology: an effective data integration and mining methodology aiding exploration of commercial petroleum plays

Shastri L NIMMAGADDA* and Heinz DREHER**
Data and Consulting Services, Schlumberger, East Ahmadi, 61001, Kuwait, and Curtin Business School, Perth, Australia
snimmagadda@kuwait.oilfield.slb.com
Curtin University of Technology, Curtin Business School, Perth, WA, Australia**
h.dreher@curtin.edu.au

***Abstract*- the success of petroleum exploration business depends on appropriate design and implementation of seismic exploration programs. Volume of seismic data instances (deduced for structural interpretation) is used for modeling multidimensional data structures in a warehouse environment. Several business rules (constraints) are applied to the design and development of a data warehouse. Seismic-time horizons, well-data, reservoir properties and petroleum production as data dimensions or classes and their associated attributes, have been used for integrating and modeling their data instances with numerous hierarchies. Ontology addresses issues of semantics and contexts (naming conventions) involved during design of business rules and building relationships among several hierarchies. Business rules inform the data integration and data mining process, especially when heterogeneous data structures are denormalized for fine-grained data structuring and facilitate the association rule mining process. Several data views have been presented for analyzing the effectiveness and deliverability of warehouse-modeled information. Design of business rules combined with fine-grained ontology structuring appears to have a definitive impact on data mining of seismic data instances, enhancing seismic data knowledge and improving geological interpretation. Petroleum ontology proves to be an effective knowledge mapping tool, which can revolutionize the petroleum exploration industries.**

I. Introduction

Structures (either in time- or depth domain) and reservoir attributes of multiple hydrocarbon bearing horizons [1]-[6], are key ingredients for any petroleum prospect delineation and evaluation [1]. These structures and reservoirs possess several properties with different magnitudes. These are basic data attributes and instances, characterized by different seismic events, which are perceivably geological events. The physical process of a seismic reflection, as demonstrated in Fig.1, establishes movement of ray-paths from successive sub-surface layers [1] - [3] (in multiple dimensions) in the form of seismograms. The later-arrival of seismic times (data instances documented in the database) yield information about deeper layers (called sediments in petroleum geology). The method is popularly known as common reflection or common midpoint, or common depth point (CDP), in which sub-surface coverage [1] is one-half of the surface distance across the geophone spread (length of survey line along which, receivers "geophones" sensors planted). Several survey lines laid on the surface are interconnected through sub-surface CDP data representation, as described in [1] and [3]. S and R are respectively source and receiver (sensor) kept on a topographic surface in different domains (Fig.1). CDP is a conceptualized domain (a key concept) based on a reflecting surface. Domain ontology understands that S and R are in different domains.

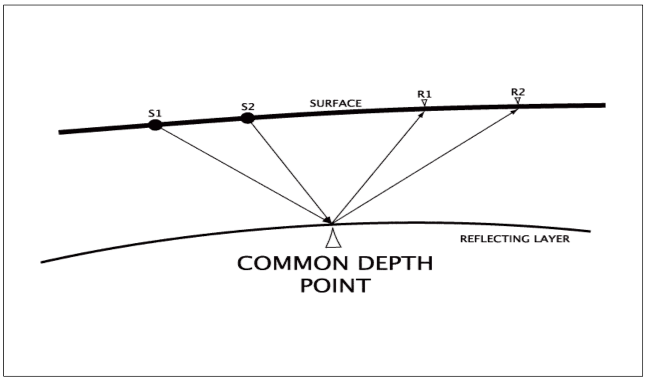

Fig. 1: Data presentation in *Common Depth Point (CDP)* domain

Ontologically, all the data instances are inter-related and organized in different ways [4]-[12], such as CDP, in which case, common midpoint dimension is logically related to other dimensions derived from receivers and sources. Similar is the case with point dimensions derivable from common receiver, common source and common offset domains, as discussed in the following sections.

Authors demonstrate (Fig.1) that a CDP is a logically conceptualized data dimension created by half distance (in a geometric sense) between sensors (receivers) and sources (points at which, shots are taken) located on the surface. But these point dimensions appear to have followed a sub-surface reflecting plane (in a geological sense, though it is not necessary to strictly follow a geometry), which is geologically interpreted as a sedimentary layer [1]-[3]. For multiple layers (horizons), multiple CDP data dimensions - a set of common receiver and source point data dimensions are described from the spread of seismic profiles (Fig.2) laid on the surface. Logically, different arrays are designed on each seismic profile in the field to remove diverse noise data patterns recorded during seismic data acquisition. These data are conventionally processed on high speed digital systems in processing centers or in the field.

Seismic data acquired in the producing or remote fields, possess multi-dimensional data representations as shown in seismograms (Fig.3), which are logically processed for interpreting new oil-plays and leads. These data are appropriately sampled and stored in high density storage devices.

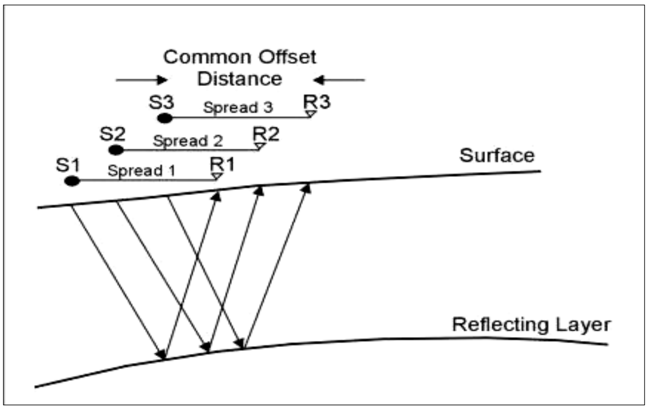

Fig. 2: Data representation in common offset domain

Seismograms (Fig.3) representing seismic data instances described in terms of amplitude, frequency and phase characteristics, illustrate dependence of each data instance amplitude value on each other's coherency attribute data.

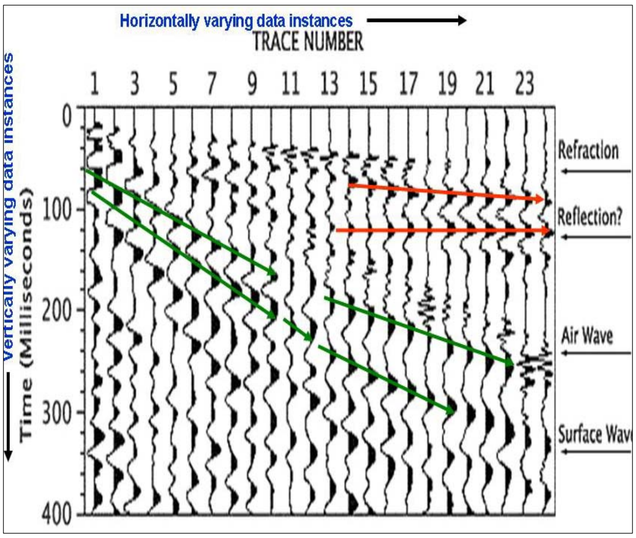

Fig. 3: Description of amplitudes and their variations on multiple seismic traces, showing trace coherency attribute (amplitude data instances dependency)

Ontology organizes and maintains the multidimensional data in different domains based on the contents and contexts. Another interesting domain is geology, in which all petro-facies that are interpreted (through relational ontology from seismic domain) by all seismic facies. Interestingly, these seismic facies possessing several amplitudes and polarities are classified to specific litho-facies or lithology (in a sub-domain of geology). Further relevant domains of geology are, structure, petrology and stratigraphy. Structural, lithology and stratigraphic [1] data patterns are interpreted in multiple data dimensions as on seismograms (as shown in Figs. 3 and 13) after logically organizing the horizon data in time and depth domains.

The seismic data are processed in other domains such as common offset, common receiver and common source, in order to see improvements in attributes of lateral, longitudinal and horizontal dimensions and their data instances. As shown in Figs. 4 and 5, depth points are typically ontologically conceptualized, so that relationships among multiple dimensions are intelligently organized and stored in a warehouse environment for integration purposes.

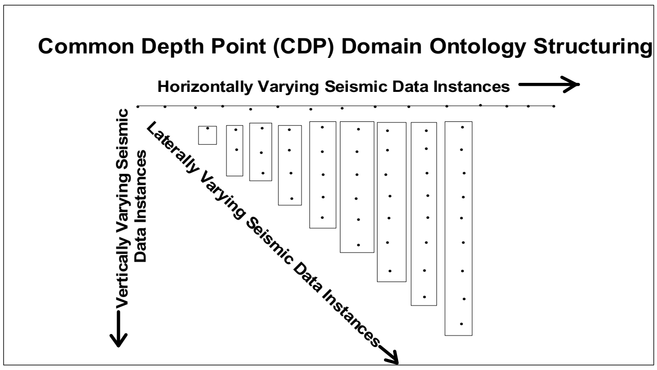

Fig. 4: Relational ontology - *common depth point* data instances representation (CDP ontology)

Different domains of similar (or dissimilar) ontologies can be described based on different geological and geophysical (seismic is the present context) data situations of different petroleum provinces, under investigation.

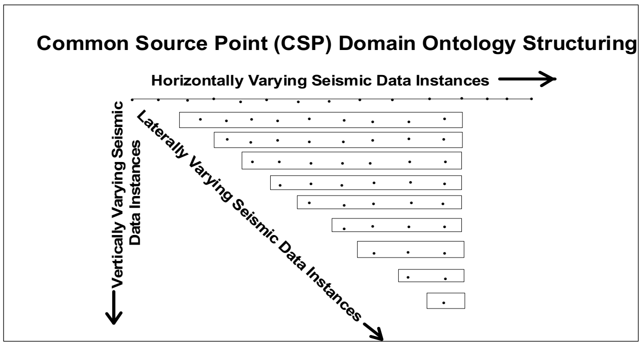

Fig. 5: Relational ontology - *common source point* data instances representation (CSP ontology)

II. ONTOLOGY MODELLING

A. Domain Ontology - design of data acquisition, processing and interpretation entities

Several hierarchies are narrated in a high level dimension, exploration from basin level to CDP or survey point level (Fig. 6).

Data acquisition is a first stage of any exploration activity, involving the different data acquisition procedures on the ground in the prospective area. Ontology DB checks for appropriate semantics, contexts, and thus informs the building of relationships among several parameters [2], [4] and [5] contributing to seismic signals (traces), and ground and source generated noises [1] – [3]. Source and geophone sensors and arrays are needed on the ground for suppressing the random, coherent and other ambient noises, before seismic data are recorded in the field vehicle.

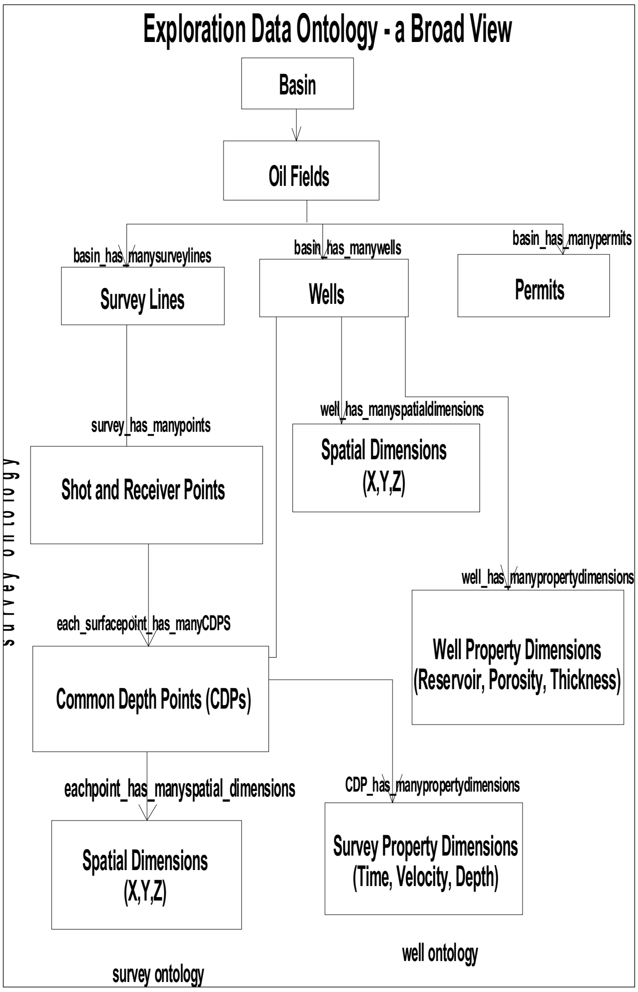

Fig. 6: Hierarchy Ontology modeling – *Basin* and *Exploration* data entities

Seismic data processing is the next domain, in which all the seismic traces are sorted and stacked (in multiple dimensions) after applying both static and dynamic corrections (in terms of milliseconds, ms), time or space varying filters and deconvolution processing. Migration is the final stage of seismic data processing, in which the true positions of sources are represented on the ground. Ontology [7] – [12] replaces all these processing (in terms of data mining) steps, reorganizing and regrouping all the traces based on the positions of the sources designed on the surface of the earth. Last stage of exploration is data interpretation, in which mined data are interpreted for exploring data patterns, correlations and trends, which enhance the geological perception and topography of individual sub-surface layers (petroleum bearing sediments, commonly represented in layers [1] and [2]). These are further elaborated in the next sections.

B. *Ontology based Seismic Designs*

The generation and recording of seismic data involve several issues. Acquisition requires many different receiver configurations (sensor dimension), including laying geophones or seismometers (sensor arrays) on the surface of the earth or seafloor, towing hydrophones [1] – [5] behind a marine seismic vessel, suspending hydrophones vertically in the sea or placing geophones in a wellbore (as in a vertical seismic profile, VSP ontology, [6] and [13]) to record the seismic signal. A source (source or shot dimensions), such as a vibrator unit, dynamite shot, or an air gun (array of shots), generates acoustic or elastic vibrations that travel into the Earth, pass through strata with different seismic responses and filtering effects, and return to the surface to be recorded as seismic data. Ontology maintains optimal acquisition of restructuring data-array patterns and adapts according to local conditions. It involves employing the appropriate source type and intensity, optimal configuration of receivers, and orientation of receiver lines with respect to geological depth-domain features and required seismic signals in seismic time-domain. This ensures that the best signal-to-noise ratio can be achieved with appropriate resolution [3] and extraneous effects such as air waves, ground roll, multiples and diffractions, all sources of unwanted noise, are minimized, distinguished, and removed through processing. Ontology understands and maintains signal and noise data patterns, separating them for data mining enhancements.

C. *Ontology based Processing Designs*

Ontology understands, examines and investigates all the CDPs and their dimension representations. As per ground-geometries, ontology identifies seismic signals in multiple directions. Seismic data are manipulated to suppress noise, enhance signal and migrate seismic events to the appropriate location in space-domain. Processing steps [1] typically include analysis of velocities and frequencies, static corrections, deconvolution, normal move-out, dip move-out, stacking, and migration, which can be performed before or after stacking of seismic traces. Seismic processing, synonymous to data mining, facilitates interpretation of exploration data, because subsurface structures of horizons and seismic reflection geometries are processed more apparently based on field-layouts and geological inputs.

D. *Ontology based Seismic Interpretation*

Interpretation of data, mined from warehoused (processed) data is key for generating petroleum prospects and risk evaluating them. In seismic geophysics [1], analysis of data is part of exploration to generate precise geological models and then predict the reservoir properties and structures of the subsurface. Interpretation of seismic data is the primary concern of petroleum geophysicists or geologists for locating wells and well-drill planning. All the seismic horizons (in seismic-domain) and geologically derived depth horizons (in well-domain) are analyzed and integrated to get consensus geological models (reservoir models), which are analyzed for hydrocarbon plays, maintaining productivity from well and health of the horizon (formation).

III. ISSUES OF DATA MANAGEMENT IN PETROLEUM INDUSTRY

Seismic data acquired in the field, represent in general, numerous dimensions. Exploration companies archive large data (multiple dimensions) in different servers, but they are poorly organized at times with redundancy and difficult to retrieve. Large volumes of datasets are loaded and distributed among PC and Unix/Linux based workstations for processing and interpreting the petroleum exploration data. Specialized software and hardware modules are used for carrying out the specialized tasks in these commercial companies. Data processing and interpretation qualities depend on easy accessing and handling of volumes of datasets emanating from different operational units of the company without any loss of information and value. Today, organizing the data intelligently for effective knowledge discovery is an art of database management. Intelligent storage of data [14]-[18] by different data structuring methods [8], [19]-[23] is much needed for more precise and rapid data manipulation. Reuse and interoperability are other issues that need to be addressed in large commercial petroleum companies, where there are service and smaller operational units, delivering commercial products. Data structures are reused [15] and with more flexibility on integrated software and hardware modules.

Ontology based data warehousing and mining technology is a solution [9], which can not only structure the data more intelligently, but also integrate and store the data for mining and analysis. Shared ontologies [14] play a vital role during intelligent integration of petroleum data.

IV. WAREHOUSING OF EXPLORATION DATA INSTANCES – METADATA DOMAIN

Data instances extracted from the field are reorganized at the data modeling centre into different data structures in multidimensional formats. One popular structure is a star-schema, logically stored in a warehouse environment for data integration, addressing the interoperability issue. These data organized in 2D/3D grids are spread across an exploration area in vertical, horizontal and lateral dimensions. These data are in the form of several dimensions (attributes), such as *location of survey, latitude and longitude, seismic times, velocities and computed depths* are data instances for several horizons or reservoirs (formations). Ontology resolves issues of naming conventions and semantics among data attributes.

Metadata serves to identify the semantic content and location (geographic dimension) of data in the seismic data warehouse and is a bridge between the data warehouse and decision support application, derived through conceptual domain ontology modeling. In addition to providing a logical linkage between data and application, Metadata pinpoints access to the information across the entire warehouse and can enable development of applications to automatically update themselves and look for data warehouse content changes or updates.

In a traditional database, a schema is described as a conceptual or logical data organization [19] of all dimensions (objects or entities) that describe all relationships between known attributes. In such a well-defined concept, the difference between Metadata and data disappears – metadata is simple data. In the context of data warehouse, metadata is needed to describe data relationships without any further ambiguity of interpreting multiple dimensions (for example, attributes from *seismic, wells* and *reservoirs (horizons)* as described in [1] and [2]).

A. *Description of exploration data types*

Seismic and well-data are typical surface and sub-surface domains discussed in the ontologies and integration of these ontologies depends on its effective conceptualization and integration of respective relationships (an integrated framework - Fig. 7) among multiple dimensions and their associated attribute properties. Data structure, representing super-type dimension, such as *seismic*, has several related multiple dimensions, which are supposed to be logically intelligible within a schema that narrates a focused seismic ontology. In seismic ontology, *time* dimension and its attribute properties are described and previously unknown relationships are built. Attribute properties of *time* dimension, are dynamic, depending upon type of structuring, within which *time* dimension is interpreted. Data instances of *time* dimension are extracted from different domains of data representation, such as CDP, COP (common offset point) dimension, CSP (common source point) dimension and CRP (common receiver point) dimension [1], which are structured in a relational ontology. To further explain interpretation of *time* dimension in different knowledge domains, several hierarchical and relational structuring of CDP, COP, CSP and CRP dimensions are ontologically conceptualized.

Similar to seismic domain ontologies, well-base ontology is described, narrating all the relationships among dimensions, such as drilled well, formation top, depth and horizon. One or two dimension IDs of this well-domain ontology must match with seismic-domain ontology. This will enable building of relationships and integration of these two different entities or dimensions. Attribute properties of these domain ontologies, are also important, since these may vary spatially with horizontal, vertical and lateral (direction) dimensions (Fig. 7).

B. *Business rules and constraints for building ontology models*

Data interpretation and knowledge discovery depend on how business rules and constraints [19] and [20] are framed and described during design of particular domain ontology and its implementation in a warehousing environment.

As narrated in Fig. 6, a hierarchy represents that each sedimentary basin has several petroleum *fields* and each *field* must have possessed one or more *drilled-well*. Each *field* possesses several seismic *surveys*, including volume of 3D datasets. Each *oil-field* dimension should have one or more *surveys* or *wells* dimensions. Each *drilled-well* must have one or more *horizon* dimensions identified for modeling and interpretation purposes.

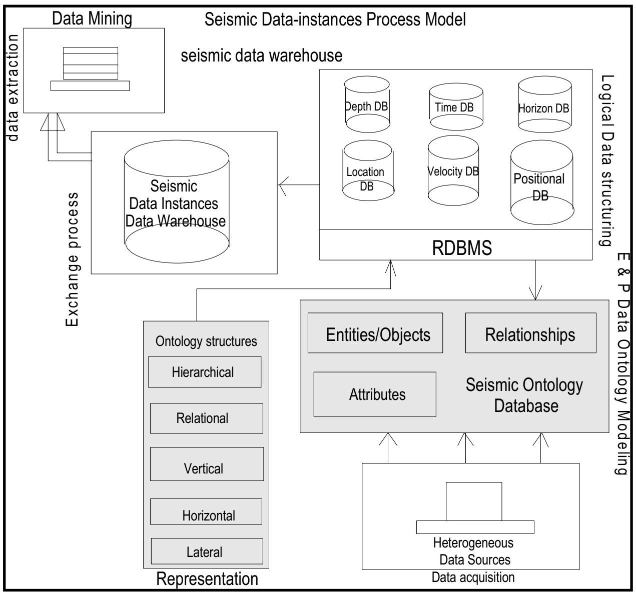

Fig. 7: Ontology-based warehouse framework for integrating exploration data instances and storing them intelligently for data mining

It is not mandatory to possess a producing horizon, but one must have *survey* and *well* dimensions. Besides these,

1. Each receiver or source point dimension has unique spatial identity, implying that for different coordinates, there may be similar or dissimilar seismic data instances in any direction (such as horizontal or vertical or lateral).
2. Each shot or source location may have been described with one or more CMP/CDP dimensions.
3. Each *horizon* (geological formation) dimension has several seismic and well data instances.
4. *Horizon* and *source* or *receiver* location identifiers must be unique.

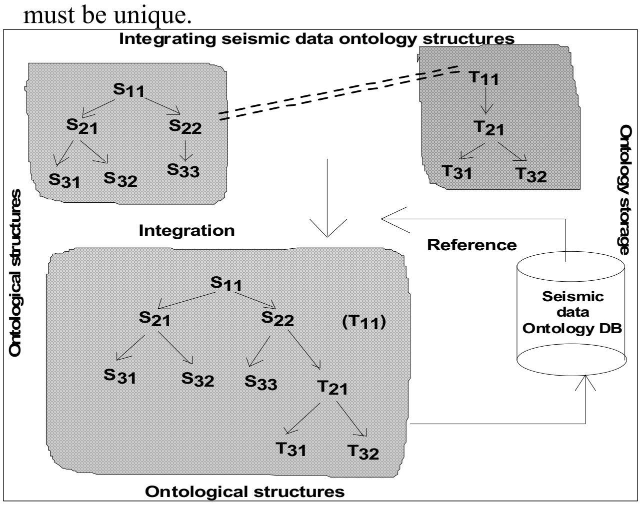

Fig. 8: Integrating seismic data instances from several seismic (or integrated from wells data) horizons

5. Dimensions derived from conceptualization (true geological realizations) of horizon properties, such as *structural-high* and *structural-low* may repeat, so data instances of these properties interpreted for each horizon may be repetitive, but with different magnitudes.
6. Attributes and their instances, such as *amplitude*, *frequency* and *phase* of *seismic time* events, for each *horizon* (geological formation) may be repetitive elsewhere in other structure with similar data instances.

C. *Ontology-base data model - CDP domain*

Seismic data instances distributed from one domain among several field grids are also associated with well placement in a different domain. Common depth points (CDP) is a vertically varying data distribution, a widely implemented modeling procedure in seismic prospecting and exploration, for exploring and exploiting depth information. Authors interpret these data distribution, as ontology conceptualization in which each depth point, varying with depth, has its association with other depth points. Both horizontally varying surface grid and vertically varying data building blocks (Figs. 7 - 9) are used. As demonstrated in Fig. 8, data instances gathered at several hierarchies (parent and child levels) are integrated, as in our case, exploration, a high level dimension, integrates child ontologies of seismic and well domains.

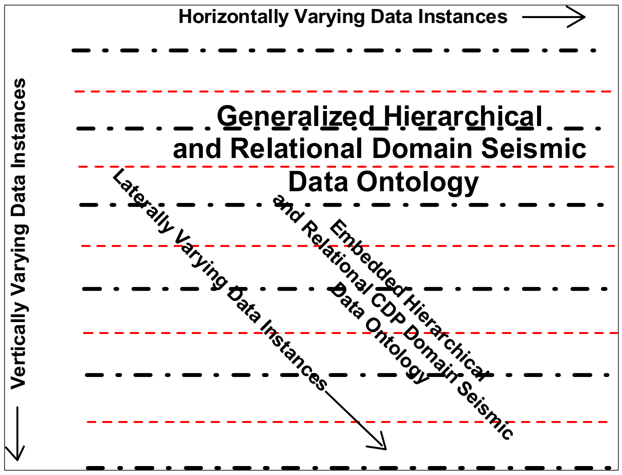

Fig. 9: a 3D seismic field layout – representation of seismic data instances in different directions (each characterizing a particular dimension)

As examined in Fig. 9, CDP data dimensions from 3D field layout area are ontologically interconnected in horizontal, vertical and lateral dimensions. All the data instances, falling in vertical blocks (as shown in Figs. 4, 5 and 7) and varying vertically, are aligned and integrated to get a particular data alignment that can classify to specific attributable structural data instances.

V. RESULTS AND DISCUSSIONS

Several strati–surfaces (horizons) are explored from processed seismic data. These horizons existing in a geological era, are not static, but are part of a dynamic petroleum ecosystem [13]. Warehoused data have been implemented, interpreting several geological structures which are used for prospect generation and evaluation. As shown in Figs. 10-15, several seismic and well data instances have been integrated using multiple dimensions. In-lines, cross-lines, point and polygon dimensions have been used for gathering different domain data and thus constructing a metadata set.

Seismic and well-data domains are key ontologies in the present study for data modeling. Warehousing approach integrates [9]-[14] dimensions along with attributes, besides building relationships and exploring useful data views for further interpreting them in terms of structural connectivity among multiple oil fields.

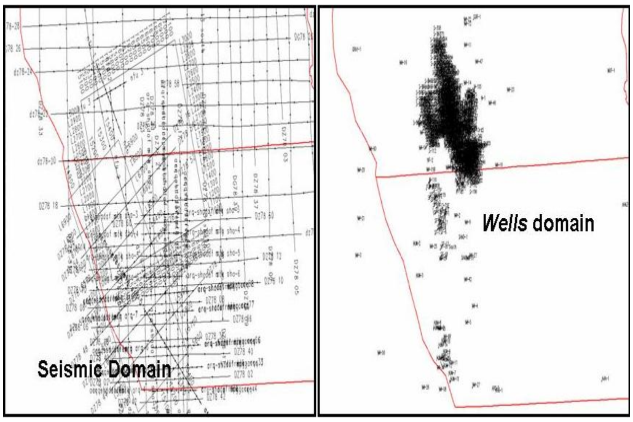

Fig. 10: Seismic and well-data instances showing with spatial dimensions

Spatial dimensions are very significant in the present study, where point, lines and contour surfaces characterize a framework, and are responsible for connectivity among fields. But these dimensions and attributes are independent of structural attributes of seismic time, depth and velocity dimensions. Figs 11-12 represent maps in seismic domain and well domain, inherently described in an ontologically integrated domain (Fig. 7).

Structural relationships have been built and interpreted along these spatial data grids, again categorizing and separating structural high and low anomalies (structure domain with high and low dimensions). These are important relationship attributes between seismic and well domain ontologies (integrated from relational and hierarchical ontology domains).

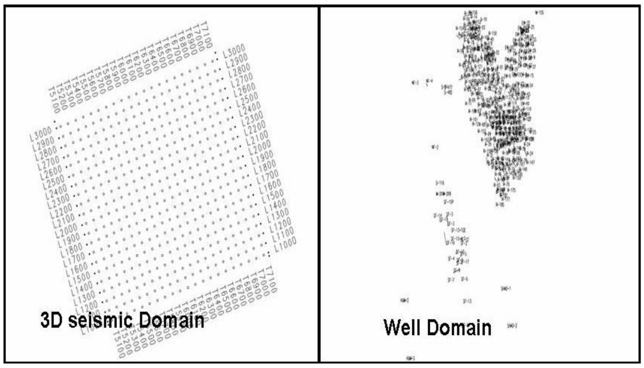

Fig. 11: Seismic data instances in 3D grid and well-data instances

Seismic line, point and polygon representing field and country boundaries shown in Figs.10-12 are used in the grid processing. Contours are generated based on grid data. Each field is interpreted and identified within a structural data contour pattern, supposedly having all structural high data instances. Seismic structural high and low data values are representative of geological structures, depending upon the velocity data attributes [6]. Other attributes such as seismic amplitudes are presented in different color patterns as narrated in the Fig. 13.

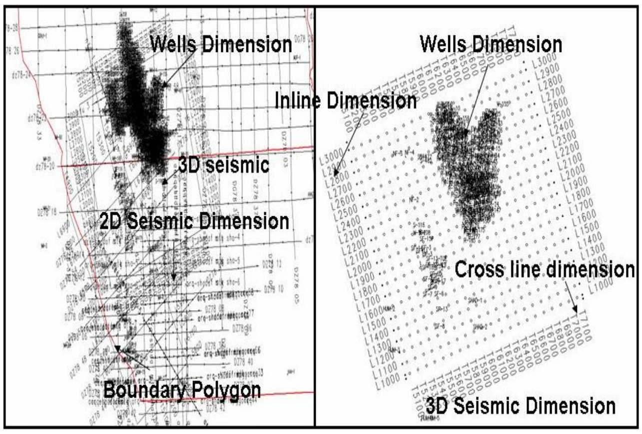

Fig. 12: Seismic – well - data domain and integration of attribute instances for – Metadata

Ontologically, each field interpreted within a petroleum system cannot be isolated, but conceptually and logically, in a strict geological sense, fields are interconnected and communicative to each other by different structural patterns. Ecologically [13], if there is any change in the structural pattern within a field, there is a corresponding affect on the other field or data pattern.

Another significant feature within an implementation of warehoused metadata is the separation of all structural-highs possessing lower seismic data values, but with different magnitudes as represented, for example, in red-green color data patterns. Structural-lows possess definite higher (deeper) seismic values, represented as blue color data patterns. These structural data patterns are valuable clues for any prospect generation and evaluation. Fig 13 is an example of a map showing oil field structural data attributes drawn from an integrated Metadata model (Fig. 7).

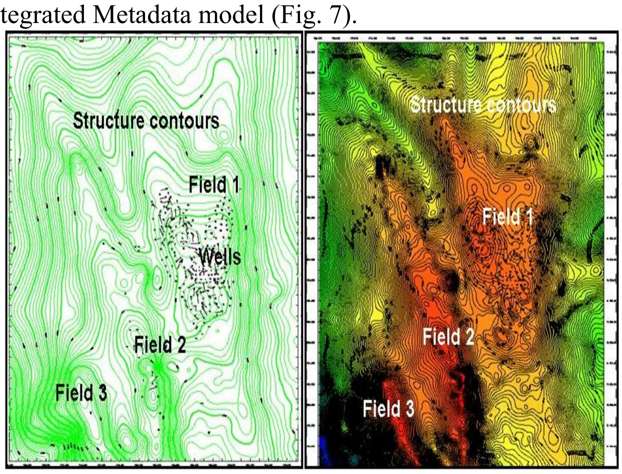

Fig. 13: Data integration - connecting fields through relational ontology (after applying constraints)

VI. CONCLUSIONS AND RECOMMENDATIONS

1. Ontology based design of data acquisition; data processing and interpretation are more effective in extracting knowledge on structural and petroleum geology domains. This is indeed a revolutionary concept in the fields of exploration and development of petroleum fields, which has more future scope of commercial research.
2. Data instances from CDP, COP, CRP, CSP dimensions can easily be structured and integrated in a warehouse environment for the purpose of interpreting seismic signals and their attributes as required by geologists, well planners, reservoir and production engineers.
3. Framing of business rules, constraints among usage of attributes within the integration process, and design of business rules have definite impact on correlation and mapping and thus on data mining.
4. Ontology based data structuring has definite advantage, especially when attributes and their relationships are conceptualized using appropriate semantics and contexts.
5. Integrating ontologically structured data in a warehousing environment has more flexibility and consistency in attribute mapping and interpretation during data mining stage.
6. Structural data views taken from implemented warehoused metadata follow definite structure shapes, in terms of seismic high and low data instances, depicting geological knowledge.
7. Integration of exploration data, modeled from different hierarchically derived multiple dimensions; facilitate the data mining process, thus extracting knowledge of commercial petroleum plays. Issues of reuse and interoperability of denormalized fine-grained exploration data structures have also been emphasized in the context of implementing ontology based warehousing in petroleum exploration industries.

This research work addresses new methodologies, which have the potential to revolutionize the exploration and resources industries worldwide. This is an on-going commercial research work at Curtin Business School, Curtin University of Technology, Australia. Industry collaboration is accommodated through the University's commercialization arm – contact the second named author.

REFERENCES

[1] Beaumont, E.A and Foster, N.H. "Exploring for Oil & Gas Traps, AAPG Publications of Millennium Edition", (1999).

[2] Gilbert, R. Liu, Y. Abriel, W. and Preece, R. "Reservoir modeling: integrating various data at appropriate scales", *Leading Edge* (2004), Volume 23 (8), p.784-788.

[3] Telford, W.M. Geldart, L.P. and Sheriff, R.E. "Applied Geophysics", (1990), Second Edition, p.100-350 and 600-750.

[4] Chakraborty, K. Mubarak, Al. H, Nimmagadda, S.L. and Ray, J. "Exploration Data Integration, an effective reengineering process for new petroleum plays in Gulf Offshore Basins", presented at the 2006 International Conference of *AAPG*, November, Perth, Australia.

[5] Chakraborty, K. Al-Hajeri, M and Nimmagadda, S.L. "3D Seismic Data Attributes Analysis for Predicting Wara reservoir qualities in the Al-Khafji Field, Middle East", *presented at the 2007 3rd North African/Mediterranean Petroleum & Geoscience Conference & Exhibition*, February, Tripoli, (2007), Libya.

[6] Nimmagadda, S. L. and Dreher, H. Ontology-based Warehouse Time-Depth Data Modelling Framework for Improved Seismic Interpretation in Onshore Producing Basins, a paper presented and published in the *proceedings of the International Petroleum Technology Conference (IPTC),* (2007), Dubai, UAE.

[7] Nimmagadda, S.L. and Rudra, A. "Data sources and requirement analysis for multidimensional database modeling – an Australian Resources Industry scenario", *presented at the 7th international conference* on IT, (2004), Hyderabad, India.

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

[9] Nimmagadda, S.L. and Dreher, H. "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/

[10] Nimmagadda, S.L. and Dreher, H. "Ontology of Western Australian petroleum exploration data for effective data warehouse design and data mining", presented and published in the *proceedings of the 3rd international IEEE conference on Industrial Informatics*, (2005), Perth, Australia.

[11] Nimmagadda, S.L, and Dreher, H. "Mapping of Oil and Gas Business Data Entities for Effective Operational Management", presented and published in the *proceedings of the 4th International Conference of IEEE Industry Informatics*, (2006), Singapore.

[12] Nimmagadda, S.L., and Dreher, H. "Ontology based data warehouse modelling and mining of earthquake data: prediction analysis along Eurasian-Australian continental plates", a paper accepted for presentation in the *International Conference of IEEE in Industry Informatics Forum*, (2007) Vienna, Austria.

[13] Nimmagadda, S.L. and Dreher, H. Ontology Based Data Warehouse Modelling – a Methodology for Managing Petroleum Field Ecosystems, a paper presented and published in the *proceedings of the 2nd International Conference of IEEE-DEST*, held in Phitsanulok, (2008), Thailand.

[14] Meersman, R.A. "Foundations, implementations and applications of web semantics, parts 1, 2, 3", *Lecture Notes presented at School of Information Systems*, Curtin Business School, 2004.

[15] Valente, A., Russ, T., MacGregor, R., and Swartout, W. "Building and (re)using an ontology of air campaign planning", *IEEE Intelligent Systems*, 14(1), (1999), pp. 27–36.

[16] Jasper, R. and Uschold, M. A. "Framework for understanding and classifying ontology applications", published in the *proceedings of the IJCAI-99 ontology workshop*, (1999), p.1-20.

[17] Uschold, M.E. "Knowledge level modeling: concepts and terminology", *Knowledge Engineering Review*, (1998), 13(1).

[18] Shanks, G. Tansley, E. Weber, R. "Using Ontology to validate conceptual models", *communications of the ACM*, (2003), 46(10), pp. 85-89.

[19] Hoffer, J.A, Presscot, M.B and McFadden, F.R. "Modern Database Management", (2005), 7th Edition, Prentice Hall.

[20] Pujari, A.K. "Data Mining Techniques", Universities Press (India) Private Limited, (2002), p 1-40.

[21] Ozkarahan, E. "Database Management, Concepts, Design and Practice", (1990), Prentice Hall Publications.

[22] Uschold, M. & Gruninger, M. Ontologies: "Principles, methods and applications", *Knowledge Engineering Review*, (1996), 11(2).

[23] Shanks, G. Tansley, E. and Weber, R. "Representing composites in conceptual modelling", *Communications of ACM*, (2004), 47 (7), pp. 77-80.