Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 1881–1890

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

An Upstream Business Data Science in a Big Data Perspective

Shastri L Nimmagadda*, Torsten Reiners and Amit Rudra

School of Information Systems, Curtin Business School, Curtin University, Perth, WA, Australia

Abstract

The rugged *geographie*s, *geomorphologies* and complex *geological* environments make the explorers more challenging exploration and production (E & P). Despite challenges, many sedimentary basins, associated oil & gas fields and E & P Ventures are productive and commercially viable. The difficulty in understanding the connectivity among multiple reservoirs is due to lack of knowledge of multidisciplinary data of petroleum systems, complicating the data integration and interpretation process. The geological and geophysical data of an upstream business are vital assets of any oil & gas industry, in particular in E & P perspective. The data are often unstructured with a variety of anomalous attributes, mingling with volumes of spatial-temporal dimension attributes and instances. In recent years, the concepts of Big Data have taken different hype in petroleum industries, because of involvement of big sized data in the data integration process. Because of the unstructured data sources, a new direction in the database organization is needed. Investigating the science behind the Big Data and their integrated interpretation of the upstream project is a principal objective of the research. In this context, various constructs and models are articulated with different artefacts. Opportunities of Big Data are explored with exploration data and business analytics, supporting sustainable E & P systems. Petroleum management information systems (PMIS) and digital petroleum ecosystems (PDE) are developed to establish a connectivity among various data sources in multiple domains and systems. The implementation of robust methodologies ascertains the significance of the integrated upstream business in the oil and gas industries that comply with the characteristics of the Big Data.

Peer-review under responsibility of KES International

Keywords: Upstream Business; Big Data; Data Science; Digital Ecosystem; Exploration & Production;

1. Introduction

A large amount of heterogeneous and multidimensional E & P data sources of petroleum bearing sedimentary basins is accumulated in many upstream companies and as shown in Fig. 1, they can be represented in digital form[10]. Presently, there is no comprehensive and robust data management and modelling methodologies for managing the E & P data sources of multiple petroleum systems.

* Corresponding author. *E-mail address:* shastri.nimmagadda@curtin.edu.au

1877-0509
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.236

The E & P datasets need a framework and workflow for integrating and risk minimizing the oil and gas businesses, in particular, the Australian contexts. The North West Shelf designated as a total petroleum system (TPS) is termed as the Super Westralian Basin, in which *shelf, slope* and *deep-basin*[10] geological data occurrences seem responsible for the connectedness through a phenomenon, "digital ecosystem"[7],[10]. Also, an assembly of sub-basins, which is a larger part of the super-basin has many petroleum systems, and each system has either known or unknown or incomplete spatial ranges. Each petroleum system may have various oil and gas fields, with hierarchical structuring of attribute data dimensions and their fact instances. Such Big Data[6, 10, 12] sources associated with the western part of Western Australia (Fig. 1) have a choice of analysing their attribute dimensions and instances in multidimensional warehouse repositories through different constructs and models.

For the purpose of data integration, various onshore and offshore data sources[10, 12] are described. The *shelf, slope* and *deep-basin* geological data occurrences of the continental basin margin are characteristic in any geological settings, where large volumes and varieties of data exist. As demonstrated in Fig. 1a, the shelf, slope and deep data events of the continental basin margin are typical in any geological setting and petroleum system development, where volumes and varieties of data exist. A scope of connecting and integrating the Big Data through domain ontologies is explored. Figs. 1b and 1c represent the digital data of the Western Australian upstream business assets. The data modelling aspects, including the description of domain ontologies of the multiple dimensions, are depicted.

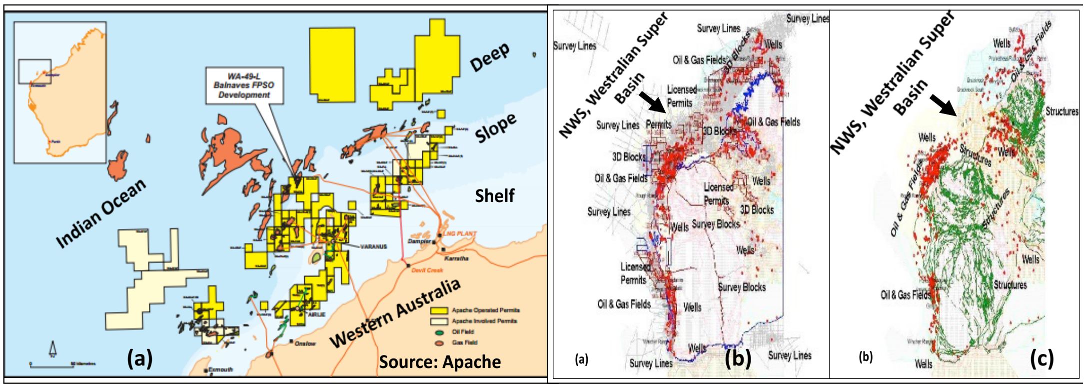
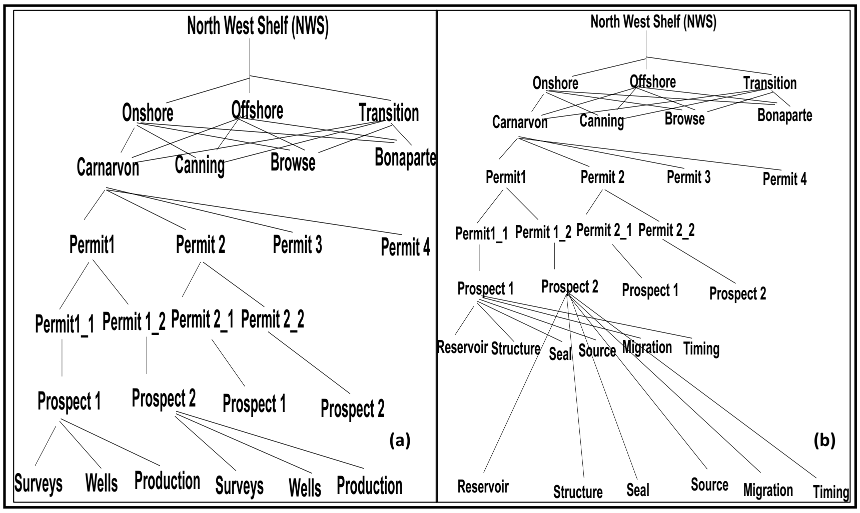


Fig.1: (a) *Shelf*, *slope* and *deep* (geological) data events of the NWS (b) Western Australian assets (c) Oil & gas fields

The description of NWS and the significance of its upstream business data science are discussed in Section 2. Description of ontologies and modelling methodologies are given in Section 3. In the context of NWS in a Bid Data scale, the data mining and visualization models are discussed in Section 4 with certain conclusions and recommendations in Section 5.

2. North West Shelf (NWS) and the upstream business data science

The collective strength of the *structure, reservoir, source maturity, seal,* and *migration* of hydrocarbons, the *timing* of deposition including their *accumulations,* constitutes the existence of a petroleum system in a sedimentary basin. They are characteristically the elements and processes of the petroleum system. The total package of the petroleum system consists of various elements and processes with a variety of their data instances and strengths. The existence of petroleum systems is ensured, and if any of the elements and or processes is missing, it is impossible to expect the oil/gas in that basin. The explorers invest in the upstream business, ensuring the presence of petroleum system and a comprehensive knowledge of data of all elements including the processes of the petroleum system. We attempt to envisage the petroleum system in a different perspective using ontology-based data warehousing and data mining technologies[10], for exploring the capability of the system. In the present study, all the elements and processes are considered as classes of a super class *petroleum system*. In the "exploration" super class, the *surveys*, *wells* and *permits* are the other representative classes. In the E & P class representation, the *production* is a key sub-class. In prospect analysis viewpoint, conceptual models are built, creating data relationships through ontological descriptions. Integrating domain ontologies of representative dimensions of multiple domains is a requirement, as demonstrated



and exhibited in different hierarchies in Fig. 2.

Fig. 2: Constructs showing (a) basin – field (oil/gas) (b) prospect connectivity in the North West Shelf (NWS) system

The Mesozoic intracratonic basin, developed as a rift/drift margin of approximately 2400km long and 140km wide, occupies a large part of NWS in the western part of Australia[10]. The Northern Carnarvon, Offshore Canning, Browse and Bonaparte are sub-basins of the NWS. The large undeveloped gas fields (in the order of 88tcf) that exist in the NWS are the focus of E & P in early 21st century. The most major nine undeveloped fields form the focus of further exploration in these basins. The Gorgon (North Carnarvon), Scott Reef (Browse), Sunrise (Bonaparte), Troubadour (Bonaparte), Brecknock (Browse), Scarborough (North Carnarvon), Bayu-Undan (Bonaparte), Chrysaor (North Carnarvon) and Dionysis (North Carnarvon) are typical fields. Besides, many undiscovered and unaccounted reserves, risk-free, particularly from deep-water prospects are present in the NWS[10]. Huge geological structures are yet to be explored. A proven Paleozoic hydrocarbon-charge system exists beyond the limits of current oil discoveries, leaving large areas of exploration including the field development opportunities of the Petrel subbasin of the Bonaparte basin. In the frontier areas, the existence of reservoir, source maturity and seal effectiveness, their combination is advantageous. An exploration risk, if it exists, it is because of lack knowledge on limits and boundaries of the linked petroleum systems. In spite of risks, recently proven discoveries are western margins of Barrow-Dampier sub-basin (in Western Australia). However, the underexplored, huge deep-water area is enormous, and untested in the Northern Beagle Sub-basin of the Carnarvon Basin, including offshore Canning and Browse basins.

3. Description of ontologies for NWS petroleum systems

The ontology conceptualization, which is an unambiguous and suitable representation of petroleum system categories of the NWS, is described in diverse application domains, moving away from the domain of artificial intelligence (AI) to desktops of domain experts[15]. Ontologies turn out to be very conversant on the World Wide Web (WWW)[1] in many domain applications. The present work has a scope of developing a Petroleum Data Description Framework (PDF), petroleum ontology, encoding the knowledge and semantics of petroleum data on web pages, to logically connect to the electronic agents that search for facts on the E&P. For the knowledge builder that extracts knowledge of the petroleum systems[4] efficiently, semantic and schematic based ontologies bring together the data conceptualizations and contextualization. Large taxonomies, categorizing the petroleum systems based on classifications of basins, each basin having various oil & gas fields, each field has one or more drilled wells with seismic vintages, and the description of their relational hierarchies are parts of the petroleum ontologies[10]. The attribute dimensions and fact instances are organized in relational and hierarchical ontological structures. For standardizing the ontologies, the domain experts, oil & gas explorers and data analysts use, reuse, share and annotate the petroleum data and information from basins and oil & gas fields. The existence of system's elements and processes that interpreted as attribute instances is characterized as classes and sub-classes in ontological perspective, initiating knowledge base constructs and dimensional models.

3.1. NWS ontology framework

In the continental basin margin areas, the shelf, slope and deep are geological events interpreted as classes and sub-classes in the respective domains of petroleum systems, besides linking to surveys, drilled wells, permits and production entities. These events that cannot be secluded are made interconnected through constructs and models. For understanding the connectivity, the basin histories and prospectivity in large areal contexts, the element and process classes and sub-classes and their linked data are integrated. The shared ontologies[8, 9, 10, 15], described in different domains handle terminologies, vocabularies including semantics while integrating the constructs and models. For facilitating the data integration process, shared ontologies that include machine-interpretable definitions of basic classes (concepts) in various domains, their one-to-one and one-to-many data relationships are developed. For examining the domain knowledge in a wider context of the NWS, data mining, visualization and interpretation artefacts are considered, interpreting entities or dimensions and their attributes. For using, reusing ontologies and extending them to other applications of the NWS basins, formal analysis of the entities and dimensions is valuable.

The pictorial digital data and conceptual models represented in Figs. 1 and 2 are used in the NWS data modelling for a unified metadata. In many health science applications, the concepts of data integration and shared logical data structures have proven successful[12]. In the E & P domain scenarios, the Canning Basin, a part of NWS has exploration and development challenges because of lack of positive domain knowledge. Though the adjoining Carnarvon basin establishes high-quality petroleum discoveries, it is vital to delineate the areal extents of the elements and processes of both the basins in their respective geographies. Ontology approach facilitates connecting the basins and petroleum systems, relating its classes[5], sub-classes and adding more facts to the existing semantics and schematics.

As an explorer, one has to adjudge the context, the ontology[9] used for, how specialized or generalized ontology is going to be, including conceptualization. The constructs and models are developed representing the NWS ontology context. For integrating the NWS contexts, development of a data warehouse is planned, ensuring the data consistency in multiple domains. All viable options are assessed and the best option that can work better predict the problem solutions with more intuitive, more extensible, and more maintainable, are analyzed. The ontology is a replica of the illustrative world of realism, and the concepts in the ontology must reveal the veracity. The initial ontologies are evaluated and debug by problem-solving methods, updating the initial ontologies. The iterative process design is concluded with the total lifespan of the ontology and its improvement. The stages and the progress of NWS ontology framework are:

3.2. Determine domains with the ontology opportunities

The domain and its choice are vital criteria in the ontology improvement, besides responding to simple questions:

- For which domain, the ontology rules well? - any coexistent domain of NWS
- For which purpose, the ontology use intended? – domain descriptions, classification and structuring
- Whether the ontologies should offer solutions and for what type of uncertainties?
- Whether the links between manifold reservoirs and their categories of oil and gas fields can facilitate new knowledge? - firm up the prospectivity, minimizing the risk of exploration & production
- Who uses and maintains the ontologies? – data analysts, ontology designers and oil and gas explorers

The notion of good and bad combinations is sorted out in the ontological development based on the description of the classes that describe diverse petroleum elements in major sub-basins. The ontologies include the classes of the inventory of petroleum systems, considering the surveys, wells, permits and production classes that related to other classes of petroleum systems. They include synonyms and semantic information for other classes of the ontology, facilitating the natural – language processing of petroleum systems and their investigation in the PDF. The *production rate* of a unique well (in a permit area) representing such element must be incorporated if the ontology helps the explorers or geoscientists to agree which petroleum system element has dominance in the framework. The basin *acreage* and *potentiality* attributes may be needed to include, if a concept or class is used in the E & P domain. The domain is described in a model (language) that is different from the type (language) of ontology users' usage, providing the required mapping constraints between the models (languages), maintaining the specialized ontologies.

3.3. Interoperability

The oil & gas explorers and ontology designers must look for alternate options of problem solutions or restraint the existing data solutions if the contextual domains need stretching or other alternatives. Use and reuse of the existing ontologies are must if one system demands interaction with applications of other petroleum systems of the adjacent basins that previously committed to specific ontologies or an orderly semantics. Ontologies are automated and imported into the PDF. Whichever formalism is expressed, many knowledge base systems can import or export the ontology in multiple contexts. The development of petroleum ontology[10] may be initiated right scratch if there are no current ontologies for petroleum systems of NWS, in which case, the scope of new constructs and models is explored.

3.4. Enumerating vital classes in the ontology descriptions

All the relevant terms (concepts or classes[5]) are listed to develop or improve an existing ontology. What are the classes and what properties do they have? The attributes such as the *name of the basin* and *names of petroleum system elements and processes* are documented. Under surveys category: *survey id*, *survey area*, survey type, type of exploration; well category: *well id, well name, type of well*; permits category: *permit id, name of permit*, *type of permit*; production category: *production id*, *type of production*, *production rate* and *pressure*, if sub-classes or concepts are available, are all itemized. At first, a comprehensive list of all super-classes is built, including their classes and sub-classes, despite having overlapped between classes or concepts and their relationship classes. The attribute dimensions may be concepts, or the concepts may themselves be classes or slots. Other steps are – class hierarchies and properties of concepts (slots) – are interconnected. Typically, limited concepts (classes) are described in the hierarchy with properties described for every class. In the ontology-design process[8, 9, 15], these two steps are vital. The other design rules are discussed here:

3.5. Defining classes and their hierarchies

The approaches in developing class hierarchy are[9, 15]:

1. To start with, the most general classes in a particular domain are defined as a *top-down* development process and successive knowledge-based classes (concepts). For example, equivalent classes are created for a broad range of concepts of a petroleum system in a basin. Onshore, offshore and transition zones are specialized sub-classes of the *basin* class. In a permit super-class, exploration permit, well approval permit, production permit or pipeline permit are sub-class categories. In the broader view of NWS petroleum system, it is a super-class, and it is characterized into petroleum systems of the Carnarvon, Canning, and Browse and Bonaparte basins as subclasses. Further, the prospect level subsystems of the Carnarvon basin's petroleum system are categorized as sub-classes.
2. Grouping of classes is done in more general classes (concepts), defining the specific classes in a hierarchy as a bottom-up construction process. For example, a sub-class interpreted as the Gorgon or Scott Reef petroleum prospect belongs to the petroleum system of the North Carnarvon Basin as a general class. The drilled well that has an oil well, gas well and condensate well subclasses is another demonstration that belongs to general class "wells". An oil well has a specific geological formation, which is again a subclass.
3. The combined views of the top-down and bottom-top approaches allow us developing a combined method. Initially, additional noticeable classes are defined, for generalizing and specializing appropriately in multiple domains. Few specific classes or concepts are included in top-level classes, which in turn are linked to a middle-level concept. Thus many middle-level concepts or classes are connected to the regional level classes.

The individual requirements and the needs of the classifications are appropriately acknowledged for each domain. Nevertheless none of these methods is integrally applicable in all practical situations[8, 9]. The top-down view of the domain is more systematized to make easy use of data events. Since the concepts in the "middle view" tend to be more descriptive in a domain, the combined approach is often simplest for ontology developers. The top-down approach may be more realistic while distinguishing the most general classification. Whereas the bottom-up approach is more suited when explicit examples are specified. Initially, the concepts are defined as classes in ontological descriptions and become the anchors of the class hierarchy, whichever approach is chosen. An instance of one class may necessarily be an instance of other classes, in such a way, the classes are organized in a hierarchical taxonomy. Every instance of B belongs to A, when a class of A is a superclass of B.

3.6. Define properties of classes – slots

The internal structure of concepts is described when once the classes are ready with the properties. For each and every class, a property attribute is described, for example, the *quality of reservoir*, *type of structure* and *production rate* in their respective classes. The property attribute is also described tagged with that class. The property attributes become slots attached to the classes. Such numerous class properties are attributed to the entire heterogeneous petroleum data of the sedimentary basin that become slots in the ontology descriptions.

- The quality of reservoir, texture and color of geological unit, are examples of the intrinsic properties
- The well name, survey id, date of production are examples of the extrinsic properties
- If the class is structured; the “parts” can be both physical and abstract (such as components of a petroleum system, the connectivity between systems and basins.

All subclasses of a class that acquire the slots of that class and their related properties are defined. The fact instances of the slots of that class representing all sub-classes are appropriately identified and interpreted. The additional slots demonstrating the classes and subclasses including in the connecting classes are also defined without any ambiguity. Analogously, the slots are labelled to the most general class that has property instances. For example, petroleum type, its composition and production rate, its PVT attribute dimensions are attached to the production class., since it is the most general class whose instances have *PVT* and *production rate* values.

3.7. Define facets of the slots

The slots have diverse phases, each is characterized by a value type, and permitting value instances, the number of cases (cardinality) and other features of the instances, the slot can engage. For example, the value of a name slot (as in "the name of the reservoir") has one string. For example, the name is a slot with value type string. A slot (as in "a basin yields petroleum from these reservoirs") can have multiple value instances, and they belong to the class "Basin". A slot is generated with value type instance of “basin” as allowed in that class. Besides, several common facets are designated here:

- The cardinality controls the number of values that a slot can hold and store. The systems are characterized based on the number of cardinalities, permitting any number of values and their instances. For example, a *tectonics* of a *basin* may be a single cardinality slot (a single tectonics process evolves a basin). The *basins* possess a specific number of petroleum systems with a precise number of elements/processes and multiple cardinality slots that exhibited in the *tectonics* process class.
- For the purpose of describing the slot values accurately, the artefacts that use the systems permit specified minimum and maximum cardinalities. A slot must have at least N values for a minimum cardinality of N. For example, the "reservoir" slot of a "basin" has a minimum cardinality of 1: each basin is made of at least one petroleum system with a variety of "reservoir". The maximum cardinality of M means that a slot can have at most M values. The maximum cardinality for the "reservoir" slot for a single variety of the petroleum systems in a basin is 1: M. Under the control of *tectonics* process, the basins are made with one or more variety of petroleum systems with multi-stacked reservoirs. Fixing the maximum cardinality to 0 is beneficial, which can show that the slot has no values in that subclass. Slot-value type*:* what type of values can fill in the slots is described. More common value types are:

- String*:* it is a tiniest instance type, characterized for slots, such as name: the value is a simple string.
- Number: it describes (both floating and integer) slots with numeric instances. For example, the sedimentary basin

produces oil at the rate of, has value type Float.

- Boolean slots: they are simple "yes-no" flags. For example, if the petroleum system exists "(yes) or not (no)" are used. Or source rocks if matured, describe "say yes or immature, say no".
- Enumerated slots*:* they enumerate a list of explicitly allowed data instances for the slot. For example, petroleum production can take on three possible data instances: oil, gas or condensate.
- Instance-type slots*:* data relationships and their descriptions between classes are analyzed. A list of allowed classes is defined by slots with value type instances. For example, a slot can come from a class reservoir that may have instances of the class basin as common values.
- Domain and range of slot*: range* is described*, when* classes allowed for slots of type instances. In current context, the class "basin" is the range of the "produces" slot. Some systems allow restricting the range of a slot when the slot is attached to a particular class. The class or classes to which the slots are attached, the properties of these slots are described in each of that domain. The "reservoir" class is the domain of the "produces" slot. In systems, where the slots are attached to classes, each slot attached to that class constitutes a domain, in which case specifying a separate domain is not required. For defining a domain and the range of a slot, the basic rules are:

1. The general class or classes are determined that respectively pointed to a domain or the range of slots without excessively generalizing them. All the classes are again interpreted in that domain and range of a slot with potential fillers in that slot. A class that fills the slots depends on the heterogeneity and multidimensionality of data, with a requirement improving the data structure grain size and its property.
2. Fairly the "basin" class is listed, instead of itemizing all possible subclasses of the "basin" class, including the range of the "produces" slot. "THING" cannot be specified to any range of the slot in a broader ontological description. A range or domain of a slot is defined including class and or sub-class with a flexibility of its inclusion or exclusion.
3. The basin and reservoir classes are in two separate hierarchical levels if the domain or range of the slot wishes to contain both classes, the reservoir class turns out to be a subclass of the basin, and consequently, the slot range inherently takes in all other subclasses of the "basin" class. The range and domain of the slot, defined in classes, contain all subclasses of class A, but not the class itself. But the range is ought to contain just the class A and not necessarily all subclasses.
4. In current context, a poor reservoir, no reservoir and such range is limited to the class reservoir itself, rather than defining the range of the slot to accommodate to a good reservoir. A more appropriate range is made while considering the class A, including classes, defining a range or domain of a slot that contains all but a few subclasses of class A. Similar business rules are applied in building reservoir models in a basin, attributing a slot to class, same as linking the class to the domain of the slot. In other words, the whole pattern is made global, ensuring all subclasses are attached to the class including their associated slot properties.
5. Fine grain data structuring that narrates the property slot can be attached to each of the reservoir classes or classifications, representative of poor or no reservoirs. The slot is committed to a more general class category "poor reservoirs", since the "poor reservoir" is narrative of fine-grain[9, 10] of data structuring that made up of atomicity property. Simplifying the fine-grain property slot may not necessarily be precise in a broader context of "basin" class because it has multitude of entities and or dimensions with varied attribute instances.

3.8. Create instances

In the preceding stage of the modelling process, individual instances of classes are created in the hierarchical ontologies[4, 10]. Selecting the class, with distinct instances of that class attaching to slot values, are required in outlining a particular instance of a class. A demonstration of a specific type of reservoir existence, for example, is a justifiable individual instance of Birdrong Sandstone- in a Barrow Sub-basin- Permian - North Perth Basin subclass and class criteria. This instance has the slot values:

- Basin/Field: Carnarvon/North Rankin
- Field_Size: 500sq km

- Reservoir_Quality: Excellent, High
- Type_Reservoir: Sandstone, Siltstone
- Porosity (%): 10-30%
- Production: 500 million barrels
- NWS_Basin_Name: Carnarvon
- Production_LifeSpan: 25 Years (based on the existing known reserves)

3.9. Defining classes in a hierarchy

Various classes and hierarchy of classes are described here. In any chosen domain, there may be many ways of representing classes, subclasses with their hierarchies. The hierarchies emerge based on the description and use of data relationships, level of details needed in the hierarchies including the compatibility of the other models and participating classes, subclasses and slots in the integration process. For generating a class hierarchy, certain basic guidelines are followed. A considerable number of new classes is generated, conforming the hierarchical ontologies.

IS-A is a designator, represented in generalizations and adaptable to E^2/R modelling as shown in Fig. 3. Various external schemas are evolved from conceptual schemas (as shown in Figs. 3 and 4) as demonstrated in E^2/R models for implementation[10, 15]. While integrating two specialized models into a generalized model, several inconsistencies may have arisen, in which case, the extended entity models (E^2/R) make them more consistent. Besides, the basic ER model is extended to include the super type/subtype data relationships and remove the inconsistencies.

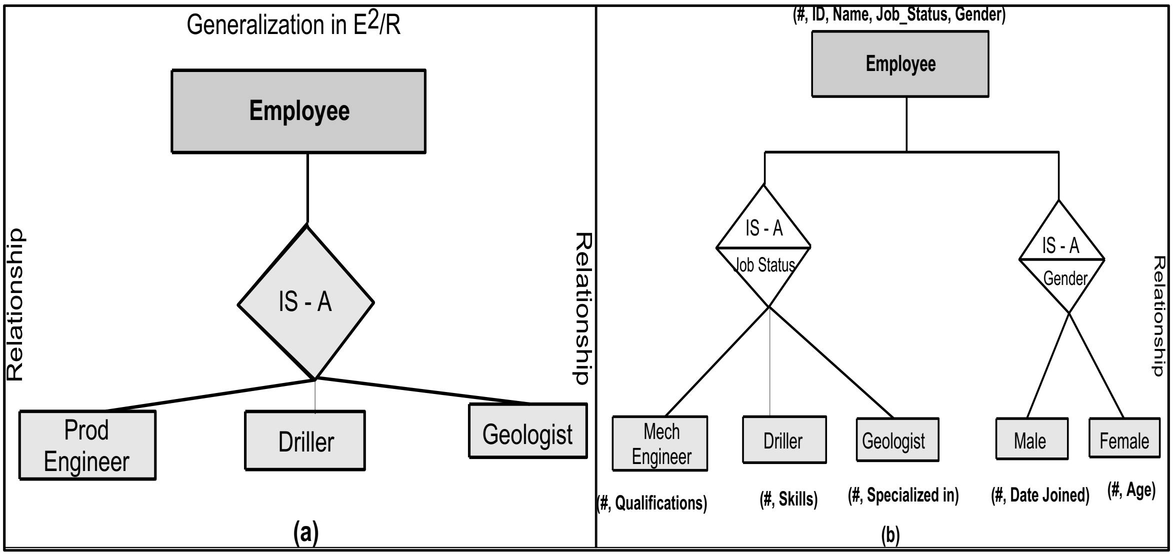


Fig. 3: ER Constructs (a) business entities showing generalization and (b) its properties with multiple classifications in E^2/R

3.10. Relationship with the extended multidimensional (M^2/R) attributes

With the arrival of new multidimensional tools[10], database modelling and its extended data warehouse version have become more refined. Additional innovative and new forms of abstractions are introduced in the conceptual modelling, keeping view the diverse applications and domains in the E & P. The complex upstream business scenarios and oil & gas company metadata can mend to more sophisticated abstract data types.

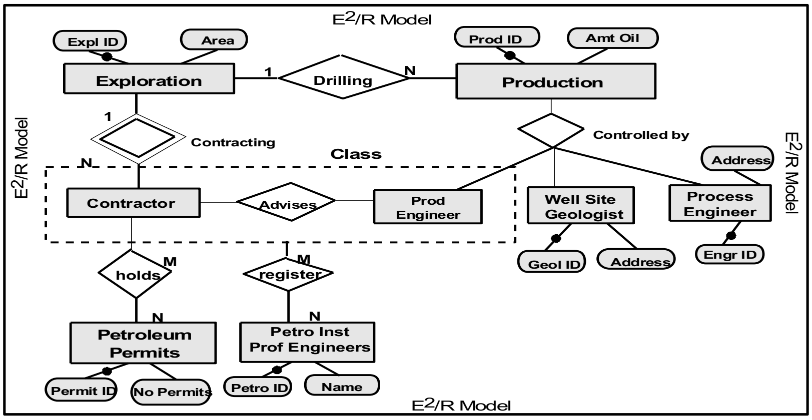


Fig. 4: A conceptual schema in E^2/R (for E&P Oil & Gas Company)

3.11. Relationship describing "is-a"

"Is-a" relationship is an expression defined for a hierarchy. B is a subclass of class A, each instance of B belongs to A. The Barrow is a sub-basin of the Carnarvon basin, in which the "reservoir" is a subclass with a slot property "coarse-grain", which is a taxonomic relationship[14]. Shared domain ontologies are integrated through an integrated framework[10, 11, 12]. Ontology designers and upstream business analysts add values to the existing knowledge of a field, petroleum system and basin in such hierarchies.

4. Data mining and visualization

Taking into account the users specific data for interpretation, different data mining schemes[2, 10] are used. They include collective tasks of categorizing data associations, classification, regression mapping, time series analysis, prediction, clustering, summarization, associative rules, and sequence patterns discovery from data warehouses. Since a single mining scheme generates anomalies or ambiguities in the interpretation, for making data mining more effective, mining rules are combined. This improves the visualization of the data views, used for interpretation.

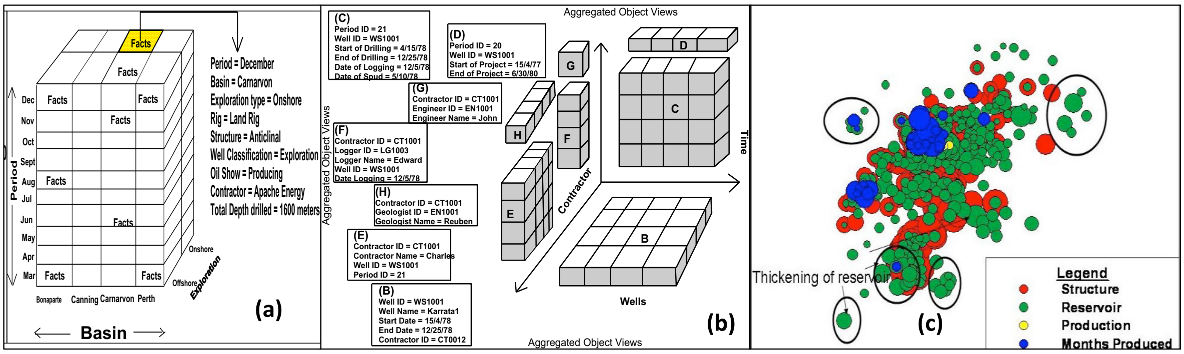


Fig. 5: (a) The cuboid metadata (b) aggregate data views (c) bubble plot views

5. Results and discussions

The bubble plot view is a simple example of the use of graphics[4] to quickly convey information about the data what they present for interpretation. The data mining needs to be creative when data views extracted from ontology-based warehouse repositories are meaningful. For accessing the aggregate data and map views, several SQL queries are run on the warehoused-metadata cubes. The bubble plots represent graphical or visual views (Fig. 5), with scores of numerical and textual data. New knowledge on reservoirs, structures and production trends is interpreted and analyzed for prospect evaluation. The bubble plot, representing bubble sizes, densities and orientations, draws several conclusions such as strengths of geological structures, reservoir and production attributes and their magnitudes, from which qualitative and quantitative properties (of reservoirs, for example) are interpreted for prospect evaluation in the in the Big Data scale. The methodology appears flexible handling multidimensionality and heterogeneity.

6. The conclusions and recommendations

The application of ontologies in the current domain is strategic, benefiting the petroleum systems and the interpretation of their elements and processes is promising. Besides, we ensure favorable combinations of different geological settings and basins for judicious application of ontologies in the upstream business data research. The knowledge base data models deduced from unknown geological regimes of the WA upstream businesses and their data sources substantiate the integration of domain ontologies in a data warehouse environment. Similar studies are recommended for investigating many unexplored or underexplored *sedimentary basins* that can deliver more productive and commercial oil & gas fields in Australia. The E & P Ventures are accordingly prioritized, based on categories of fields and basins for investment. In the areas, where petroleum ontologies are described, investments may be promising. If ontology-based data models are judiciously applied in the NWS, the benefits and returns on investments are immense.

7. Future scope and outlook

New Big Data opportunities are explored, extending the robust data modelling methodologies in areas where unconventional reservoir systems and hydrocarbon depletions are expected. A regional ontological model is planned for the entire NWS basin tectonics in Big Data scale, which can be beneficial to oil & gas explorers, upstream business researchers and data analysts in downsizing the costs and adding values to the investments.

References

1. Barrett TD, Jones JD, Yuan J, Sawaya M, Uschold T, Adams D, Folger D. RDF Representation of Metadata for Semantic Integration of Corporate Information Resource. In Proceedings of Real World RDF and Semantic Web Applications Workshop held in conjunction with WWW-2002.
2. Berson A, Smith JS. "Data warehousing, data mining & OLAP", Mc Graw – Hill Education (India) Pty Ltd, 2004, pp. 205-219, 221-513.
3. Biswas G, Weinberg J, Li C. ITERATE: A conceptual clustering method for knowledge discovery in databases, *Artificial Intelligence in the Petroleum Industry*, Paris, France, 1995, pp.111-139.
4. Castañeda Gonzalez O. J, Nimmagadda SL, Cardona Mora AP, Lobo A, Darke K. On Integrated Quantitative Interpretative Workflows for interpreting structural and combinational traps for risk minimizing the exploratory and field development plans, presented and published in the *Bolivarian Geophysical Symposium* proceedings, held in Cartagena, 2012, Colombia.
5. Chaudhri AB. Object Database Management Systems: An Overview in "BCS OOPS Newsletter", 1993, No.18 Summer '93, USA.
6. Cleary L, Freed B, Elke, P. "Big Data Analytics Guide: 2012", Published by SAP, 2012, CA 94607, USA.
7. Damiani E. Key note address on 'Digital Ecosystems: the next Generation of Service Oriented Internet", IEEE-DEST, Phitsanulok, Thailand, Feb 2008.
8. Fernández López M. Overview of Methodologies for Building Ontologies, Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999.
9. Gruninger M, Uschold, M. "Ontologies and Semantic Integration," to appear in Software Agents for the Warfighter, the first in a series of reports sponsored by the US Government Information Technology Assessment Consortium (ITAC). Edited by Jeff Bradshaw, Institute for Human and Machine Cognition (IHMC), 2002, University of West Florida.
10. Nimmagadda SL, Data Warehousing for Mining of Heterogeneous and Multidimensional Data Sources, Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG, 2015, p. 1-657, Germany.
11. Nimmagadda SL, Zhu D, Rudra A. Knowledge Base Smarter Articulations for the Open Directory Project in a Sustainable Digital Ecosystem, WWW 2017, Perth, WA, Australia.
12. Nimmagadda SL, Rudra A. Managing the Embedded Digital Ecosystems (EDE) using Big Data Paradigm, ed. B. Kei Daniel, Big Data and Learning Analytics in Higher Education, DOI 10. 1007/987-3-31906520-5_5, Springer International, Switzerland 2017.
13. Pujari AK. "Data mining techniques", University Press (India) Pty Limited, Hyderabad, India, 2002.
14. Raj RK, Ramesh SR. COMPUTATIONAL ONTOLOGIES AND INFORMATION SYSTEMS: I. FOUNDATIONS, Communications of the Association for Information Systems (Volume14, 2004)158-183.
15. Uschold M, Gruninger M. Ontologies and Semantics for Seamless Connectivity, SIGMOD Record, Vol. 33, No. 4, December 2004.