Journal of Business Research 89 (2018) 143–158

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

On big data-guided upstream business research and its knowledge management

Shastri L. Nimmagadda[a,*], Torsten Reiners[a], Lincoln C. Wood[b]

[a] School of Management, Information Systems Discipline, Curtin Business School, Curtin University, WA, Perth, Australia
[b] Department of Management, University of Otago, Dunedin 9054, New Zealand

ARTICLE INFO

Keywords:
Upstream business
Heterogeneous and multidimensional data
Data warehousing and mining
Big Data paradigm
Spatial-temporal dimensions

ABSTRACT

The emerging Big Data integration imposes diverse challenges, compromising the sustainable business research practice. Heterogeneity, multi-dimensionality, velocity, and massive volumes that challenge Big Data paradigm may preclude the effective data and system integration processes. Business alignments get affected within and across joint ventures as enterprises attempt to adapt to changes in industrial environments rapidly. In the context of the Oil and Gas industry, we design integrated artefacts for a resilient multidimensional warehouse repository. With access to several decades of resource data in upstream companies, we incorporate knowledge-based data models with spatial-temporal dimensions in data schemas to minimize ambiguity in warehouse repository implementation. The design considerations ensure uniqueness and monotonic properties of dimensions, maintaining the connectivity between artefacts and achieving the business alignments. The multidimensional attributes envisage Big Data analysts a scope of business research with valuable new knowledge for decision support systems and adding further business values in geographic scales.

1. Introduction

Upstream business data science and its analysis is the current motivation of business research. The existence of several volumes and varieties of either online or offline data events motivates the present business research. The heterogeneity and multidimensionality challenges of spatially and periodically varying resource data are examined in upstream businesses. Resource business data refers to data sources associated with oil and gas and mineral and mining entities, which are often spatial-temporal. The periodic dimension consists of composite attributes such as *day*, *month*, *quarter* or *year*. Managing and delivering accurate information in various operational units are challenging tasks for making timely business decisions. From exploration to production and marketing phases, volumes and variety of data exist with information workflows (Agarwal & Dhar, 2014) among various operational systems. The issues (Castañeda Gonzalez, Nimmagadda, Cardona Mora, Lobo, & Darke, 2012) associated with the existing tools and technologies of business data science require a large amount of logical data storage to address the heterogeneity of the data amassed among hardware and software platforms. Flexibility to construe multiple dimensions and their attribute instances can be envisioned through the data structuring stage.

Within the resources industry, exploration and production data dimensions are typical. Operational data, updated operational data, archived and external data, and unstructured data are characteristic in the resources industry. In addition, information on spatial-temporal, non-geometric, geometric to non-geometric and fully geometric spatial dimensions (entities/dimensions/objects) is representative and critical to geographically varying information systems, particularly with mining and oil & gas businesses. This information is critical in describing the heterogeneity of and minimizing the ambiguity involved in the use and reuse of the knowledge-based constructs and models in the exploration business and reducing the risk of losing crucial knowledge of the resources industry.

In the present business research, the contribution is organized with the description of Big Data characteristics (Debortoli, Muller, & Brocke, 2014; Dhar, Jarke, & Laartz, 2014; Kudyba, 2014) in an Australian resources industry and how the proposed methodologies can handle both real-time and historical data heterogeneity and multidimensionality challenges. In Section 2, the issues, problems and motivation of the current business research are described. The research questions and objectives are outlined in Section 3. The significance of the research and the research contribution are provided in Section 4. The data modelling methodologies and the integrated framework are given in Section 5. The results and discussion with evaluable utility properties of the constructs and models used in the framework are

* Corresponding author.
E-mail address: shastri.nimmagadda@curtin.edu.au (S.L. Nimmagadda).

https://doi.org/10.1016/j.jbusres.2018.04.029
Received 22 September 2016; Received in revised form 21 April 2018; Accepted 23 April 2018
0148-2963/

provided in Section 6. An online business data scenario is discussed with a motivation for increasing its commercial research and trade value. The data mining, visualization and new business knowledge interpretations are provided with a decision tree mining model, evaluating the articulations through utility properties (Venable, Pries-Heje, & Baskerville, 2016). Conclusions and recommendations are given in Section 7.

2. Issues, challenges and motivation

In response to chaotic business data management situations in industries, new data management, storage and data processing capabilities have been investigated with restrained spatial-temporal dimensions. Previous studies (Castañeda Gonzalez et al., 2012; Schermann et al., 2014) suggest that the data warehousing and data mining technologies with Big Data contexts (Cleary, Freed, & Elke, 2012; Pyne, Rao, & Rao, 2016) are new direction in solving problems in the resources industries. The early research performed on data modelling, data warehousing and mining methods in various industry applications have scalability and flexibility issues. Gornik (2000) provides a classical example of data warehouse design and implementation in an airline industry. Meersman (2004) describes ontologies with a case study in the petroleum industries. Dimensional data exist in these businesses, but in bits and pieces, with diminishing data quality, and they are sometimes incompatible with the data access protocols of various software and hardware platforms (Nimmagadda, 2015). Operationally, it has been challenging to find the precise data and information required for exploration and field development tasks, including information on drilling and production entities. Structuring the data with periodic and geographic dimensions is another issue for which more joins are required to bring the query results into a single table (Moody & Kortink, 2003). Digging large volumes of data is a compelling motivating factor (Sagiroglu & Sinanc, 2013) with challenging heterogeneity of exploration and production data. In every data model, the periodic dimension (online vs. offline) is allowed in structuring and integrating the data structures through multiple attribute dimensions. Users may be interested in both the current and historical data (Longley, Bradshaw, & Hebberger, 2001; Li, 2010), and their structures may vary with business rules. For example, it may be necessary to record or document the real-time petroleum permit status or its history. Is it necessary to record and document the current prices of all petroleum products or exploration costs and their monthly, quarterly or yearly cost fluctuations for decision support systems?

The periodic description of bubble plot views drawn from multiple data sources and shown in Fig. 1a demonstrates the quantum of upstream business data available in the current research. The periods at which the attribute *number of drilled wells* is reported to have trends similar to those of active seismic and petroleum permit campaigns at peak periods designated I, II, and III in Fig. 1a. At similar periods, the data sparsity, heterogeneity, multidimensionality and granularity are observed with increasing and decreasing bubble sizes and their densities. In a 2D bubble plot, the diameter of each bubble varies in size, providing a means of representing and analysing an additional dimension of data. Evaluating the Big Data characteristics in periodic and geographic dimensions and interpreting their varying attribute strengths (with bubble sizes) are key criteria and foci of business data science. As demonstrated in Fig. 1b, the data characteristics signify more data issues, for which a holistic mapping and modelling approach is needed. The periodic bubble plot views represent the data volumes with varieties and their anomalous data instances, as examined in Fig. 1c. The geographic areal extents presented with an extractable data value in percentage instances motivate researchers to create business values in geographic contexts. Hundreds of such dimensional attributes are identified and, at places, conceptualized and contextualized for modelling and building connections between attributes and their structures. In geological and geophysical (G & G) surveys (Nimmagadda, Dreher, & Rudra, 2015), millions of *point*, *line* and *areal contour* data instances that vary with space and time-period are described. For example, data events associated with seismic and drilled-well entities rapidly change with spatial-temporal dimensions.

There is need of a centralized multidimensional warehouse repository that can motivate the working centres, with the collaboration of teams locally and globally, so that they can access the relevant quality information with ease. The processed data are shared and accessed by multiple users for making timely business decisions (Gregersen & Jensen, 2002). Although the approach has a broader scope, the present study is limited to the design and development of data models, addressing the issues associated with spatial-temporal dimensions and implementing the models in the upstream business research.

3. Research questions and objectives

The research questions and objectives are framed as follows, keeping in view the issues, challenges and motivation:

1. How are the spatially and periodically varying business data attributes (either online or offline) structured, considering the volumes

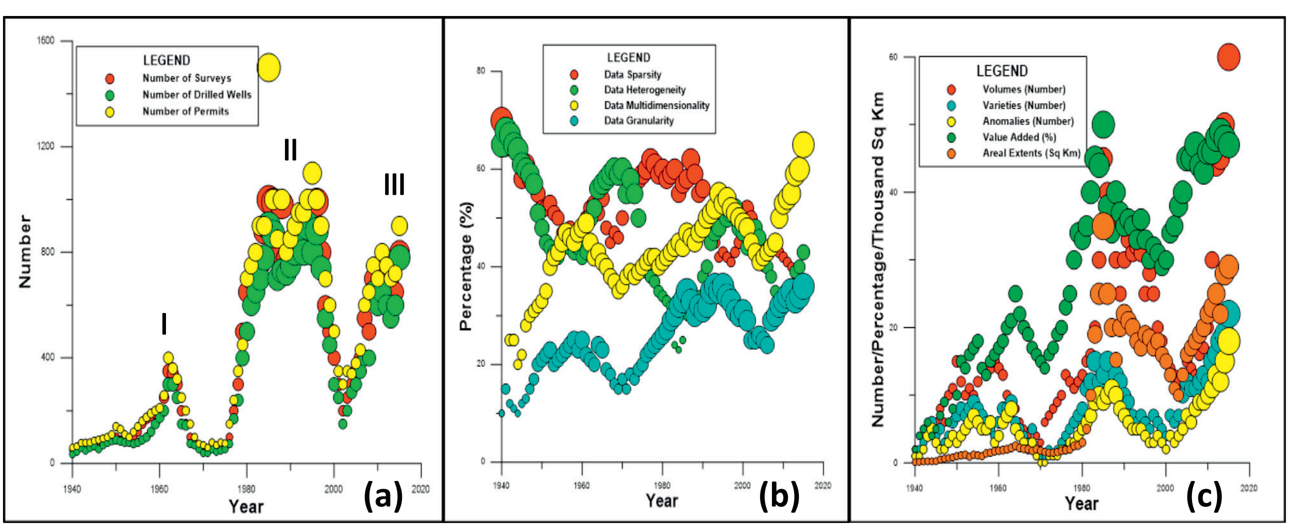

Fig. 1. Bubble plots of *the number of oil and gas wells*, *survey* and *period* attribute dimensions.

and varieties of data?
2. How can each system connect to other systems in an upstream business in a data warehouse repository perspective?
3. Keeping in view the heterogeneity and multidimensionality, how can the data be integrated?
4. How are the connections explored from the metadata? Can the correlations, trends and patterns of data extracted from metadata be visualized and presented for interpretation?
5. Are the proposed methodologies better for organizing, analysing and interpreting the spatial-temporal data in a Big Data focus? How are they significant in the business research?
6. How can robust methods add value to resource business projects?

Objectives such as the performance of databases, ease of accessing data and use of warehouse repositories in the resources industry are other foci. In this context, adopting the Big Data paradigm (Agarwal & Dhar, 2014) ensures the following:

1. The development of data models that distinguish real-time, instantaneous and historical data documentation issues;
2. The description of how periodic attributes affect the cardinality of relationships in resource business data;
3. The diagnosis of how changes in business rules influence data modelling and integration;
4. The presentation of spatial-temporal data in cuboid structures for knowledge building;
5. The analysis of how the Big Data characteristics impact the overall business research.

The business activities and functions of resources industries are diverse, especially in the Australian business contexts, where several associated business chains need an alignment. The objectives of the current research are designed while keeping in view the data volumes and varieties of the upstream project. The size of data instances at a petabyte scale can make the distributed exploration and production businesses more competitive and challenging.

In addition to various attribute dimensions in multiple domains that affect the portability of the constructs and models, business rules and constraints impact the model flexibility criteria. Even geographic data dimensions that characterize the constructs may pose data integration issues in conjunction with periodic dimensions. Such critical factors can guide business researchers while articulating the methodological framework and its associated data models, including data mining, visualization and interpretation artefacts. The size and scale of Big Data motivate the veracity of the construct design and development. Especially with the historical data, the model granularity is likely in control of periodic dimensions and their atomicity.

4. Significance of the upstream business research

Hundreds of attribute dimensions and their fact instances of upstream business transactions accumulate online daily. They are again either structured or unstructured in multiple domains, complicating the data integration process while building warehouse repositories.

As shown in Fig. 2, several such type- and sub-type dimensions exist within Big Data characteristics, generating hundreds of conceptualized and contextualized data anomalies between dimensions that make up the connectivity between attribute dimensions. In the present study, the warehoused data have the following additional significant dimensions:

a. Longitudinal dimension - time-variant business data
b. Lateral dimension - cross-sectional data from several sedimentary basins, each being thousands of sq·km in area and/or data from several business organizations.

The additional dimensions depicted in the data warehouse schemas

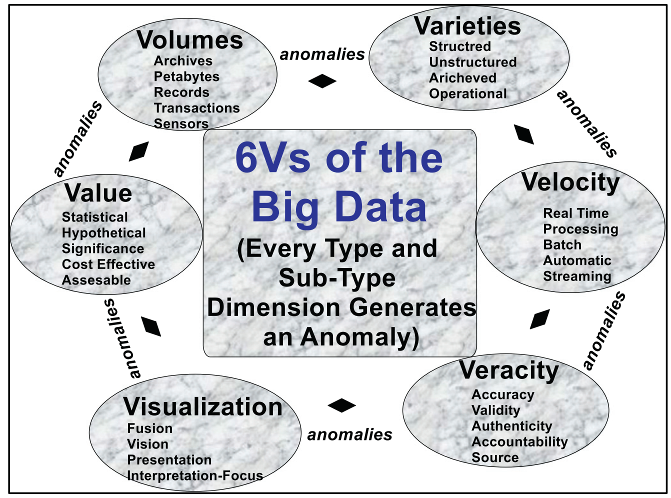

Fig. 2. Description of Big Data characteristic type and sub-type dimensions.

are intended to explore the scope of data mining, visualization and data interpretation. Time series data represented in a 4D time-lapse (Dodds & Fletcher, 2004) domain (spatial-temporal) can significantly minimize risk when making crucial financial decisions. The technologies provide immense business value in both private and public sectors of petroleum industries. There is an opportunity to extend and apply the technologies in multinational companies. What these industries have in common is the need to analyse extremely large datasets, produced and collected by a diverse range of methods, to understand particular phenomena (Mattison, 1996). As in the scientific field, an effective means of analysing a variety of data is to use visualization and summarize the data, highlighting the trends and patterns in such a way as to interpret the data for new business knowledge.

By geographic location, the state of Western Australia (WA) is the largest producer of minerals, oil and gas deposits among all the other states in Australia. The economic growth of Australia (Longley et al., 2001) depends on the exploration and exploitation of WA's natural resources (Dodds & Fletcher, 2004). However, the exploration data from the states of Victoria, Queensland, New South Wales, South Australia, the Northern Territory and WA are considered in the data integration process. The existing data models are compared (Debortoli et al., 2014; Dodds & Fletcher, 2004; Nimmagadda, 2015) for assessing the viability of exploration and the interoperability of models in the resources industries.

Typically, representing the spatial-temporal data trends in visual analytics is vital for increased understanding of business growth geographically (Cleary et al., 2012). The information needed for data science and the speed at which the warehouse repository can deliver query results further motivate business researchers. For example, *geologists* post formation tops (geological attributes) for immediate use by reservoir engineers to calibrate their models and test the *date* information in the production department. Business managers graphically visualize and investigate the neighbourhoods of drillable-well locations explored at different periods. The exploration staff optimizes their data interpretation values by knowing what seismic lines have the most appropriate acquisition and processing abilities to deliver quality business outcomes, and they use their results to plan new surveys, when the existing surveys are found inappropriate (Dodds & Fletcher, 2004) to use. Exploration data scientists may need to access the drilled-well data that are acquired in different periods to facilitate the extraction of new domain knowledge from production histories. Similarly, drilling staff may interpret the exploration data to cognize the geological information qualities and their depths in a total field area. Exploration and drilling staff need information on *periodic production rates*, quality of *crude petroleum* attributes and their data instances.

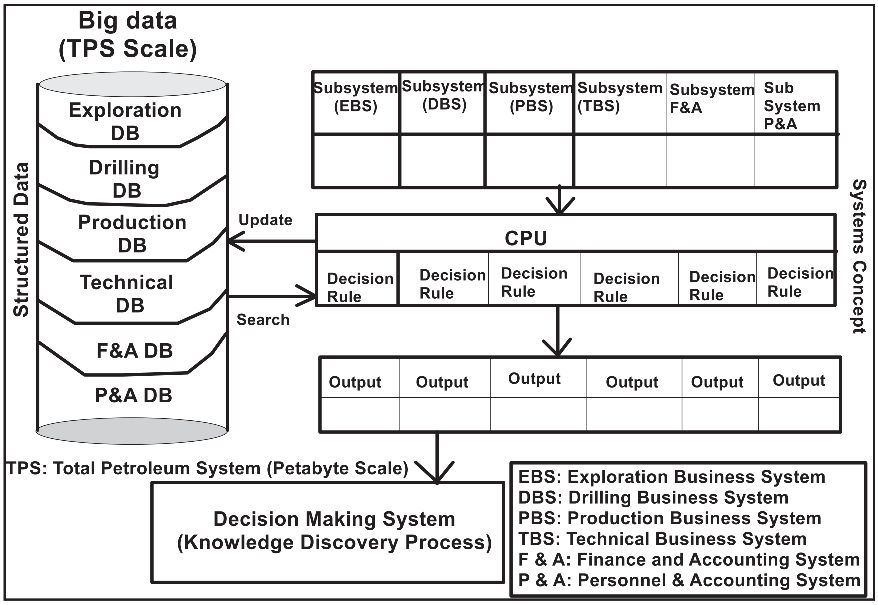

Fig. 3. System connectivity in Big Data perspective.

To address the research questions and objectives (1 and 2 in Section 3), the Big Data paradigm (Nimmagadda, 2015) initially allows recognition of the existence of volumes and varieties of data instances from various operational units. The operational database system (Fig. 3) considerably covers the present-day activities and functions of exploration, drilling and production entities. For connecting business subsystems and data processing tasks, the business rules are designed to meet the desired total petroleum system (TPS) concept. In the data warehouse perspective, this system is at the petabyte scale and is a knowledge-based (petroleum) information system (Vaishnavi & Kuechler, 2004; Nimmagadda & Rudra, 2017) meant for contributing to decision support systems. For real-time business data manipulations, petroleum information systems (Nimmagadda & Dreher, 2012; Schermann et al., 2014) capture the business data, managerial decisions and complex queries from outside vendors. The queries generated from large warehouse repositories are presented to exchange information among corporate-, regional-, project- and unit-level operators. The data views are intended to be analysed using statistical correlations, regression analysis and data mining schemes (Gupta & Gupta, 2009; Provost & Fawcett, 2013; Pujari, 2002) for future forecasts of the resources.

For data warehousing and mining, large volumes and varieties of data are typically modelled in dimensional logical star-schemas (Pujari, 2002) to connect and integrate with other interrelated logical schemas of multiple domains. Initially, a conceptual model is described in the form of an ER logical model, demonstrating the connectivity (Castañeda Gonzalez et al., 2012) among exploration, drilling and production entities and their data attributes. Resource data, when integrated with spatial and periodic data attributes, become large with varying spatial data characteristics. In the integrated project environment, the large-sized data move at a faster rate (Debortoli et al., 2014) in between workstations and among project centres. Irrespective of the size of the volume, the accuracy and veracity of data are monitored in various modules so that the validity of systems and their connectivity are well examined. The connectivity is explored among various entities, as demonstrated in Fig. 3, integrating them into multidimensional models with periodic contexts. In an example, the 2D plot views are examined for patterns and trends of attributes associated with the *cost of exploration* business and *discoveries* made at various periods of time in different geographic contexts (Dodds & Fletcher, 2004). A similar periodic data analysis is intended in the well-drilling and production campaigns and their attribute data dimensions. The user can request a drilled well that has the culture and seismic information for a particular period or range of periods in the plot and high-quality map views, as described in Dunham (2003). In this context, the use and reuse of the data structures are evaluated (Schermann et al., 2014) through utility properties, the systems' approach and knowledge of the connectivity between geological data events that can lead to the interpretation of unexplored resources in geographic dimensions.

4.1. Contribution to business research

The contribution is divided into four important areas:

1. *Knowledge-based methodological articulations:* Building intelligent decision framework articulations envisages the development of multidimensional warehouse repositories. Making data models more flexible and robust can meet the challenges of heterogeneity and multidimensionality, offering Big Data business solutions. The flexibility and robustness hold good for both historical and/or online data sources.
2. *Adding value to Business Research:* Adding new knowledge to the exploration ventures, data interpretation projects in particular, has immense value in validating the methodological framework in various business contexts. The technological advancements have supremacy in managing expensive business operations, thereby cutting costs and turnaround times. It includes delivering quality products and services in multi-client environments. In addition, multidimensional warehouse repositories, associated artefacts and their utility evaluation properties help maintain and deliver sustainable business value in spite of obstinate business rules and constraints.
3. *Increased Scopes and Opportunities:* The Big Data paradigm motivates business researchers with new scopes and opportunities in multiple chains of industry and distributed business environments. Data science of real-time upstream business events encourages researchers to inculcate new data interpretation ideas that support the knowledge discovery process. Spatial data science opens new avenues of research in upstream businesses. Irrespective of data sources and their types in multiple domains, data integration is an added scope and benefit in making new business alignments.
4. *Beneficiaries of Business Research:* Exploration and field development

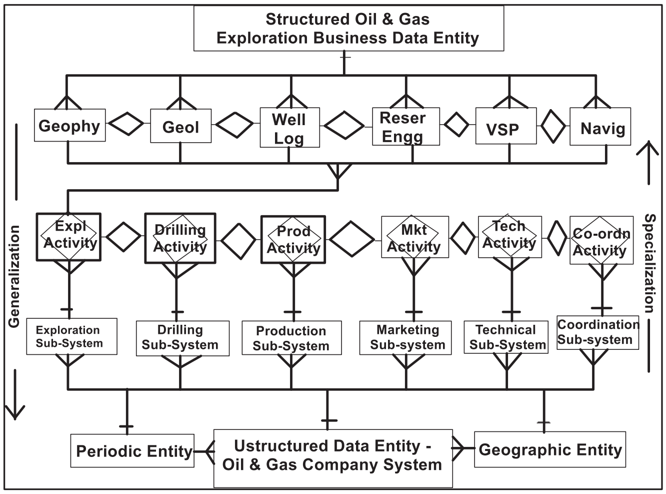

Fig. 4. Conceptual modelling of an Upstream Company.

are part of current business research in industry-contextual domains. Mining and oil & gas explorers, business data managers and data scientists are real beneficiaries of the current business research.

5. Data modelling methodologies

In addition to real-time data instances, the volumes of historical exploration and production are added assets of the business organization, especially in learning and connecting the data models. How the integrated data models impact the current data trends are assessed. It is necessary to document and maintain them periodically, ensuring the knowledge from past data has an impact in the data modelling. Ontologies are built (Meersman, 2004) graphically in the form of knowledge-based entity-relationships (ERs) so that the logical models (Moody & Kortink, 2003) can link different entities associated with the activities and functions of oil and gas businesses. As described in Fig. 4, a conceptual model is designed that combines multiple roles and business activities of exploration, drilling, production and other technical entities of the Oil & Gas Company through ER constructs with several business rules imposed on them. A large Oil and Gas Company is envisaged as a generalized unstructured data entity connected to various data attributes, including periodic and geographic data entities. Characteristically, although there are geophysical, geological, well-log, reservoir engineering, vertical seismic profiling (VSP) and navigational and periodic data entities, they are separable (Castañeda Gonzalez et al., 2012) and composite in nature. The data entities are interpreted in a way to enable use and reuse their attributes in building constructs and models and exploring connections among subsystems with individual subtype data entities. At the specialization level, an exploration business subsystem handles several such forms of structured data from denormalized data relationships.

The models built for mineral and petroleum exploration and production are expected to generate granularity characteristics and graphical visualizations that apprehend their cardinalities explicitly. In multidimensional repositories, logical data structures are aimed at using fine-grain structuring to visualize and link the attribute dimensions and their instances at an atomic level. Data acquisition, data structuring, storage and data mining components (Coronel, Morris, & Rob, 2011; Puschmann & Burgess, 2014), visualization and interpretation are other value-added artefacts in the integrated knowledge management system. Sequences of events (Debortoli et al., 2014; Nimmagadda & Rudra, 2017) are described in all the workflows as per the processing requirements, including importing and exporting the events into warehouse repositories. Second, the designs of relational and multidimensional data structures are compatible, facilitating the data storage environment. Third, the design of several business rules to link and integrate various conceptual and logical data structures are detailed in the following sections.

5.1. Modelling the constructs with business data

Two simple examples are illustrated that demonstrate the periodic aspects differently in the current problem:

An upstream oil and gas company requires document and storing periodic and geographic dimensions (Nimmagadda, 2015) as provided by the contractors. The business rules are as follows:

1. Each exploration holds a production license
2. Each production may also hold an exploration license.

An upstream oil and gas business desires to store petroleum permit information (Longley et al., 2001) and the information on contractors who hold their licenses. The business rules are as follows:

1. The conceptual models for both examples look rather similar, as shown in Fig. 3. The cardinality is the same, but their optionality is different. Let us introduce the time aspect, for example, the situation, which demands a five-year history of exploration, production and contractor permits in the repositories. The analysis reveals that in the first example, the relationship changes into many-to-many, since the production during the course of time varies significantly. An associative entity is required to hold the dates when exploration records need connections with production profiles (Figs. 4 and 5).
2. In the second example – the contractor/permit problem – the relationship does not change over time. One contractor holds at least one permit/license. What could be the reason for a change in the data relationship? In the exploration/production example, the relationship is said to be transferable; that is, production held by exploration can be transferred to another type of exploration over time. In other words, when a prospect is explored by a set of exploration techniques, it can now be exploited by another set of exploration methods when a development plan is proposed over that time period. The relationship changes to many-to-many when the time aspect is considered.
3. In the third example, the relationship is not transferable, that is, permits always belong to one and only one contractor. If the contractor with existing licenses wants to add some other licenses, a new permit order must be placed. It is a somewhat simple criterion to regulate when deciding if one-to-one or one-to-many relationships develop into many-to-many relationships over time. If they are transferable, which is more common, many-to-many relationships are established, and if they are not transferable, they are not.

A typical contractor-survey-production business data entity relationship is interpreted by appraising the connectivity among *contractor*, *survey* and *production* entities, as shown in Fig. 5. The conceptualization attributes are described through one-to-many and many-to-many data relationships, which envisage a close relationship between exploration and production entities. Several such other conceptualization and contextualization attributes are interpreted through entity relationship constructs and models. When referring to the conceptual model, as in Figs. 4 and 5, the following question may arise: how can one distinguish the current and survey histories of the unstructured Big Data situations? The detailed conceptual model, as shown in Fig. 5, does not separate the survey records. The current survey history from these occurrences in the associative relationship (a conceptualized survey-activity) is null where the start and end dates of surveys are stored. Alternatively, two data relationships between contractor and survey entities are kept in the constructs, one for the current and the other for the historical surveys.

In addition, other attributes, such as *the cost of exploration*, the

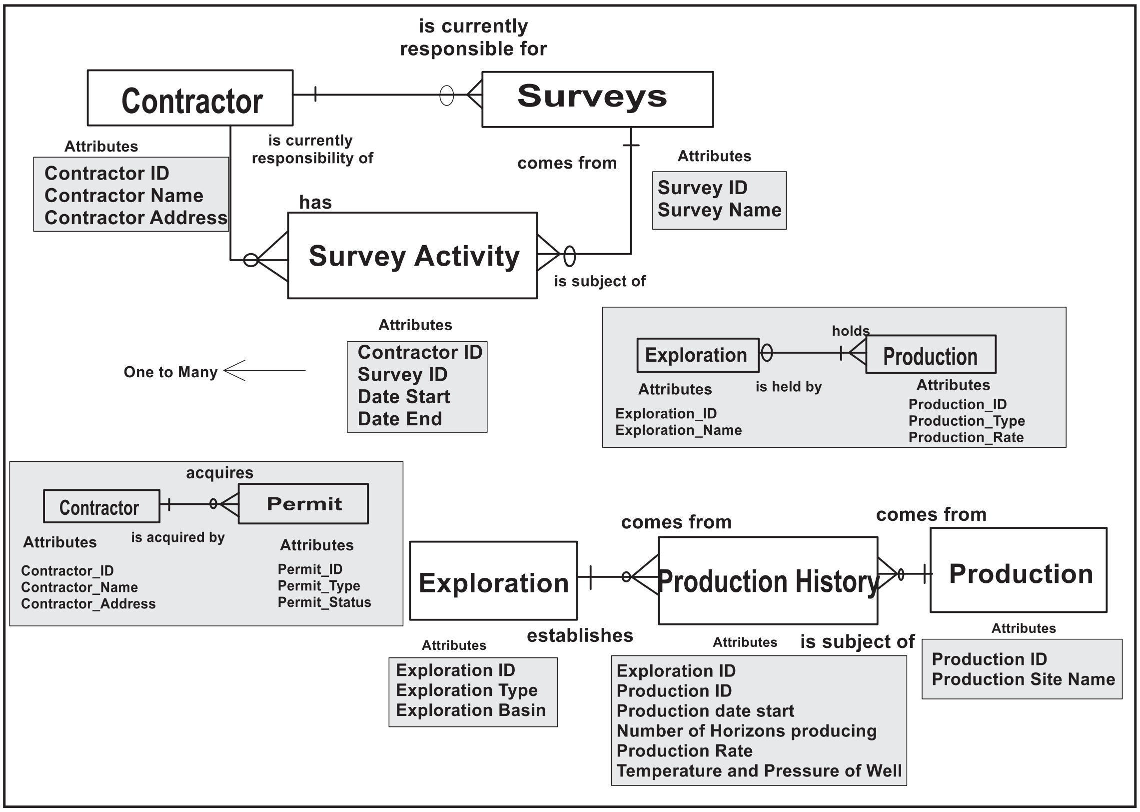

Fig. 5. A conceptual model, a contractor-survey-production problem of a business entity.

drilling cost per metre for an *exploration type*, the *casing type* used in borehole management (drilling business), *borehole-depth*, the *daily hire rate* charged by the contractor and the *time period* taken to explore and *drill-the-well*, are representative attributes in the upstream integrated business research. Occasionally, more than one *casing type* is used in the same borehole management. The same contractor may have several leases or permits in the related oil & gas fields and their linked sedimentary basins (Nimmagadda, 2015). The business rule attributes are incorporated in modelling the constructs along with various other attribute characteristics that are affected by their data relationships.

5.1.1. Using a dependent entity of the business data

The business rules play strategic roles in shaping the data relationships among either real-time or historical data. In this case, one or more business rules may exist:

- One-to-one relationships become one-to-many or many-to-many
- One-to-many relationships become many-to-many

The data to be stored in the repositories include the instances associated with the *exploration costs*, *petroleum production data*, and *mineral discovery* attributes. More precisely, they are *drilled-well* data acquired from drill sites, well-site installations and drilled-well expenditures in different spatial-temporal dimensions. Again, a conceptual model is drawn with an exploration entity or dimension describing attributes (Figs. 3 and 4) such as *exploration ID*, *exploration name* and the *start date of exploration*. The exploration manager decides to acquire and document *the end date of exploration* and the *costs of exploration* attributes and their instances for decades of exploration work performed with petroleum production attribute instances, for which a more detailed production analysis report is requested either by a driller or production data analyst. This model cannot deliver the required data; the present model needs to adapt and meet the new requirements. The changes are shown in Fig. 5. The new entity or dimension described from exploration histories has a dependent entity that keeps track of different *exploration costs* and forms of *production* over periods of time and geography. Note that the attribute *start date of exploration*, as a single occurrence of exploration, has turned into an attribute with multiple occurrences existing in the exploration data histories. The attribute *start date* is the date of a particular exploration work that has come into effect. One needs to make decisions or choices at this stage about recording *exploration costs*, *production rates* and *drilled-well pressures* attributes. There may be two possibilities:

- Periodic change of exploration and production costs or
- Changes in geological entities or dimensions based on geography.

The choice depends on the frequency of change and the importance of recording every change that has occurred in the data relationship. In the case of production changes, it may be appropriate to record every change if an average production rate changes every week. However, in the event of a stock exchange, it may not be feasible to record every detail that affects the share prices of oil and gas, keeping in view the volatility of the market. One may resolve the reporting periods of the changes for that day in mid-session or at the end of the day. For exchange rates, it is usually the daily rates that may be relevant to the value of the dollar that may vary based on the price of oil and gas in the market. It is from the users' perspective that the reporting frequency of the data or their spatial and periodic data relationships in the warehouse structures are recorded and documented.

5.1.2. Changes in data relationship cardinalities

In the data modelling process, one-to-one and one-to-many relationships can turn into many-to-many data relationships, based on conceptualization and contextualization (Meersman, 2004). At times, it is not right to make changes in business situations in which cardinality (Coronel et al., 2011) plays a key role in the modelling process. There are two reasons for cardinality change:

Table 1
Survey attribute dimensions and instances.

Description - surveys	Number	Data size (GB)
Number of survey volumes	35	7
Number of 2D survey lines used in each volume	350 (40,000 Line km)	0.078
3D survey lines for each volume	150 × 150 (350 sq·km)	0.3
Number of sources used in each survey	30	0.008
Number of survey vessels used for each 2D/3D survey	37	0.065
Number of survey types in each volume	23	0.008
Number of survey companies and contractors	75	0.06
Number of survey permits for each 2D/3D volume	975	0.072
Number of exploration types for each volume	3	0.005
Number of basins	15	5
Period	10 years	0.006
Cumulative number of surveys	1282	15

- Changes in business rules
- Business data need to be updated frequently (for data analytics)

5.1.3. Changes in business rules

The changes in business rules often affect the data constructs and models, thus impacting the data integration process. As demonstrated in Tables 1–7, *surveys*, *drilled wells* and *permits* generate a number of varying intricate business data rules. In the *survey-contractor-production* context, the company stores a list of business rules (as derived from Tables 1–7) from surveys and contractor data details that are used for data acquisition and processing in the exploration industry. The current business rules are the following:

- Each contractor acquires one or more surveys
- Each survey is looked after by at least one contractor

The business models for this problem, with sample attributes, are shown in Figs. 4 and 5. Awarding licenses or permits to different contractors is an on-going business process. Two weeks after successful implementation of the tables in the relational databases, another contractor with the approval of the management of the resources company decides to acquire more surveys and adds to the existing databases for building additional exploration knowledge and value for sweet spot analysis. One or more contractors hold surveys or licenses associated with data acquisitions. Such situations are unpredictable and/or unavoidable (Dunham, 2003; Nimmagadda, 2015) in the upstream business, implying that the data relationships become many-to-many. Additional data attributes need to be stored, such as *dates* and *time periods*, when a contractor acquires a particular survey suggesting that many-to-many relationships can resolve an associative entity business. Changing business rules by many contractors or employees associated with employers for that survey is conceptually a simple change to the problem, but a subsequent review of the entire database structure is needed. The business rules are examined periodically and cognitively to assess the probability of changes in the short or medium term. If a change is likely to happen, it needs to be incorporated into the database structure sooner and earlier. Similar rules may be conceptualized and contextualized based on the data attributes described in Tables 1–7.

Table 2
Drilled-well data attribute dimensions.

Description - wells	Number	Data size (GB)
Number of wells in each 2D/3D volume	335	3
Number of periods for each volume	1200 (months)	0.1
Number of rigs for each volume	5	0.005
Number of riggers for each volume	55	0.065
Well status (criteria number)	15	0.003
Number of structures for each volume	10	0.007
Well completion status (criteria number)	17	0.01
Well deviation (criteria number)	8	0.01
Well side tracked (criteria number)	4	0.005
Well classification number: depends on type of survey	3	0.005
Well oil (gas/condensate) producing rate	7 × 5 = 35	0.006
Well formations (number of formations encountered or interpreted, depends on the area and type of exploration)	25	0.028

Table 3
Permit data and attributes.

Description	Number	Size (GB)
Permit dimension	750	7
Permit status	150	0.065

Table 4
Data properties (average estimates).

Data description	Percent (big data)	Basin size (%)
Sparsity	51	47
Heterogeneity	46	45
Multidimensionality	43	44
Granularity	33	26

Table 5
Cumulative business data.

Description	Number	Size (GB)
Period (1940–2015)	75 years	75 years
Volumes	1284	25
Varieties	581	0.075
Anomalies	403	0.2
Geography, thousand sq·km	719.84	0.29
Number of attributes	805	0.05
Number of rows/facts	59,376	17.5
Number of dimensions	977	0.06
Cube metadata size, Kbytes	152,137	10
Production	774 MMBOE	6

Yes: relationship or connectivity in a metadata structure.

Table 6
Major dimensions and their connectivity.

Dimensions/facts	Survey facts	Well facts	Permit facts
Period	No	No	No
Contractor ID	Yes	Yes	Yes
Permit ID	Yes	Yes	Yes
Exploration ID	Yes	Yes	Yes
Basin ID	Yes	Yes	Yes
Exploration type	Yes	Yes	Yes
Navigation ID	Yes	Yes	Yes
Field ID	Yes	Yes	Yes
Prospect ID	Yes	Yes	Yes
Geophysical ID	Yes	Yes	Yes
Geology ID	Yes	Yes	Yes

Yes: relationship or connectivity in a metadata structure.

5.2. Dimensional modelling of periodic data attributes

For the purpose of constructing and connecting databases, ontology modelling is still a popular approach through ER representations of entities and their relationships. However, for the purpose of

Table 7
Interpreted petroleum systems and evaluated business data attributes.

T	W	B	PRS	RES	SRC	STR	SLS	MIG	Tim	CLS	ACCM
Conv.	50	5	5000	5	2	7	2	2	3	3	500 MMBOE
Unconv.	25	3	100	2	1	3	1	1	2	2	50 MMBOE

Conv: Conventional; Unconv: Unconventional; T: Type of Field; W: Number of Wells, B: Number of Basins; PRS: Average Flow Pressure of the Well; RES: No of Reservoirs; SRC: Number of Sources; STR: No of Structures; SLS: Number of Seals; MIG: Number of Migration Pathways; TIM: Time of Deposition; CLS: Number of Classes; ACCM: Average Accumulations.

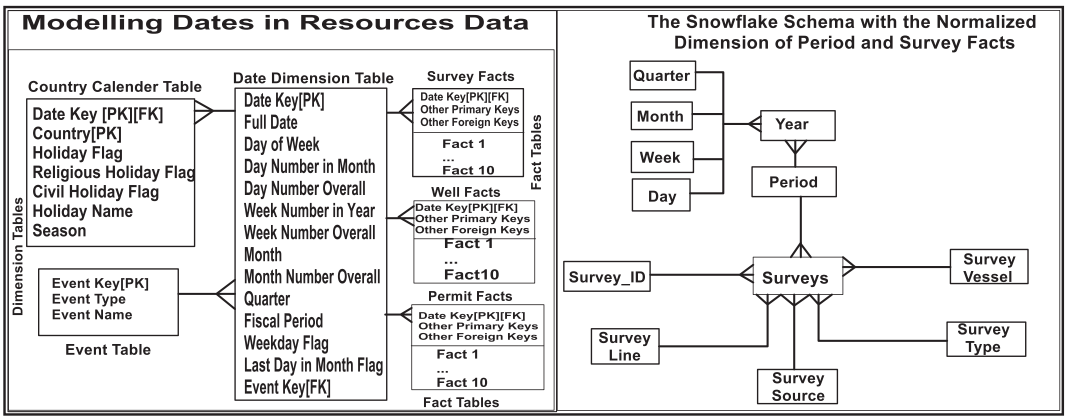

Fig. 6. Multidimensional star schemas depicting periodic dimensions.

compatibility, scalability and flexibility, dimensional schemas are designed for generating a warehouse repository that can accommodate the unstructured, heterogeneous and multidimensional data. Efficient use of storage and analytical processing are added advantages of the dimensional modelling. In the present study, a group of entities and their corresponding attributes are considered to draw dimension models (Pujari, 2002). The conceptual ER models are converted into various dimensional schemas (Moody & Kortink, 2003). Various objects created by the user, such as tables, queries, forms, reports, views, business rules and constraints, can also be used in object-oriented modelling (Nimmagadda, 2015). The star schemas presented in Fig. 6 describe the structure of the periodic attribute dimensions of the upstream data. The most common database models that define the data relations have one-to-many or many-to-many relationships. The cross-referenced key attributes that link the tables represent the relationships between entities and/or dimensions. Primary and foreign key attributes link multiple databases, providing easy access to large volumes of data. A sample of multidimensional schemas involved with the date dimensions is described for an upstream business, as shown in Fig. 6. The multidimensional star schemas are illustrated with exploration data attributes involving surveys, drilled-wells and permits, including periodic dimension attributes. For example, several dimensions and facts are documented for each survey, connecting it to a generalized periodic dimension.

As illustrated in Fig. 6, database contents vary with time period (Gregersen & Jensen, 2002). For example, the databases contain the product information in which the unit price is changed for each product when material and exploration costs change as per market situations. If only a current price is required to change, then the price is modelled as a single-valued attribute instance. However, for accounting and billing purposes, there is a need to preserve the history of all expenditure data and the period over which each expenditure has had an effect on the contents of the database. The result is a multivalued composite attribute named ExplorationCost_History. The components of ExplorationCost_History are ExplorationCost and Effective_Date (Fig. 6). An important characteristic of such a composite attribute is that it has multivalued attributes, and the composite attributes go together, linking multiple models. Each value of the attribute ExplorationCost is time stamped with its effective date.

5.3. Integrated framework articulation

The adaptability of the integrated methodological framework is highlighted in chains of business contexts and scales while dealing with the heterogeneity and multidimensionality of volumes and the variety of data sources in resource businesses. The existing pitfalls associated with business alignments and chains of operations force the businesses to undertake research in building integrated knowledge-based articulations. The logical and physical schemas made up of entity-relationship and multidimensional data relationships facilitate the adaptability of artefacts in the data integration process, easing the complexity of Big Data systems and strengthening their implementation in business ecosystems.

An integrated framework is articulated with intelligent workflows involving Big Data dimensions and their fact data instances. These facts are characteristically seismic, well-log, reservoir and production business data dimensions. The data are populated in the tables of multidimensional warehouse repositories. Slicing and dicing are performed to obtain the data views from warehouse repositories for interpretation. As an example, the areal extents of the quality of *porosity* and *permeability* attributes pinpoint the new prospect locations in the investigated area.

The methodology is innovative for upstream business researchers, especially when businesses vary in space and time dimensions. The Big Data, including the online business data analytics, are significant features of the research, simplifying the business operations where geographically distributed operational units need connections through physical and logical data organizations.

To address the research question and objective 3 in Section 3, the integrated framework is further evolved to accommodate the spatial and periodic domain ontologies and integrate them in multiple business applications. The resource companies that do business at several locations in Australia and in several countries perform transactions on different dates. A country calendar table is added to document the characteristics of each date for each country. Thus, the date is a foreign key in the country calendar table, and each row of the country calendar

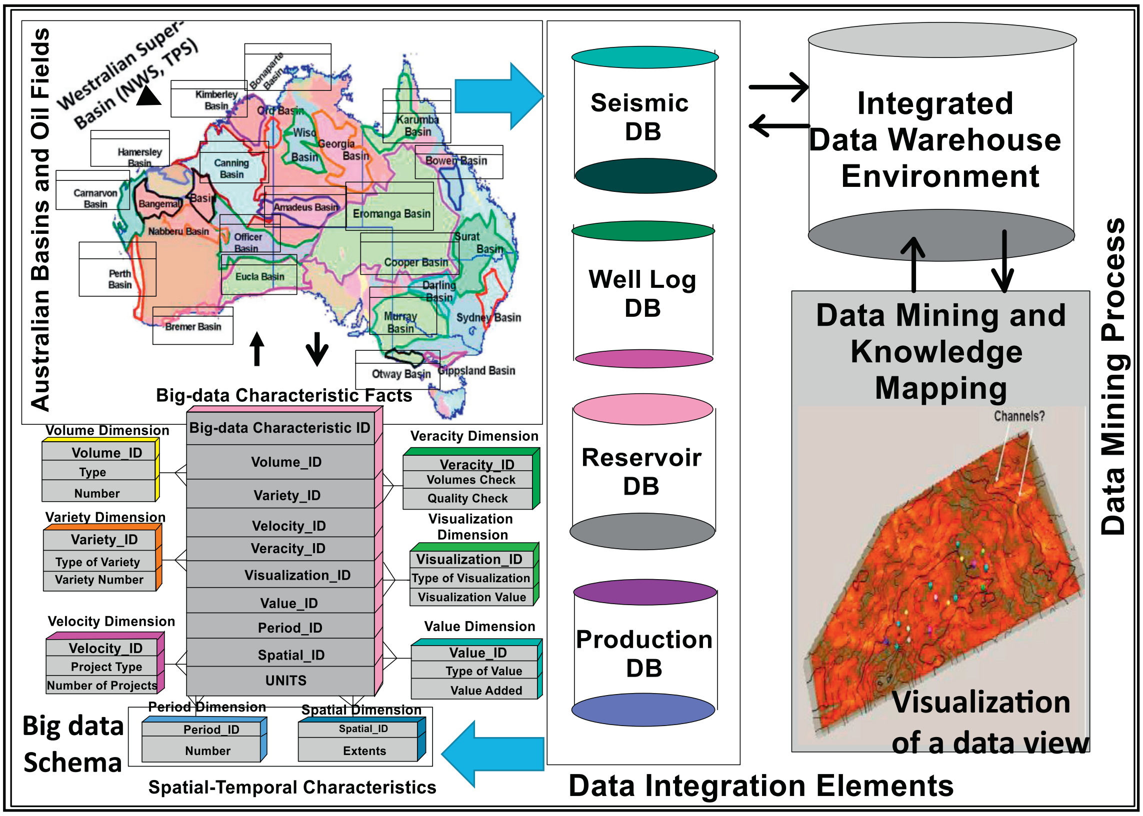

Fig. 7. Methodological framework and workflow.

table is made unique by a combination of the date key and the country, which form the composite primary key in that table. A particular event may occur on a given date (for simplicity, no more than one special event may occur on a given date). Event data are normalized by creating an event table, so each descriptive datum for each event (e.g., Christmas, Good Friday) is stored only once. The time dimension is made relevant to both exploration and production data, including sweet spot discovery data, implying that the period dimension is linked to different operations and activities of the resources industry.

As shown in a workflow in Fig. 7, the data sources in a Westralian Super Basin (Longley et al., 2001; Li, 2010) have volumes and a variety of attribute dimensions and fact instances, enabling business researchers to place all the information in a single multidimensional repository, thus allowing data mining and visualization tasks performed at multiple locations and in real time.

6. Results and discussion

There is an enormous demand for structured and quality information in large operating companies, for which the scope of adapting new technologies and solving problems associated with organizing complex business data events are explored. Pitfalls have been observed (Nimmagadda, 2015) in aligning businesses and the chains of operational units that have posed the data integration challenges. Many researchers outline the innovative integrated methodologies in industrial application scenarios (Vaishnavi & Kuechler, 2004, 2007). Several data types exist in multiple domains, with a large amount of online and offline data in various geographic regions. This fact has motivated the current research and the design of the research questions. Several interviews and consultations have been conducted with a variety of oil and gas production and service companies to clarify the research questions and objectives. The integrated methodological framework thus articulated with various data modelling, data warehouse structuring, data mining, data visualization and interpretation artefacts, serve the purpose of handling the heterogeneity and multidimensionality of business data in different application and knowledge domains.

6.1. Data mining, visualization and knowledge interpretation

In this section, the research questions and objectives 4 and 5 of Section 3 are focused on. Like any other data, upstream business data instances change with time and space. The simple attributes derived from the composite spatial-temporal dimensions are included in the multidimensional modelling.

The field exploration, prospecting, appraisal and development stages produce an enormous amount of information and knowledge at multiple levels of system investigation and analysis, each level adding information for integration and value to business data analysis. Different data views are extracted from metadata for domain-knowledge and its interpretation in decision support systems. The knowledge obtained in all case studies is ensured with meaningful interpretation, implementable in different application scenarios. For example, an *element* within a petroleum system is found to be more productive, and its areal extents discovered are large enough that similarly strong attributes are predicted in other fields of associated systems, including in unconventional petroleum systems. The explanatory, confirmatory and exploratory data mining procedures (Nimmagadda, 2015; Provost & Fawcett, 2013) play a decisive role in the increased availability of knowledge from a system and the effectiveness of its interaction with associated systems. The knowledge acquired in a system has an impact on perceiving the knowledge of other related (or associated) domains, for which data models are described for effective data mining, visualization and interpretation (EMC Education Services, 2015; Nimmagadda, 2015). A generalized knowledge-based process model as given in Fig. 8 that depicts a workflow for modelling the data sources, including business data interpretation and its implementation in the upstream businesses.

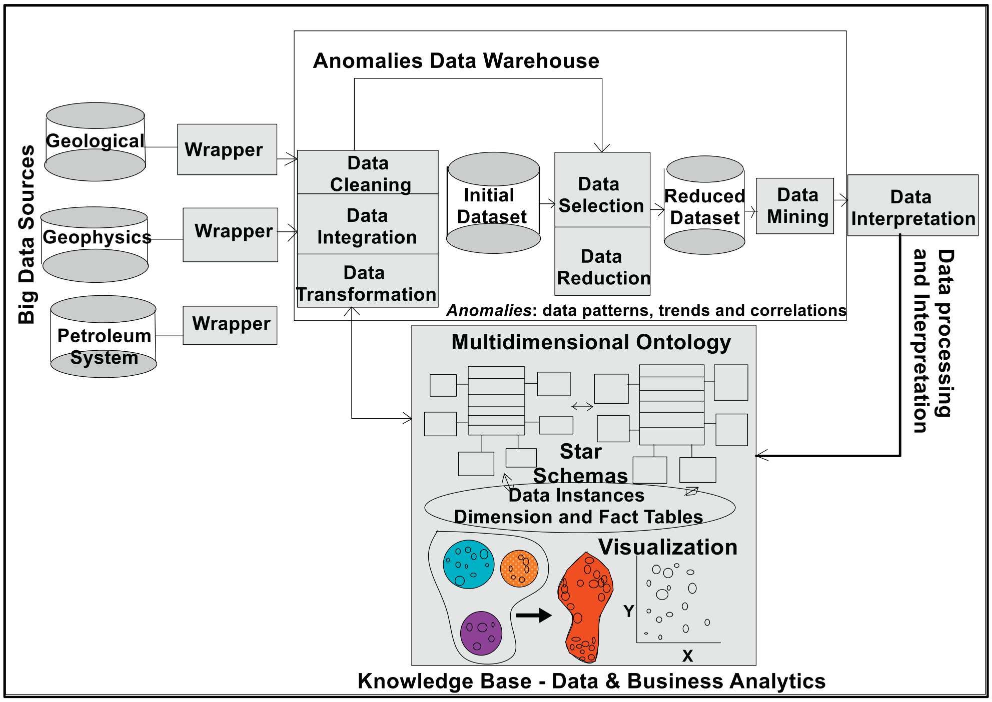

Fig. 8. Knowledge management and discovery workflow.

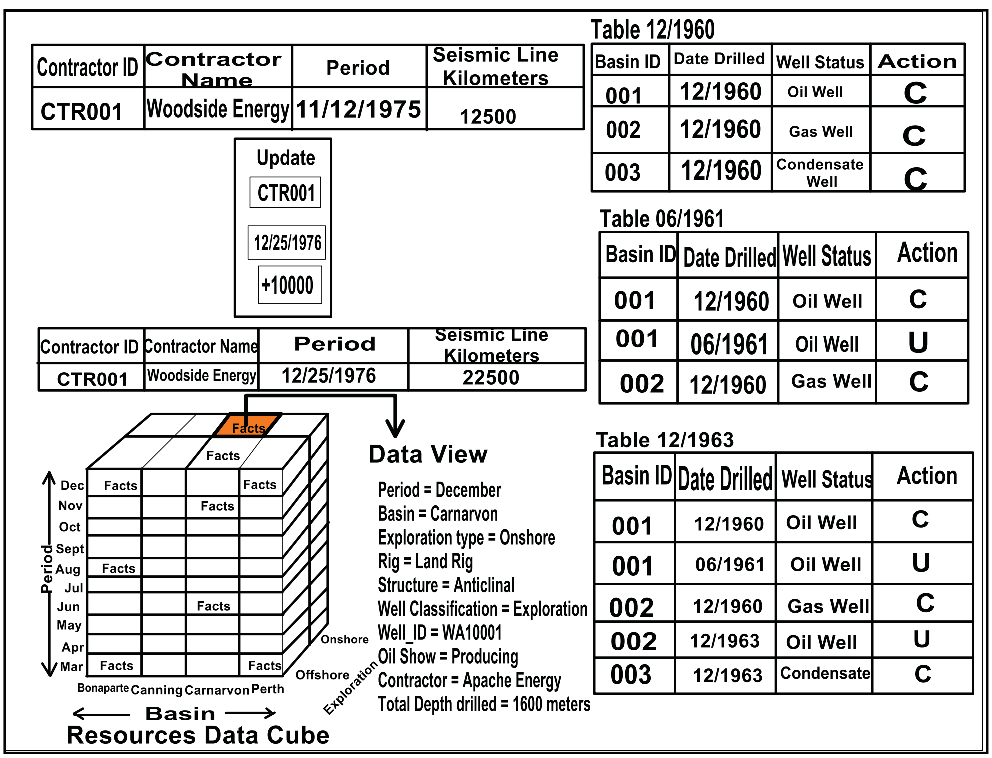

Fig. 9. Cuboid structure and its data views.

For knowledge mapping, fine-grained data schemas are well-fixed criteria for which the data relationships are denormalized to obtain finer data views as per user-end queries. There are two ways of observing fine-grained data models. One is a multidimensional cube (Moody & Kortink, 2003; Pujari, 2002; Fig. 9) in which each cell contains one or more attributes. In other words, in dimensional modelling, the attributes are categorized in such a way as to reach their atomic levels. Based on "Commit" and "Update" actions in the cuboid metadata, interpretable data views are extracted for visualization and knowledge mapping. The users (typically geoscientists, engineers and reservoir modellers) of the data warehouse thus summarize the data as aggregated data views. As shown in Fig. 9, the data views are presented for different *time periods*, *surveys conducted*, *contractors*, and *wells-drilled* attributes for interpretation. In a dimensional cuboid structure, each cell holds data relevant to the intersection of all its dimensional values. For example, a cell might contain many drilled wells in a particular period and a basin, with a specific number of oil- and gas-producing horizons. Similar cuboid data may be viewed from a repository that may represent other star-schemas. Such periodic data are typically aggregated (Fig. 9) with appropriate time intervals, yielding a large volume of equally spaced time-series data. Such data views are analysed using classical data mining schemes, such as statistical tools, as given in Gupta and Gupta (2009) and Witten and Eibe (2005).

In the multidimensional data warehouse repository, the schedules on each day of the drilling or mining of a borehole or a mine, if completed ahead, are documented. The operations of each truck or rig used in the drilling or mining operations are pursued. Each lease is given a unique permit number. To achieve more-efficient operations (Dodds & Fletcher, 2004), the information on the movement of drilling rigs in-between the adjoining concessions is incorporated with updated exploratory drilling campaigns. User involvement is necessary to achieve satisfying results and present them to the decision support systems. From corporate metadata, access to the databases is requested to create data views by various working groups on exploration, drilling, production and marketing entities at different project sites. The actions taken on the database operations are tabulated as shown in Fig. 9. This procedure enables us to explore the data views for interpretation of the innovative knowledge from newly added conceptualized and contextualized attribute dimensions in the warehouse.

Several aggregates presented in Fig. 10 facilitate the users' perception on visualization and interpretation artefacts. Several aggregate data views represented in Fig. 10(a) to (f) signify the seismic and contractor attributes with periodic dimensions. The visualization of metadata cubes and their data views that represent reservoir and production trends facilitate the interpretation, enabling the use and reuse of knowledge-based exploration and production in the evaluation of upstream business research.

Several survey-drilled-wells-permits data are used in the data modelling process. If all survey-drilled-wells-permits were combined for a total of ten volumes in the framework (Fig. 7), the total Cube Metadata Size would be $10 \times 1100 \times 350 \times 300 \times 360$ bytes. After adjustment with complete transactions, if considered in a particular periodic dimension, the total data warehouse size could be on the order of 415 GB. The descriptions of the volumes and varieties of data used in the modelling are given in Tables 1–7. Issues such as data sparsity, heterogeneity, multidimensionality and granularity and their sizes that affect the modelling, as well as the overall integrated framework, are evaluated. The common attribute dimensions for connecting the fact tables (for each volume) involved in the E & P evaluation are *number of surveys*, *number of drilled wells*, *number of permits*, *exploration type*, *the number of basins* and *the number of companies including production*.

6.2. Design of decision tree mining structures

Identifying multidimensional classifications in large datasets is a significant problem (Nimmagadda, 2015) from the data mining perspective. Databases have a number of records and sets of classes, each record belonging to a given class. The problem of classification is to decide the class with which a given record fits. The classification problem is concerned with generating a description or a model for each class from the given dataset. For example, similar production data instances in specific reservoir regimes can depict a particular classification. These classified regimes are supervised by training the datasets. Using the training sets, the classifications create the descriptions of the classes. The descriptions help classify the unknown records and sets of the databases. In addition to training, the datasets are tested, and these tests can be used to determine the effectiveness of a classification. Decision tree mining is a classification scheme that generates a tree and sets of rules representing the model for different classes in a given dataset. To develop the classification, the sets of records available are divided into two disjoint subsets – a *training set* and a *test set* (Fig. 11a). The former is used to describe the classifier, and the latter is used to measure the accuracy of the classifier. The accuracy of the classifier determines the percentage test examples that are correctively classified. As described in Fig. 11a, decision tree structures generate understandable rules, especially among petroleum system elements. These tree structures handle numerical and categorical attributes and provide a clear indication of which element or process of a system is important for predicting/forecasting the classification.

In the case of unconventional oil & gas fields, the decision tree mining analyses which horizon (a producing layer) has dense fractures, indicating the strength of the reservoir. The answer can assist in planning future boreholes and their placement in the investigated area.

In the current research, three attributes are found to be significant - porosity, permeability and kerogen content/total organic content (TOC). Attributes whose domain comes from the non-numerical

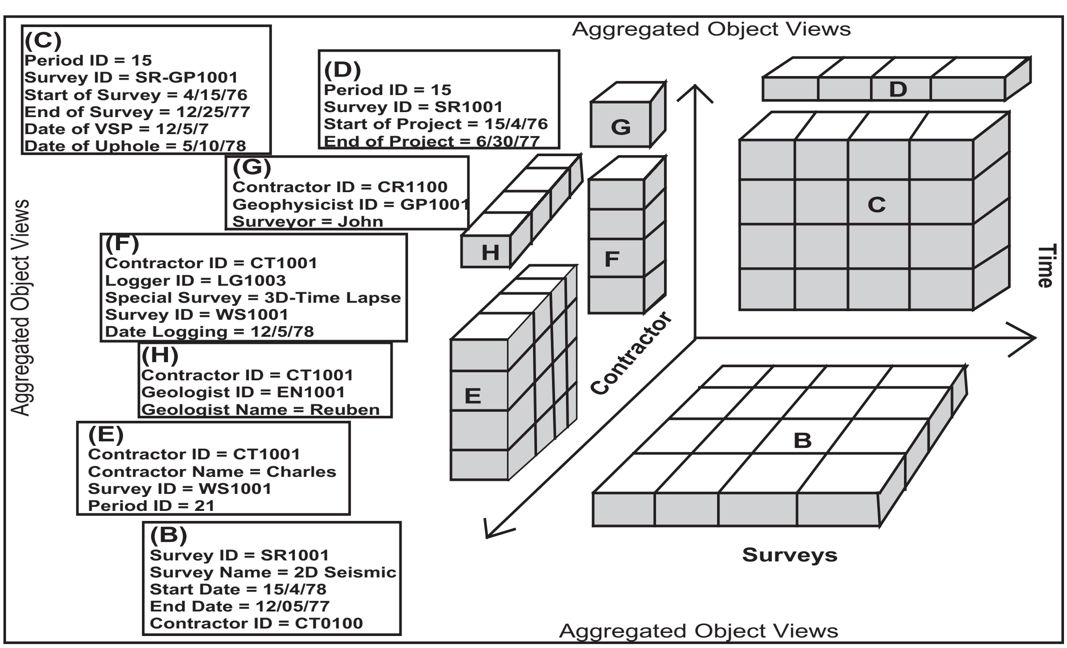

Fig. 10. Data views and aggregates extracted from metadata cubes.

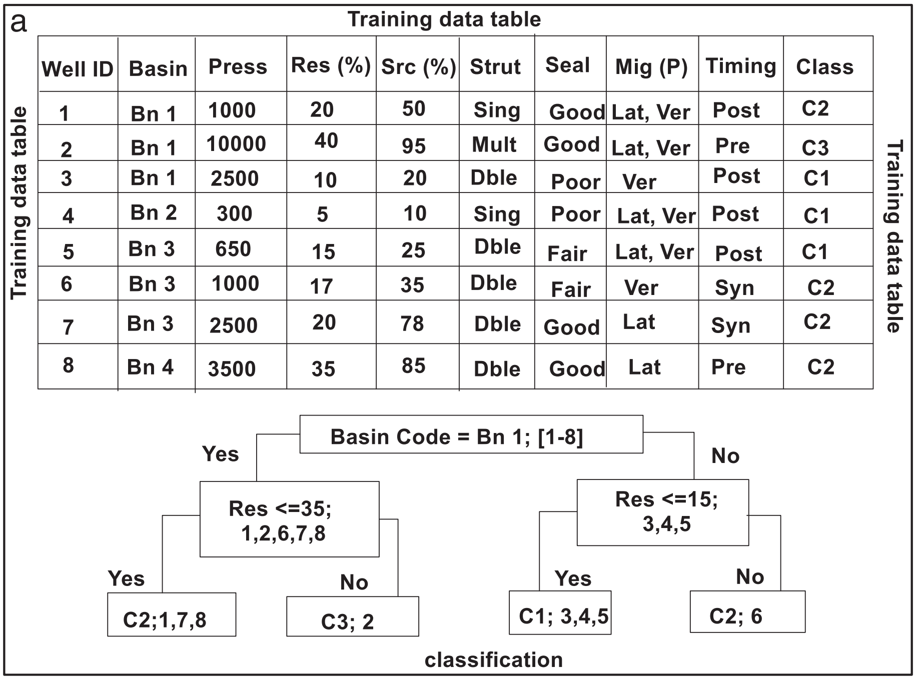

Fig. 11a. A decision tree structure generating mining rules.

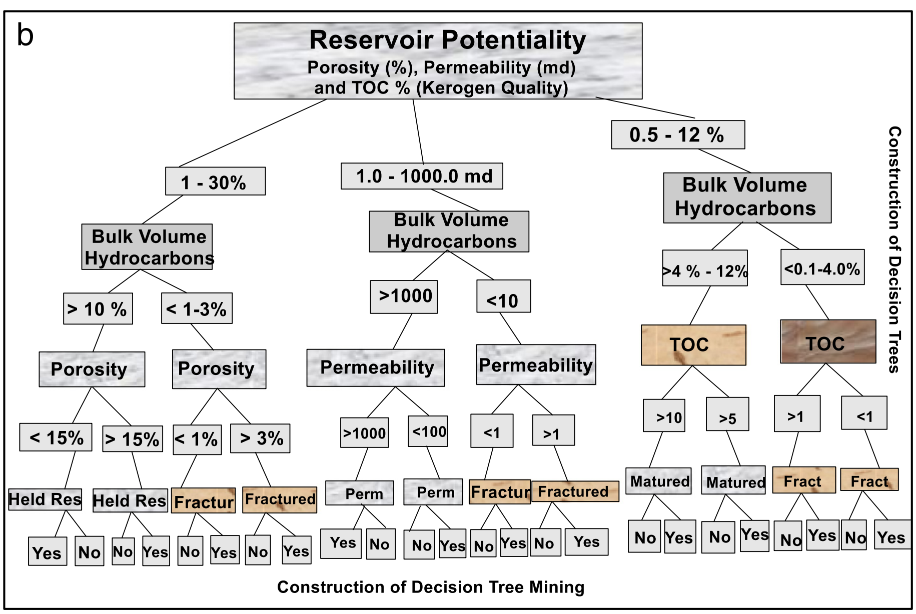

Fig. 11b. Multidimensional decision tree structure evaluating the petroleum system.

attributes are called categorical. Fig. 11b shows the construction of a decision tree mining model in which various mining rules associated with porosity and kerogen instances and their cut-offs are described. Favourable data instances of porosity, permeability and kerogen content that contribute to the description of various fractures and their categorizations are interpreted. Interpreters make use of such information models as decision-making tools for ascertaining which type of rock and kerogen content contributes to commercial hydrocarbon accumulations with favourable reservoirs. Among the major strengths of decision tree mining models are that they can generate more logical and interpretable data mining rules and can provide flawless clues to which fields have significance in predictive classification and for future descriptive exploration and production ventures.

To build queries and data views from warehouse repositories, specified *drilled wells* and or *seismic profiles* are chosen that are of interest in the current attribute interpretation, which adds value to the overall upstream business research project.

6.3. *Analysis of evaluation properties*

The current research is significant for explorers involved in upstream businesses and for data management professionals and information system engineering specialists. The research problem solutions are implementable in industries where data heterogeneity and multidimensionality issues exist. For the kinds of risks and uncertainties observed in the E & P, the business data models are significant in the exploration, appraisal and field development stages. The integrated methodological framework can risk minimizing the economics involved in expensive exploratory drilling campaigns. Upstream business operations in multiple hierarchies (Nimmagadda & Rudra, 2005; Rudra & Nimmagadda, 2005) can easily be performed and achieved the business goals. Various functions and activities associated with operational units globally can be aligned smoothly both online and offline. The evaluable properties of the artefacts used in the framework are relevant to the direct benefits and improvements needed for future data structuring in various domain applications. In this section, the value of the holistic and integrated methods is corroborated, addressing the research question and objective 6 of Section 3. The constructs and models used in the current domain application are evaluated for which utility properties (Venable et al., 2016; Venable & Baskerville, 2015) such as *yearly percentage use*, *reuse*, *effectiveness* and *interoperability* are considered. How workable the artefacts and attributes are in the current domain research and whether they can holistically add value to the Big Data project, including to its *productivity* entity, are examined. Three significant peaks, I, II and III, are interpreted in the bubble plot views. Overall, an increase in percentage use, reuse, effectiveness and interoperability property instances is observed with swelling in bubble size, implying their cumulative attribute strengths with the period dimension. These peaks corroborate with the peak values observed for the volumes and varieties of the surveys, the drilled wells and the permit attribute dimensions; they are illustrated in Fig. 1a.

As shown in the bubble plot in Fig. 12a, these significant peaks are interpreted from the estimated percentage values of the *use*, *reuse*, *effectiveness* and *interoperability* properties. These attributes corroborate with the percentage occurrence of *sustainability*, *productivity* and *value addedness* attributes, as demonstrated in Fig. 12b with attributes such as the *number of surveys*, *drilled wells* and *permits*, as illustrated in Fig. 1a. The utility properties of the integrated framework are further evaluable with the potentiality of the petroleum systems and their modelling. The system elements and processes interpreted using the Big Data volumes and varieties of E & P data sources are translated into conventional and unconventional commercial petroleum resource ventures, as given in Table 7.

6.4. Online business data analytics and business value

Delivering quality data and timely information to users in the distributed business environment is important for online business analytics and achieving business research value. The recent innovations in information and communication technologies (Nimmagadda et al., 2015) motivate the performance of business processes electronically from far-off places. The data warehouse is a backend application of e-business activities. Developments in Internet, extranet, and Intranet services provide better access to and interfaces with data and information portals. XML technologies bridge the gap of exchanging documents geographically. The data views extracted from oil and gas data warehouse repositories are exchanged between different operational business units through XML documents or even in simple ASCII and excel files. The cross-sectional data from different oil company, data processing and data analytics scenarios provide value for performance indicators, predicting the economic and technological trends in spatial-temporal dimensions. XML technology is poised to make integrated data connectivity a reality in the oil and gas business. A major step towards reaching XML application in the petroleum industry comes from the cooperation of several major company's software vendors and managers who contribute in building e-business applications for commercial offerings.

Digital clouds are added tools in which all the digital data from multiple domains (or ecosystems) are captured in a single repository, with access to the data being offered in different geographic regions through Internet media. "The digital cloud" is the storage and delivery of digital data from on-demand computing resources—including everything from applications to data centres—over the Internet (Nimmagadda et al., 2015). In the present study, an attempt is made to describe digital data through e-clouds from multiple domains (in geographic centres), as shown in Fig. 13. Big Data are featured in digital media, and the Big Data characteristics that make the integrated framework work in multidimensional distributed business environments support the concept of digital business ecosystems. The Hadoop distributed file system (HDFS) stores volumes of data and provides access to various applications online. The constructs and logical schemas use the digital cloud in multiple domains. Map Reduce and Spark utilities are used for quick data analytics. The purpose of data mining, visualization and interpretation is to provide new knowledge using the online data analytic tools.

Improving the accuracy of information and reducing the search time on cloud computing servers are key rationales of business data analytics. Supportive technologies comprise agents (Erdmann & Rudi, 2001) for searching the data, data delivery agents using meta-data languages (e.g., UML, XML and HTML), and knowledge representation tools, including agent-based software engineering (Jennings, 2000). The standardized documents characterize the simplicity and cost-effectiveness

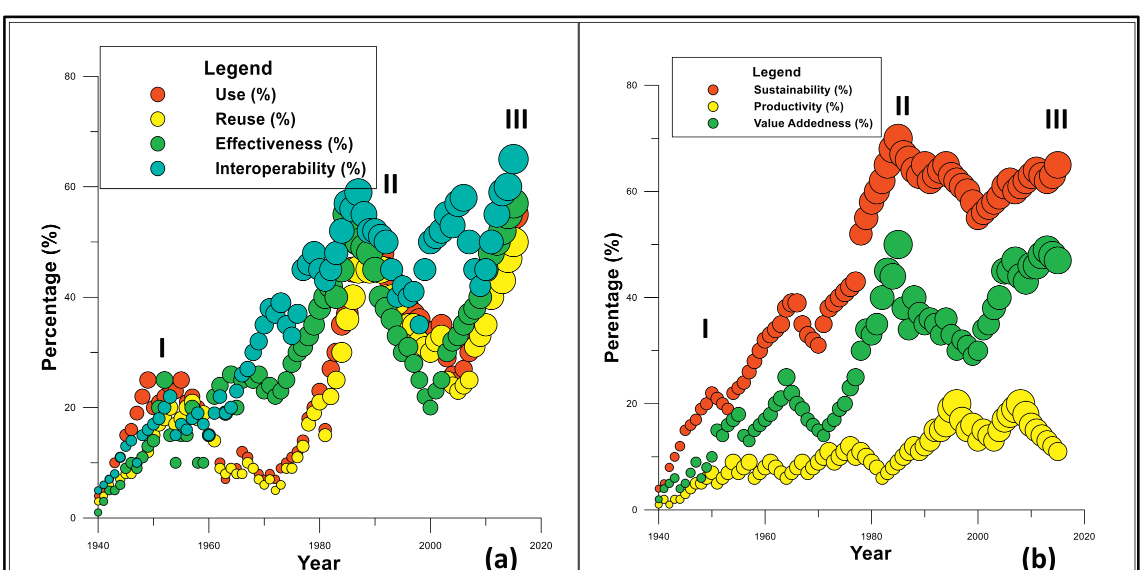

Fig. 12. Bubble plots, evaluating the utility properties of the integrated framework.

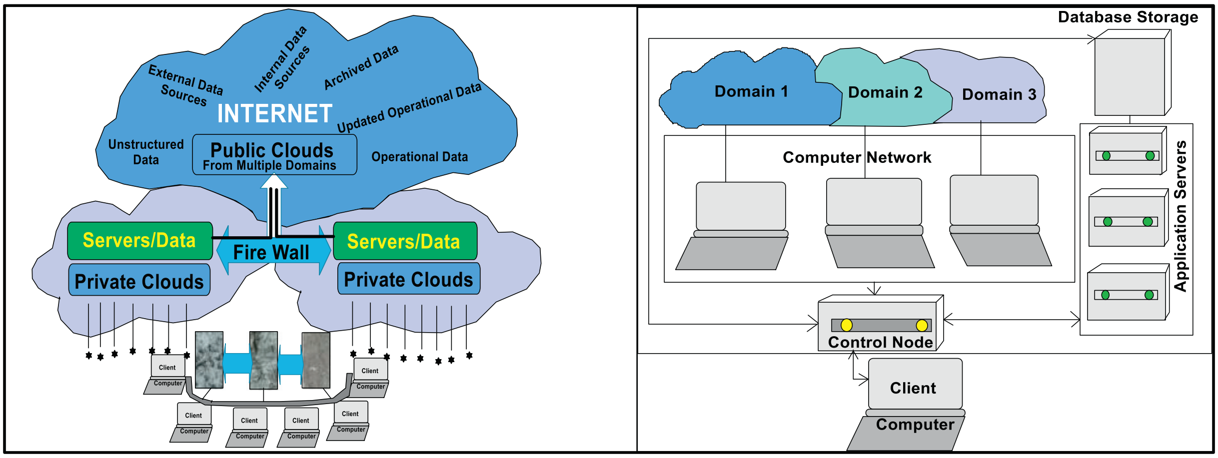

Fig. 13. Digital clouds from multiple domains and online business data delivery.

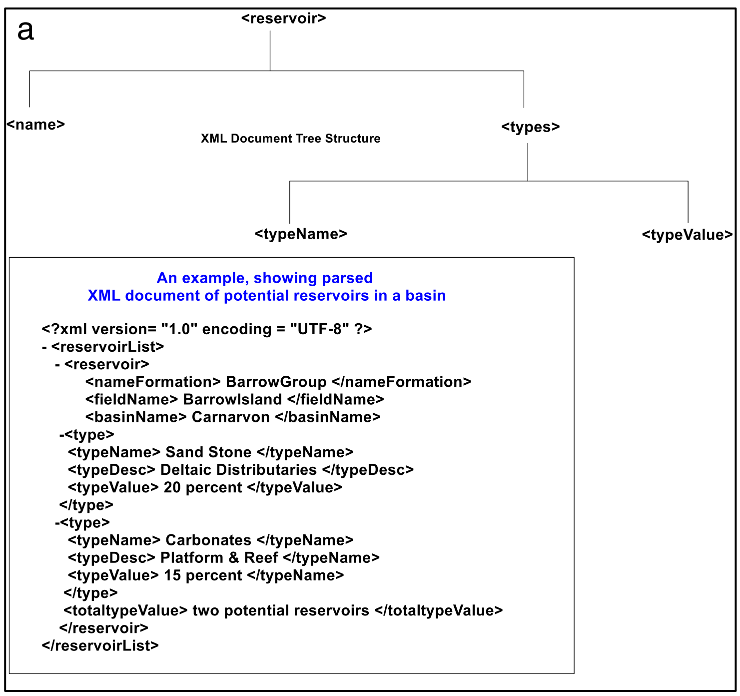

Fig. 14a. An XML view description carrying *reservoir* information.

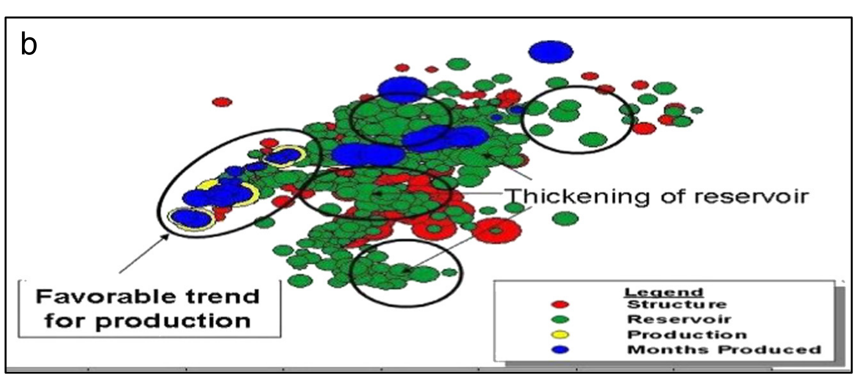

Fig. 14b. A knowledge-based graphical view of producing reservoirs drawn from warehoused business metadata.

in transferring data between disparate workflow systems. Oil and gas companies facilitate the bringing together of data and workflows, easing business transactions and offering quality customer services. All the document exchange procedures are compatible with the data warehousing technologies. Business analysts trigger the warehouse to access pieces of information on the petroleum data of a drilled well, and the warehouse identifies the description of the data view. The search engine acts to locate the data from that particular data warehouse repository for real-time transactions.

The data views interpreted from metadata carry useful information on prospect (sweet spots) identification and prospect analysis. Erdmann and Rudi (2001) provide data view structures with an ontological focus. To transmit data views geographically, the knowledge-based agents are structured in XML codes. Hori and Ohashi (2005) and Erdmann and Rudi (2001) demonstrate the use of XML technologies for preparing and delivering the research outcomes online. Real-time and interactive visualization and interpretation of multidimensional data views in different geographic regions are advantageous in spite of transmission challenges. For online deliveries, XML schemas (Heather, 2004) are designed that support the rule-based agents in the reservoir domain in carrying and delivering the data and information to multiple operational centres. A typical XML document that carries the *reservoir* attribute dimension and its instance is shown in Fig. 14a.

An XML schema for conveying information on the type of reservoir and its qualities in a multi-client environment is described in Fig. 14b with new domain knowledge and is ready for sharing information online.

7. Conclusions and recommendations

Oil and gas business data sources are made of composite dimensions, with *period* and *space* dimensions that include details on the *day*, *month*, *quarter* or *year* and their instances. The *periodic* data structure is complex and requires more joins to bring the query results into one table. Similar to the periodic dimension, the *space* dimension has composite attributes. The use of data attributes in the entity and dimensional modelling and their integration in a warehouse environment are demonstrated. The performance and ease of access to warehouse repositories, as desired by the user are assessed. The business data analysis provides interesting trends, correlations and patterns within the multi-dimensional metadata cubes. These trends are useful for knowledge extraction and interpreting useful information, particularly for predicting future forecasts of resource data. It is advantageous to have the periodic dimension in the warehouse structuring of the resource data. Data modelling focused on Big Data facilitates the use of characteristic instantaneous data views for interpretation. Both periodic and geographic dimensions affect the cardinality of the data relationships. The business rules described in resource data structures are flexible, ensuring the ease of use and reuse of the data structures. Big Data business research demonstrates its implementation in industry, as the volumes, variety, heterogeneity, and multidimensionality address the granularities of the data structures. Spatial and periodic dimensions have a significant impact on the visualization and thus interpretation of new knowledge that adds value to upstream business research projects.

The use and reuse of the constructs and models in various application domains and systems are significant utility properties of artefact evaluations. They are demand-driven integrated framework articulations that not only impact business life cycle optimization but also draw attention to resolving challenges associated with multiple domains and their attribute dimensions. Although business research is more focused on the resources industry, the novelty lies in building the methodology and calibrating the use, reuse and interoperability criteria of the models associated with chains of resource industries. An integrated framework that is built for managing complex data systems drives the resource businesses effectively, even extending to other domains of research.

The overall outcome of the current research validates integrated framework articulations, including the interpretation of domain knowledge from Big Data metadata structures. From conceptual modelling to its implementation, the spatial-temporal attribute dimensions play a significant role in revealing new business knowledge through the interpretation of several data views extracted from warehoused metadata and their associated repositories. The Big Data-guided integrated framework can make business research sustainable, productive and valuable to multiple chains of an industry. The integrated methodology has scope for further research and its implementation in various other applications, such as healthcare ecosystems, modelling of the total environment and disaster management domains.

Acknowledgements

The authors are grateful to Professor Heinz V Dreher and Dr. Amit Rudra, School of Information Systems, for useful discussions while preparing the manuscript and critical comments. The authors are thankful to the international reviewers for detailed periodic reviews that helped us improve the quality of the manuscript. Our sincere gratitude to the School of Management and Information Systems Discipline for necessary facilities provided during preparation of the manuscript and permitted us to publish it in the journal.

References

Agarwal, R., & Dhar, V. (2014). Editorial—Big data, data science, and analytics: The opportunity and challenge for is research. *Information Systems Research, 25*(3), 443–448.

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

Cleary, L., Freed, B., & Elke, P. (2012). *Big data analytics guide: 2012.* CA 94607, USA: SAP.

Coronel, C., Morris, S., & Rob, P. (2011). Database systems, design, implementation and management. *Course technology*. USA: Cengage Learning.

Debortoli, S., Muller, O., & Brocke, J. V. (2014). *Comparing business intelligence and big data skills, bise – Research paper.* Springer Fachmedien Wiesbadenhttp://dx.doi.org/10.1007/s12599-014-0344-2.

Dhar, V., Jarke, M., & Laartz, J. (2014). *Big data, Wirtschaftsinformatik.* Springer Fachmedien Wiesbadenhttp://dx.doi.org/10.1007/s11576-014-0428-0.

Dodds, K., & Fletcher, A. (2004). Interval probability process mapping as a tool for drilling decisions analysis – The R&D perspective. *The Leading Edge, 23*(6), 558–564.

Dunham, H. M. (2003). *Data mining, introductory and advanced topics.* Prentice Hall Publications (10–200 pp.).

EMC Education Services (2015, 27 Jan.). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. *Business & economics*John Wiley & Sons (432 pp.).

Erdmann, M., & Rudi, S. (2001). How to structure and access XML documents with ontology. *Data & Knowledge Engineering, 36*(3), 317–335 (Elsevier Science Publishers, B.V, Amsterdam, The Netherlands).

Gornik, D. (2000). Data modelling for data warehouses. *Rational software white paper*www.rational.com/worldwide.

Gregersen, H., & Jensen, C. S. (2002). Conceptual modelling of time-varying information. http://powerdb.net/database.

Gupta, C. B., & Gupta, V. (2009). *An introduction to statistical methods* (23rd revised edition). New Delhi: Vikas Publishing House.

Heather, A. W. (2004). *A quick reference of more than 300 XML tasks, Terms and Tricks, from A to Z.* New Delhi, India: Firewall Media, Laxmi Publications Pty Ltd.

Hori, M., & Ohashi, M. (2005). Applying XML Web Services into Health Care Management. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*.

Jennings, N. R. (2000). On agent-based software engineering. *Artificial intelligence*. *Vol. 117*. *Artificial intelligence* (pp. 277–296). Elsevier Publishers.

Kudyba, S. (2014). *Big data, mining and analytics: Components of strategic decision making.* New York: CRC Press.

Li, G. (2010, December). *World atlas of oil and gas basins.* USA: Wiley-Blackwell (496 pp.).

Longley, I. M., Bradshaw, M. T., & Hebberger, J. (2001). Australian petroleum provinces of the 21st century. In M. W. Downey, J. C. Threet, & W. A. Morgan (Vol. Eds.), *Petroleum provinces of the 21st century, AAPG memoir*. *Vol. 74*. *Petroleum provinces of the 21st century, AAPG memoir* (pp. 287–317). USA: AAPG.

Mattison, R. (1996). *Data Warehousing Strategies, Technologies and Techniques.* NY, USA: Mc-Graw Hill Publishers (100–450p).

Meersman, R. A. (2004). *Foundations, implementations and applications of web semantics, parts 1, 2, 3, lectures at School of Information Systems, CBS, Curtin University, Australia*.

Moody, L. D., & Kortink, M. A. R. (2003). From ER models to dimensional models: Bridging the gap between OLTP and OLAP design, part 1 and part 2. *Journal of*

Business Intelligence, 8(3)http://www.tdwi.org (Summer Fall editions).
Nimmagadda, S. L. (2015). *Data warehousing for mining of heterogeneous and multi-dimensional data sources.* Germany: Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG1–657.
Nimmagadda, S. L., & Dreher, H. (2012). On new emerging concepts of Petroleum Digital Ecosystem (PDE). *Journal WIREs Data Mining Knowledge Discovery, 2012*(2), 457–475. http://dx.doi.org/10.1002/widm.1070 (Wiley Online Library).
Nimmagadda, S. L., Dreher, H., & Rudra, A. (2015). *Big data hype in petroleum industries, Foundations, PPDM, Canada*.
Nimmagadda, S. L., & Rudra, A. (2005). *Data Mapping Approaches for Integrating Petroleum Exploration and Production Business Data Entities for Effective Data Mining, a paper presented and published in the proceedings of 3rd Kuwait International Petroleum Conference and Exhibition (KIPCE2005).* Kuwait City.
Nimmagadda, S. L., & Rudra, A. (2017). *Managing the Embedded Digital Ecosystems (EDE) using the Big Data Paradigm, a book chapter published in a book entitled "Big data and learning analytics in higher education, current theory and practice" Ben Kei Daniel.* Switzerland: The Springer International Publishinghttp://dx.doi.org/10.1007/978-3-319-06520-5.
Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.* O'Reilly Media, Inc. (27 Jul. 2013 - Computers - 414 pp.).
Pujari, A. K. (2002). *Data mining techniques.* Hyderabad, India: University Press (India) Pty Limited.
Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communications, 8*, 1690–1709.
Pyne, S., Rao, P. B. L. S., & Rao, S. B. (2016). *Big data analytics: Methods and applications.* Springer (16 Nov. 2016, Computers, 276 pp.).
Rudra, A., & Nimmagadda, S. L (2005). Roles of multidimensionality and granularity in data mining of warehoused Australian resources data. *Proceedings of the 38th Hawaii International Conference on Information System Sciences* Hawaii, USA.
Sagiroglu, S., & Sinanc, D. (2013). *Big data: A review in Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp.42–47), IEEE*.
Schermann, M., Hemsen, H., Buchmüller, C., Bitter, T., Krcmar, H., Markl, V., & Hoeren, T. (2014). *Big data, an interdisciplinary opportunity for information systems research.* Springer Fachmedien Wiesbadenhttp://dx.doi.org/10.1007/s12599-014-0345-1.
Vaishnavi, V., & Kuechler, W. (2004). Design research in information systems. Retrieved from http://www.isworld.org/ResearchDesign/drisISworld.htm.
Vaishnavi, V., & Kuechler, W., Jr (2007). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology.* NY: Auerbach Publications, Boca Raton, FL, Taylor & Francis Group.
Venable, J., & Baskerville, R. (2015). *MEDS: A methodology for evaluation in design science, a workshop.* Perth, Australia: Tech Park, Curtin University.
Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research, research essay. *European Journal of Information Systems, 25*, 77–89.
Witten, I. H., & Eibe, F. (2005). *Data mining practical machine learning tools and techniques, Morgan Kaufman Publisher* (2nd edition). New York: Elsevier.

Shastri L Nimmagadda is presently a research fellow at School of Information Systems, Curtin University, Australia. Shastri worked for the Schlumberger Company in multiple geo-markets worldwide as an Expert in Geosciences. Shastri worked for Wafra Joint Operations Petroleum Company in Kuwait (with Chevron Energy) and Santos JO in Sydney Australia. He previously worked for several petroleum operating and service companies in India, Australia, Uganda, Kuwait, Abu Dhabi, Egypt, Malaysia, Colombia, Indonesia and Russia. Shastri did his M Tech and PhD in Exploration Geophysics from the Indian Institute of Technology, Kharagpur. He obtained Master of Information Technology with distinction from the Curtin University of Technology, Australia. Shastri received his 2nd PhD in Information Systems, Curtin Business School, Curtin University. His basic and applied research interests are Big Data modelling, data integration, multi-dimensional warehouse repositories, processing, interpretation and knowledge mapping including research in domain applications. He is a professional member of AAPG, SEG, ASEG, SPE, EAGE, IEEE, AIS, AEG and IGU. He already published and presented more than 100 research and technical papers in various international journals and conference proceedings in the areas of geophysics, oil & gas exploration and information systems.