On Big Data guided Unconventional Digital Ecosystems and their Knowledge Management

(Completed research paper)

Shastri L Nimmagadda
Lincoln C Wood
Torsten Reiners
Dengya Zhu

Abstract

Establishing the reservoir connections is paramount in exploration and exploitation of unconventional petroleum systems and their reservoirs. In Big Data scale, multiple petroleum systems hold volumes and varieties of data sources. The connectivity between petroleum reservoirs and their existence in a single petroleum ecosystem is often ambiguously interpreted. They are heterogeneous and unstructured in multiple domains. They need better data integration methods to interpret the interplay between elements and processes of petroleum systems. Large-scale infrastructure is needed to build data relationships between different petroleum systems. The purpose of the research is to establish the connectivity between petroleum systems through resource data management and visual analytics. We articulate a Design Science Information System (DSIS) approach, bringing various artefacts together from multiple domains of petroleum provinces. The DSIS emerges as a knowledge-based digital ecosystem innovation, justifying its need, connecting geographically controlled petroleum systems and building knowledge of oil and gas prospects.

Keywords: Big Data, Unconventional Petroleum Reservoirs, Design Science, Digital Ecosystem, Data Mining, Visual Analytics
Introduction

In an ecosystem scale, a sedimentary basin is a bowl-shaped geosyncline (Castañeda Gonzalez et al. 2012), formed by multiple petroleum systems that hold large volumes and varieties of spatial-temporal Big Data. The gas/oil shales, coal bed methane, heavy oil and gas hydrate resources are prevalent as unconventional reservoirs (Li 2011). The Unconventional Digital Petroleum Ecosystem (UDPE) is envisaged as an ecological setting/system (Dovers et al. 2001), which may have embedded with numerous conventional and unconventional petroleum systems from large size sedimentary basins. For collaborations and exploring the connectivity between systems, we acknowledge the concepts of digital ecosystems as described in Li et al. (2012) and Chang and West (2006).

![Figure 1. An Unconventional Digital Petroleum Ecosystem (UDPE) Environment](image)

Each of these basins can generate and produce a large amount of data from oil and gas fields (Cleary et al. 2012; Durham 2013) and multiple petroleum systems. Each field may have numerous oil and gas producing wells; each drilled-well has different unconventional frackable reservoir pay zones, and each pay-zone is interpreted with different fluids - either oil or gas and both. In the data organization perspective, the attribute dimensions are characteristically represented in different hierarchies. We view that both conventional and unconventional reservoirs linked digital data coexist in a single petroleum ecosystem (Figure 1).

In a Big Data scale, a basin is an ecosystem with several composite schemas and associated supertype and subtype dimensions, interpreted with different oil/gas-plays (Nimmagadda and Dreher 2012). They are conceptualized and contextualized attribute dimensions especially in regions where there are no geological boundaries between basins. The entire narrative is a description of an unconventional digital petroleum solution. The focus of the current research is to articulate an integrated ecosystem framework, Design Science Information System (DSIS) with different artefacts evolved from the generic Design Science Research (DSR) approach (Venable et al. 2016). Design, develop, and implementation of multidimensional artefacts are the motivation of the DSR, and its adaptability is characteristic in digital ecosystem representation. In a digital oil field situation, the dimensions are logically structured with ontological descriptions that make the data relationships in different artefacts or schemas more connectable, flexible and extendable to larger scale ecosystems such as super basins (Nimmagadda and Dreher 2012). The vital phenomena are to understand the connectivity between conventional and unconventional petroleum systems. As demonstrated in Figure 1, the digital ecosystem issues are resolvable, in addition to the semantic and syntactic data challenges through various design science artefacts (Vaishnavi and Kuechler 2007). In that analogy, the authors interpret the sedimentary basin (Li 2011) as digital petroleum ecosystem, in which the digital data in multiple dimensions are brought together for integration in a warehouse environment. In the present research, issues associated with existing digital data organization, significance and motivation of proposed methodologies and how they address the presentation of digital petroleum ecosystem challenges are provided. Different artefacts articulated...
in the integration framework are evaluated, defending their development in the unconventional digital ecosystem scenarios.

**Literature Review and Research Gaps**

The literature survey explores the research gaps in the current application domains. In comparison with biological ecosystems, Li et al. (2012) describe the concept of a digital ecosystem and how its interdependent systems interact and exhibit as self-organizing, scalable and sustainable behavior in various application scenarios. Yu (2008) provide various domain applications that use digital ecosystems and technologies. Sabry and Krause (2012) explore developing digital ecosystems by providing power consumption models over cloud computing as analytical tools for managing energy optimizations. Henningsson and Hedman (2014) develop a digital ecosystem technology transformation framework for integrating theories of business and technology ecosystems. They illustrate the use of framework through a case study of transformation in the digital payment ecosystem. Porteous and Morawcynski (2017) illustrate the digital ecosystems as super-platforms, by citing Google, Facebook, and Amazon and Chinese Companies, Alibaba Group and Tencent Holdings. They focus on financial and customer ecosystems. Hasan and Kazlauskas (2009) develop a theoretical framework to explore the ecosystem approach to investigate and practice in ICT issues in the climate change problems. The cloud vendors make use of principles of digital ecosystems providing a paradigm to address distributed computing, resources management and security requirement challenges (Briscoe and Marinos 2009). So far the literature survey discusses concepts of digital ecosystems, theoretical framework, design and implementation in business, energy, climate and services sectors. We interpret similar concepts of digital ecosystems as a constituent of multiple domains with the representation of entities and dimensions (Dovers et al. 2001). Unconventional reservoir ecosystems possess diverse domains, and we adopt a design science information system framework, articulating the unconventional reservoir ecosystems and their management with information system artefacts.

**Problem Statement, Issues and Challenges**

Without the knowledge of the data, an uncertainty exists interpreting favorable connectivity between systems in a commercial petroleum province. Other challenges are poorly managed data integration methods. Inadequate spatiotemporal information and knowledge on areal extents of systems make oil and gas exploration operations risker, for which knowledge-based decision supportive digital ecosystem solutions are required. The data characteristics as interpreted in Figure 1 can support to an assembly of relational, hierarchical and network type of complex spatiotemporal data structures (Khatri and Ram 2004; Ozkarahan 1990) and applications in the depiction of multiple ecosystems. But, the heterogeneity of the data sources in multiple domains poses critical data and visual analytics challenges. In addition, with increasing exploration and production activities (Durham 2013) in petroleum-bearing provinces, the sedimentary basins emerge with volumes and varieties of Big Data sources in many upstream petroleum companies. Massive storage devices are needed for warehousing the data instances in the petabyte scale (Nimmagadda et al. 2019). In data modelling scenarios, managing thousands of such attribute data instances, connecting them with a similar number of fact tables is a tedious process (Coronel et al. 2011). Integrating multidisciplinary data in spatiotemporal dimensions is a significant challenge in managing digital petroleum ecosystems. For example, in a hierarchical data structure, the unconventional digital oil and gas fields are made up of: each field has a number of surveys, acquire volumes of data instances from millions of sensors in spatial dimensions (Castaneda et al. 2012). In several drilled-wells attributes, each well has a number of horizons (geological formations) and each survey has a number of survey profiles. The digital oil and gas fields can be characterized in hierarchical and relational domain ontologies (Shanks et al. 2003).

**Research Questions and Objectives**

The introduction and description of the problem statement motivate us identifying research gaps and framing the following research questions:

1. Why do we need knowledge-based DSIS as an unconventional digital ecosystem solution?
2. How do we manage the Big Data, easing the heterogeneity and multidimensionality of exploration data, and facilitating the petroleum upstream company? What is its scientific sanctity?
3. How do we evaluate and implement the unconventional petroleum ontologies as digital ecosystem solutions?

As a part of research question 1, for interpreting the connectivity between unconventional petroleum systems, Design Science Information System (DSIS) is needed to articulate different artefacts of Big Data in digital ecosystem solutions. Further, the research is aimed at implementing the Big Data paradigm in broader contexts of digital ecosystem scenarios and challenges. The heterogeneity and multidimensionality of the data are examined in multiple domains to analyze issues of systems’ connectivity. For example, multiple domains of the UDPE (Figure 1) may have closely related and unifiable attribute connectivity. Based on the research questions, the following research objectives are described:

1. **Articulate an integrated ecosystem framework:** Different domains, types and sub-types of attribute dimensions are identified. We propose a robust and flexible methodological DSIS framework (Venable et al. 2016), articulating the UDPE. It is further assessed how the DSIS articulations can effectively intervene the reservoir models for decision support systems.

2. **Share a common understanding of the structure of information and knowledge:** It is one of the common goals in developing domain ontologies with respect to multidimensional data structures. Domain experts, data analysts, oil and gas explorers and project managers should be able to share knowledge and use/reuse it in multiple domain applications. Different plot and map views with aggregated user queries facilitate their interpretation in new knowledge domains.

3. **Enable the use and reuse of constructs/models and domain knowledge:** Whether the developed artefacts have an impact on the overall DSIS framework in comparison with the UDPE needs to be ascertained. Models in several domains/systems need a new direction that signifies aggregated data view representations in spatial-temporal dimensions. The digital ecosystem representation includes the construction of models in time-and depth-intervals with their associated measures and units. If a group of explorers develops ontologies in detail, other domain experts can reuse them in related domains. As an example, the domain knowledge acquired in a particular field targets a specific model or vice versa and it may be used or reused in another similar/dissimilar ecosystem.

4. **Digital ecosystem explores the connectivity between multi-stacked reservoirs of the unconventional shale-gas basins:** Investors in petroleum industries explore for new and sustainable reserves in unconventional shale basins. This objective facilitates promotion of new opportunities and even the scope of developing carbon emission ecosystem contexts while implementing the digital ecosystem solutions.

**Significance and Motivation**

The shale gas, coal tar sands, tight gas sands and coalbed methane (Durham 2013) are unconventional resources, though they are abundant around the world, using the latest IS/IT tools, natural gas operators or explorers have not exploited the viable unconventional reservoirs. It is partly due to scarce and sparse geological and engineering information that can regulate the natural gas policy and market promotion. In addition, there is a shortage of expertise and technical know-how needed to develop the unconventional deposits successfully. The technology is so far limited outside North America (Durham 2013). The Big Data tools (Cleary et al. 2012) motivate us to develop new data integration methods in UDPE settings.

**Methodology**

Oracle-driven database software is used to build an integrated framework (Inmon 2005; Lee et al. 2006). The unconventional petroleum systems possess digital data to store in high-performance computing databases. For establishing the connectivity between systems, the data relationships are described as attribute dimensions in multiple domains. Within the context of a petroleum ecosystem, each element is interpreted as a dimension; at places, the connectable data attribute events are conceptualized and contextualized in geographic dimensions. Ontologies described as data relationships (including constraints and business rules) represent multidimensional data models, addressing the integrity and consistency including schematic and semantic heterogeneities. In other words, to analyze an unconventional petroleum ecosystem, ontologically described data relationships are interpreted for envisaging the reservoir connectivity. The notion of connectivity has come from the data integration process (Castaneda et al. 2012; Gilbert et al. 2004). The attribute elements, such as structures, reservoirs, seals, source and processes, such as migration and timing of occurrence or existence of the elements and oil and gas accumulations
from multiple systems and how they are inherently connected in the sedimentary-basin scale can describe a digital ecosystem representation. Referring to Research Questions 1 and 2, an approach exploring connections among multiple reservoirs and traps within a petroleum ecosystem is sought through investigation of chain attribute dimensions. The design of an integrated conceptualized framework, with compatible artefacts, systematize the research outputs through DSIS. However, high-level tools of Hadoop Distributed File System (HDFS) can match the artefacts of DSIS, for managing the unstructured Big Data sources (Shvachko et al. 2010). As shown in Figure 2, knowledge-based multidimensional data constructs and models are proposed for articulating them in the integrated DSIS framework (Indulska and Recker 2008; Vaishnavi and Kuechler 2007). The conceptualization and contextualization attributes of various entities and dimensions allow the data analysts to connect systems into innovative multidimensional repositories. The fine-grained multidimensional data structuring with added domain ontology descriptions may be required to make the DSIS approach more robust, flexible and adaptable in an environment, where varieties of business rules and constraints rapidly change, for which the constructs and models need constant updates.

Figure 2. DSIS Framework for Ecosystem Development, with Artefact Descriptions

The schemas described for Big Data characteristics, 7Vs attribute dimensions have been linked to the activities and outcomes of the DSIS research framework, to iterate the functions of the artefacts in the UDPE. The research activities and the deliverable outcomes are described in cuboid matrix form. For each research activity, the expected outcomes are depicted. The cells of the cuboid structure stores data or metadata about the Big Data linked petroleum digital ecosystems. Various sub-schemas are made, linking the framework to facilitate connectivity (Figure 2). Different attribute dimensions and their instances of Big Data are represented in various cells of a cuboid structure, as described in the DSIS framework. The design science approach has been discussed more in Vaishnavi and Kuechler (2007). For designing and developing the current research framework in the UDPE scenarios, different artefacts developed are discussed in the following sections.

**Petroleum ontology descriptions in Big Data assemblage**

A large amount of data are routinely acquired using digital portals in spatial-temporal dimensions (Khatri and Ram 2004). A sizable number of seismic sensors are laid onshore and offshore areas, where volumes of data are acquired...
in different formats. Managing the growth of volumes and varieties of data is no longer possible with traditional databases (Coronel et al. 2011). The conventional reservoirs of productive basins do produce oil and gas even without integrated workflows. Why are the conventional technologies setbacks in exploring and exploiting the unconventional digital ecosystems? In worldwide unconventional shale-gas basins multiple petroleum (information) systems do exist with a variety of elements and processes and volumes of their fact instances. For modelling, the fact instances are taken from storage and documentation media. In a Big Data project, these basins characterize with volumes and varieties of data, whereas the velocity and veracity attributes are represented as project performance tools. The variability is another attributing characteristic of the Big Data, contributing to interpretation in the form of anomalies between characteristics as demonstrated in Figure 3.

The anomalies may have been generated in between different Big Data characteristics. In Geology & Geophysics perspectives, the anomalies are the deviations of the geophysical responses due to physical properties of rocks hosting oil and gas deposits. The entities or dimensions emerge as conceptualized and contextualized attribute instances, detectable as divergent events from normal instances (Nimmagadda and Dreher 2012). The visualization is a kind of graphics artefact that can benefit the interpretation, and adding values to geological knowledge discovery. For example, instances associated with the Hydrocarbon Indicators (HCI) can attribute to seismic anomalies, providing possible development of porosities within unconventional petroleum provinces.

The growth of unstructured data in many applications motivates the authors to adopt the Big Data tools in unconventional digital ecosystem contexts. As illustrated in Figure 3, we interpret various Big Data dimension attributes, which we consider relevant to the data sources of unconventional digital petroleum ecosystems and their associated attributes. 7Vs are popular in describing such Big Data, which are characteristic, representing in spatial-temporal controlled digital petroleum ecosystems. Among them, volumes and types of data are typical in their representation as illustrated in Figure 3 and Table 1. As an example, we examine the unstructured data sources in various unconventional digital petroleum ecosystem scenarios as illustrated in Table 1.

Table 1. Big Data in an Unconventional Digital Ecosystem

<table>
<thead>
<tr>
<th>Field</th>
<th>VO</th>
<th>Size</th>
<th>VA</th>
<th>DM</th>
<th>CS</th>
<th>C</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nimm</td>
<td>15</td>
<td>90</td>
<td>20</td>
<td>90</td>
<td>15</td>
<td>1000</td>
<td>100</td>
</tr>
<tr>
<td>Shak1</td>
<td>12</td>
<td>72</td>
<td>10</td>
<td>55</td>
<td>13</td>
<td>1095</td>
<td>109</td>
</tr>
</tbody>
</table>
Field: Name of oil/gas field; VO: Number of Volumes; Size: gigabytes; VA: Varieties (Number); DM: Data Models (Number); CS: Composite Schemas (Number); C: Cells in Cubes; V: Number of Data Views.

The volumes and varieties of data, as identified in Li (2011) are used to build schemas and their structures. Typical schemas include star-, snow-flake and constellation types (Coronel et al. 2011). Based on varieties of data in multiple domains, various schemas are characterized. A particular schema is chosen that can make the data mining, visualization and interpretation effective. Models are integrated into the data warehousing environment, and the composite schemas are substantially reduced in small numbers, which can be easy to manage in warehouse schemas, as displayed in Table-1. Volumes represent geological and geophysical data, as characterized in spatial-temporal attribute instances in Big Data scale. Varieties are the types of data characterized in the volumes in diverse domains. Seismic, gravity, magnetic, electrical, electromagnetic, geochemical, reservoir engineering are typical varieties in geophysical domain representations. Similarly, several types of data are deduced in geology and reservoir engineering domains (Gilbert et al. 2004).

As per Research Question 2, we analyze the unconventional digital ecosystems that exhibit the heterogeneity and multidimensionality characteristics in the data systems. The amount of data and information considered in describing the data models and warehouse schemas are shown in Table 1, simplifying large size ontological structures and their descriptions. The seismic and drilled-well data systems with hundreds of attribute dimensions and volumes of fact instances are part of the Big Data. The data models represent various constructs, logical and physical schemas. The digital data derived from conventional and unconventional reservoir types may have been responsible inherently embedding multiple systems, implying that the data associated with these systems may have multiple links with geographic attribute instances. The unconventional shale gas, shale oil and coal bed methane (CBM) resources’ data (Creties et al. 2008) can conceptually be modelled and unified, even with sustainable conventional digital ecosystems. In this context, we interpret sustainability (Burke 2013) as a process of enduring the resource with its spatial-temporal dimensions, as illustrated in Figure 4. The challenge is modelling and integrating such real-time Big Data events, and their knowledge management in operating and service companies. The symbol indicates anticipated connectivity between systems. Each system may have multiple links with spatial dimensions. Typical data described in a unique digital ecosystem are shown in Figures 4a and 4b. The Number of surveys, drilled wells, and oil wells are distinctive attribute dimensions of the digital ecosystems and as indicated I, II and III, three peaks of attribute instances describe their strengths at different periodic intervals.

Figure 4. (a) Typical Data Used in the Modelling Process (b) Conceptualization of an Ecosystem Representing the Unconventional Big Data Sources

Oil and gas fields occupied by large geographic regions are often in terabyte scale (Nimmagadda et al. 2019; Li 2011). The concept of a digital ecosystem in the DSIS context is to bring data attributes together from multiple domains and connect through different multidimensional schemas. The seismic and borehole data attributes and
their instances are usually represented together in spatial-temporal dimensions. The standardized conceptual model for the internal representation of petroleum ontologies is based on the work done in Eapen (2008). The UDPE precisely describes a series of events in the contexts of petroleum exploration that can be used to depict the reservoirs, or drillable fracks in any unconventional petroleum field. The petroleum ontologies in the UDPE contexts describe:

1. Unconventional reservoir or/ seismic sequence and structure information.
2. Structure or /unconventional reservoir integration process.
4. Unconventional reservoir/ or structure internal and external data associations to petroleum accumulations.
5. Constraints, affecting the final unconventional petroleum trap/s or seal confirmation.

One of the schemas presented in Figure 5 demonstrates the connectivity between various attribute dimensions. For building multiple data structures, volumes of data instances are considered articulating the DSIS framework. In knowledge-based warehouse repository, hundreds of logical multidimensional data relationships do emerge while analyzing digital ecosystems and connecting them to the DSIS framework. In the data schemas, there are multiple dimensions narrated and interpreted conceptually with various relationships, physically linking the fact tables in one-to-one, one-to-many (as indicated in a symbol—) and many-to-many data relationship types (Coronel et al. 2011). As demonstrated in a dimensional model in Figure 5, two fact tables are connected through common attribute dimensions. For storage, adaptability, flexibility of dimensions, the snowflake schemas are used in the modelling. Though it is easier to implement the schemas, performance may have been reduced. However, we limit the number of schemas, which may be optimum to achieve the connectivity including their use and reuse may be extended in diverse domains, depending upon the varieties existing in data volumes.

![Figure 5. Star-Schema Model Representing Semantics Relationships and their Connectivity Among Various Attributes of the UDPE](image-url)
remove the potential conflicts and constraints arising during conceptualization and interpretation of terms in various data sources and provide a structured vocabulary that integrates all data and knowledge in a unified domain (Nimmagadda et al. 2019). Petroleum ontologies specify different topological relationships among structure, reservoir/source, to illustrate how mandatory semantic constraints are represented. In the case of an unconventional reservoir, the source does not have a necessary constraint with trap or structure but has with fracture-reservoir (present within the source rock itself), and every petroleum source must topologically touch one or more instances of a fractured reservoir. Producing unconventional oil or gas field must contain at least one fracture-reservoir or frac-connectivity among several fractures, unlike reservoirs do have a mandatory relationship with the trap in a conventional producing field.

Integration of unconventional petroleum ontologies

Tools and concepts used for designing and developing data-warehouses in different domains are given in Inmon (2005) and Lee et al. (2006), and they are in correspondence with the composite schemas presented in Figure 5 that support our current data integration process. As described in Research Questions 1-3, the need to integrate petroleum systems’ data from multiple systems and sources is described in Castaneda et al (2012). The data warehouse approach brings together unconventional petroleum systems’ digital data from different oil and gas fields. In ecosystem scenario, data from geological, geophysical and geochemical domains are integrated into a data warehouse repository. The data warehouse approach (Inmon 2005) can track the connections between petroleum systems in spatial-temporal dimensions. It also allows processed (knowledge-based) data to share among professionals and even distribute geographically in the form of digital clouds (Briscoe and Marinos 2009). As shown in Figure 6, the Big Data-guided DSIS is framework articulation of the UDPE setting, formulated for implementing research outcomes and validated by various evaluation properties (Venable et al. 2016). The artefacts are building blocks of the overall UDPE setting with which the DSIS articulations are customized.

**Figure 6. A Schematic Workflow of Big Data Organization**

In addition, the integrated articulations allow different petroleum systems communicate and interact each other directly within framework. As illustrated in Figure 6, the data warehouse designers not only get the support of the ecosystem concepts, but define the scope, depth, comparability and accuracy of data entering into the warehouse. The range of data refers to types of petroleum systems, and their linked geological, geophysical and geochemical
data in multiple periods (time-dimension), and geographic locations (space-dimension). The depth of data refers to the level of details needed in the modelling. For adaptability, data in similar and dissimilar attribute dimensions, separate sites should adopt to same classifications. No matter how different data are collected across sites, it is paramount to integrate data schemas from multiple sites and domains in the warehouse repositories. As per Research Question 2, for easing complexity and security, it is crucial that petroleum system analysts and geo-modellers use compatible software systems to map and model data in a way, they are compatible to repositories (Figure 6). Accuracy is desired for types of data in any given geological situation, and it is a requirement for agreeing to a credible metadata in the warehouse. A sedimentary basin is a typical example, in which the data dimensions are from seismic, drilled-well, petrophysical and production domains and they are connectable to warehouse repositories through DSIS articulations. The framework described in Figure 6 allows us generating the necessary logical data structuring models to interconnect various domain ontologies and their structures.

**Implementation of the Framework**

As a part of Research Question 3, several logical rules and constraints applicable for scaling UDPE models and multiple levels of information stored in ontology models are incorporated in the warehouse schemas. Further, for queries, information retrieval and presentation, we consolidate the metadata outcomes in different attribute data cubes as illustrated in Figure 7 in the following sections. For implementing DSIS articulations (Nimmagadda et al. 2019), the data views retrieved from metadata cubes of UDPE are analyzed for any data correlations, patterns and trends. Data visualization and interpretation artefacts are examined for attribute analysis and geological knowledge discovery in ecosystem contexts. Largely, the DSIS has a decisive role in strategizing the exploration and field development in particular in digital reservoir solutions, making huge impacts in the integrated interpretation projects, especially during prospect identification and risk evaluation stages (Nimmagadda and Dreher 2012).

**Validation of DSIS Approach in the UDPE**

The data interpretation and knowledge discovery are the final stages of reservoir management project, facilitating the implementation of the DSIS (Research Question 3). The analysis of reservoir distribution in different knowledge domains of UDPE and basin scenarios must support the artefacts used in the DSIS framework. The multiple reservoir connections and their areal extents are interpreted with seismic attributes at volume and surface levels. Interpretation of different map and plot views support the frackable shale reservoir connections explicitly at field and basin levels. As shown in Figure 7, the computed seismic attributes are evaluated, especially the reservoir cubes typically presented in space and depth dimensions. The volume and surface attribute cubes are used for mining slices and dices (Pujari 2001) for interpretation of new geological features attributable to UDPE in the study area. Various data mining schemes (Matsuzawa and Fukuda 2000) are considered for ascertaining the reservoir potentiality in the unconventional reservoir contexts.

![Figure 7. Exploring Unconventional Fracture Reservoir Connections among Attribute Cubes](image-url)
**Ecosystem Visual Analytics and Big-Knowledge**

Data transformation tools, such as statistical regressions, bubble plot graphical representations can translate the Big Data into big-knowledge. The data instances interpreted for structure, reservoir, source and seal elements (Figure 2) are documented in several classifications. Set theory and data classification approaches (Matsuzawa and Fukuda 2000) suggest that several permutations and combinations are possible to make sense of each grouping of elements and processes and interpret each petroleum system in the ecosystem scale. For example, the geological structure and its attitude attribute instances are used in building structure related ontological descriptions. Similarly, the specific reservoir attribute instances documented for an interpreted horizon (of fracks) in a basin are structured (in data structuring sense) through reservoir ontology descriptions (Shanks et al. 2003). For finer details of digital ecosystems in petroleum industries, we emphasize fine-grain structuring for effective data mining. Slicing and dicing are usually performed on metadata for tracking knowledge-based attribute dimensions interpreted in basin contexts. The strength of element and process attributes of petroleum systems and their connectivity are assessed through high resolution map views. Frackable reservoirs interpreted in structural compartments appear to have reservoir connections with associativity between fracturing attributes as illustrated in Figure 8a.

**Rule-Based Decision-tree Mining Model**

For implementing and evaluating the models, a Microsoft decision tree mining with grapher tools is used to explore and analyze the information from homogeneous branches of a tree, finding mining rules, with conditional controls and constraints. A decision tree mining is a classification scheme (Pujari 2001) that generates a tree structure with a set of mining rules or business constraints, representing models in different classes in a given dataset. As an example, in unconventional reservoir plays, based on porosities and kerogen content, several numerical and categorical attributes are deduced for evaluating the fracks (fractured reservoirs). Hydrocarbon plays and non-plays have common categorical attributes in the data cubes. Several leaf nodes are described in the decision-tree; each leaf node represents a mining rule.

The decision tree rules deduced from a model in Figure 8b are:

1. Rule1: If the shale has more than 5% porosity, it is a **play**.
2. Rule 2: If the TOC is less than 5%, the shale has a **play**.
3. Rule 3: If the TOC and porosity are each less than 5% and 5%, the permeability is less than 0.1%, the shale **play** has a relatively poor fractured reservoir
4. The Kerogen type III more than 3%, with less than 5% porosity, permeability less than 0.1% and TOC less than 5% appear more favorable
5. Rule: 4: Based on Rule 1, Rule 2 and Rule 4, if the shales are good fractured reservoirs, then Rule 4 holds good.

![Figure 8.](image.png)
6. Rule: 5: If the attributes are not favorable as narrated in Rule 3, then the shale play does not hold good.

The data available in the public domain (Li 2011) are used to test the accuracy of the classifier. Accuracies of rules are calculated from training, test datasets and Rule 4 emerges 90% accurate, compared with the accuracy of the other rules. TOC and permeability are other weighing attributes supporting Rule 4. Different multidimensional shale reservoirs (reservoir plays from various fields) and their properties have been analyzed. Each attribute has unique shale reservoir properties. Though the strength of individual characteristics matches with the specific play-dimensions, the overall attributes are different as demonstrated in Figures 8b and 8c. It is a schematic view drawn for a frack model, aiding the decision support system, which type of fracking and kerogen content best classify and suit to making a valuable financial deal with unconventional digital oil solutions.

**The Visualization and Value attained from Big Data Analytics**

The metadata models are further analyzed for meaningful geological information; exploration products are presented for interpretation and knowledge discovery. The potentiality of digital ecosystems is examined for the value of knowledge. As described in Table 2, the hydrocarbon potential exists in the shale bearing digital ecosystems. Production Index (PI) indicates the maturity of the unconventional ecosystem that can produce commercial and viable oil and gas resources.

**Table 2. Potentiality of Unconventional Digital Ecosystem**

<table>
<thead>
<tr>
<th>Field</th>
<th>T (ft)</th>
<th>K (mg/g)</th>
<th>PO (%)</th>
<th>PE (mD)</th>
<th>SW (%)</th>
<th>PS (%)</th>
<th>T</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nimm</td>
<td>1</td>
<td>150</td>
<td>1.5</td>
<td>150</td>
<td>70</td>
<td>500</td>
<td>45</td>
<td>0.18</td>
</tr>
<tr>
<td>Shak1</td>
<td>2</td>
<td>300</td>
<td>2</td>
<td>245</td>
<td>60</td>
<td>400</td>
<td>35</td>
<td>0.20</td>
</tr>
<tr>
<td>Kak1</td>
<td>3</td>
<td>550</td>
<td>0.5</td>
<td>320</td>
<td>75</td>
<td>600</td>
<td>60</td>
<td>0.35</td>
</tr>
<tr>
<td>Mak1</td>
<td>5</td>
<td>600</td>
<td>1.25</td>
<td>455</td>
<td>80</td>
<td>700</td>
<td>55</td>
<td>0.46</td>
</tr>
</tbody>
</table>

T: TOC (%); K: Kerogen (mg/g); PO: Porosity (%); PE: Permeability (mD); SW: Water Saturation: percentage; PS: Pressure (PSI); T: Reservoir Thickness (ft.); PI: Production Index. The instances extracted from Big Data derived metadata is tabulated in Table 3. Quality of structures or traps and the frackable reservoirs that hold the hydrocarbon content is evaluated.

**Table 3. The Value Extracted from the UDPE**

<table>
<thead>
<tr>
<th>Field</th>
<th>Number of Structures</th>
<th>Number of Fracks</th>
<th>Number of Wells</th>
<th>Volumes (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nimm</td>
<td>5</td>
<td>158</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Shak1</td>
<td>4</td>
<td>245</td>
<td>11</td>
<td>150</td>
</tr>
<tr>
<td>Kak1</td>
<td>6</td>
<td>357</td>
<td>9</td>
<td>154</td>
</tr>
<tr>
<td>Mak1</td>
<td>3</td>
<td>634</td>
<td>7</td>
<td>200</td>
</tr>
</tbody>
</table>

MB: Estimated million barrels of oil equivalent.

For a sedimentary basin, the knowledge-based domain ontologies are described in several multidimensional data structures and their strengths evaluated in the UDPE business contexts (Shanks et al. 2003). We have plotted porosity attributes of multi-stack unconventional reservoirs in an investigating area in a bubble plot view (Figure 8a). The size of each bubble has significance in terms of its characteristic porosity attributes, described for an unconventional reservoir in a sedimentary basin. Interestingly, in the current study area, the unconventional reservoirs follow certain porosity attribute trends, allowing the explorers to explore new opportunity areas of prospective locations. It has been possible because of smart DSIS framework articulation and its implementation in Big Data scale, where large size geographically controlled frackable reservoirs are successful exploration targets. One of the characteristics of the unconventional digital ecosystem, PI is analyzed in the context of shale gas potentiality. The PI is a composite attribute, made up of free hydrocarbon percentage and residual petroleum contents.
potential of the source element (Table 2). PI of unconventional formations (geological) ranges from 0.16 to 0.46, with an average of 0.25, indicating the matured source is sufficient enough for generation and production of hydrocarbons in the shale gas investigating area.

The Conclusions, Limitations and Recommendations

The documentation, organization and integration of Big Data sources and their characteristics support the relevance and representation of unconventional digital ecosystems in the UDPE. Description of domain ontologies, their intelligent storage and integration in a warehouse environment are unique for digital ecosystem representation. Data modelling, schema selection, data warehousing and mining and visualization and interpretation artefacts articulated within the DSIS framework bring together the unconventional digital ecosystems. Different domain ontologies deduced for data sources and integrated with shale gas, shale-gas processes, structure, reservoir capacities and geologic characterization attributes are added values to unconventional resource projects and their research outcomes. Feasibility and applicability of exploration and field development are further assessed for each sedimentary basin with potential hydrocarbon bearing geological structures. Advantages of the use and reuse of data structures are emphasized. The effectiveness of data mining and interpretation of data views drawn from data warehouse depends on logical modelling and mapping of multiple data dimensions and their attributes. Attribute dimensions such as structure (including fracks causative to faulted structures and their compartments), reservoir, and seal appear critical in assessing the potentiality of the existence of unconventional digital ecosystems and associated shale-gas plays. We recommend implementing the methodology in such basins where geologically favorable shale gas prospects exist. A common agreeable conceptual representation of unconventional petroleum ontology is needed to resolve the heterogeneity in petroleum data sources to make meaningful data and information exchange among geographically located data warehouse repositories. Other inferences are:

1. Multidimensional data structures created in a sedimentary basin scale, are more flexible, and they are usable and reusable in different knowledge domains. However, the data qualities and their reconciliation are challenging while placing fact instances in the data models.
2. Data warehousing in the current context is suitable when considering large size sedimentary basins that comprise of multiple petroleum systems in countries where various oil and gas fields share common elements and processes to make connections between systems through digital clouds and their computing nodes.
3. The articulations of Big Data-guided DSIS framework, simulated as a UDPE setting, is a way-out guiding the petroleum explorers as knowledge-based digital reservoir solutions. The constructs, models, and methods are the final deliverables of the current research application.
4. Interoperability is assessed in the form of capacity to hold two or more petroleum systems by composite schemas in different sedimentary basins in a petroleum province. Data attributes and properties drawn among several dimensions show significant associations, trends and relationships among multiple data attributes such as structure, reservoir, production and other geological, geophysical and geochemical attribute dimensions considered for visual analytics and interpretation.

References


Cleary, L, Freed, B, and Elke, P. 2012. “Big Data Analytics Guide: 2012”, Published by SAP, CA 94607, USA.
Big Data Guided Digital Ecosystems and Knowledge Management


Durham, S. L. 2013. An unconventional idea, open to interpretation, AAPG Explorer, March Series, USA.


