POKMS: A Process-oriented and Ontology-based Knowledge Management System in Exploration and Development

YANHONG ZHAO[1], HONGQI LI[1], LIPING ZHU[1] and RUI HAN[2]
[1]Petroleum Engineering and Data Mining Lab, College of Geophysics and Information Engineering
China University of Petroleum (Beijing)
Geology Building 605, Fuxue Road NO.18, Changping Dist, Beijing, 102249, CHINA
yh.zhao.cup@gmail.com
[2]Institute of Computing Technology, University of Chinese Academy of Sciences, Beijing, CHINA

Abstract: - As information technologies have become widely applied enabling technologies in different engineering disciplines such as oil/gas exploration and development, various (especially legacy) applications are deployed with different data models and they constitute a business process that must be accomplished through the coordination and cooperation of different specialized departments. However, heterogeneous data sources together with different data models give rise to two key challenges in automating the process. First, acquiring and standardizing data is error-prone. The second, and more important, challenge is that small human errors will affect decision significantly. Without a domain-specific business process enabling platform, decision makers could not acquire progresses on each activity in time, and thus delaying the transferring and sharing of the produced results among different departments. In this paper, we propose POKMS, a process-oriented and ontology-based approach for knowledge management and heterogeneous data integration. POKMS first automatically builds the exploration and development ontology from the epicentre data model, which is a global petroleum industry data model published by the petro-technical open standards consortium. Subsequently, POKMS uses the domain ontology to define the basic entities in business process modeling. Following this approach, we built a web-based knowledge service platform providing graphical tools for users to visually design the business processes. Using this platform, decision makers can acquire progresses on each activity instantly, thereby timely transferring and sharing results produced by each activity among departments, experts, and non-experts. Furthermore, end users, in particular non-experts, can reuse the domain knowledge and monitor the processes of the ongoing projects, and thus help improve the efficiency of decision making.

Key-Words: - Ontology; Heterogeneous data integration; Business process modeling; Knowledge management.

1 Introduction

To adapt to the rapid development of information technology, various (especially legacy) applications are deployed with different data models. These applications form a business process that need to be accomplished through the coordination and cooperation among different specialized departments. Facing a large number of heterogeneous data models, a general model for data exchange is urgently needed. Data integration provides a mechanism to logically integrate data of various and heterogeneous sources to realize data exchange and sharing across different applications. Meanwhile, in the knowledge economy, knowledge is a major driving force for organizational change and wealth creation, hence effective knowledge management is an increasingly important source of competitive advantage and a key to success of modern organizations [1]. Nevertheless, knowledge does not exist in isolation, it is created and utilized during the execution of business processes, and therefore it can only work well within the context of business process combination. A business process is a combination of particular professional activities of creating values. Business process management can optimize enterprise operation processes, promote the collaboration of all departments, improve working efficiency, and reduce cost of the business.

In the field of exploration and development, there are two challenges in automating efficient process management with knowledge sharing. First, there are heterogeneous data sources with different data models, and extracting ontology from these data sources is very time consuming. For example, Smalley and Espeland in BP pointed out it took BP geologists 44% of time to find data and control data quality, and 50% of time to interpret and analyze [2]. The second issue is how to efficiently facilitate management of business processes and process-related knowledge. Without an enabling domain-specific business process platform, decision makers could not acquire progress on each activity in time, and thus results produced by each activity could not be transferred and shared between departments smoothly. For such an issue, combining automatical data acquisition, domain-specific business process, and process knowledge in a cooperating environment is of great significance.

Previous work fails to solve the above issues. First, the automatic or semi-automatic ontology construction from the existing resources, especially RDB [4, 5], is facing a large amount of heterogeneous data schemas. This means the quality of ontology extracted from RDB is difficultly to guarantee . i.e., data duplication and redundancy still occur in many legacy databases. Second, traditional business modeling methods only support diagrammatic and mathematical modeling [6], and no systematic and consistent approach is currently available to represent the business processes in the oil exploration and development filed. Moreover, process-related data are simply defined by the entity type and the relations between entities are seldom considered [7, 8]. Third, state-of-the-practice business and knowledge management cannot achieve integrating automatically data acquisition, domain-specific business processes and process-related knowledge [7, 9 and 10]. In addition, no graphical process representation tools are available for users to design the business process. Therefore, these systems cannot be applied to the oil exploration and development filed due to their limitations and domain-specific requirements. Table 2 summarizes the limitations of state-of-the-art and state-of-the-practice systems.

Our contributions are three-fold as follows. First, we employ the epicentre data model [11] as the standard for data exchange and data sharing among different applications in oil exploration and development. We then create the mapping rules between the epicentre data model and Web Ontology Language (OWL) ontology to automatically extract concepts, attributes, and relations for the ontology construction.

Meanwhile, we propose a business process model, which divides each business process into activities. Each activity is composed of input, output, role and some processing logics. We use ontology to describe the basic entities related to each activity, such as input and output entities. Meanwhile, the process-related knowledge and the process instance knowledge are also represented by ontology. A process-oriented knowledge net is formed through the relations defined in the exploration and development ontology. This net provides a common data interpretation for different applications and realizes the data exchange and data sharing during process execution.

Finally, we have designed and implemented a working process-oriented knowledge management system, called POKMS (Process-oriented and Ontology-based Knowledge Management System), to include ontology maintenance, business process management and knowledge query modules. Currently, this system has been deployed at Shengli oilfield in China.

This paper is structured as follows. We review related work in Section 2. In Section 3, we describe our approaches on ontology construction and process modeling. In Section 4, we introduce the knowledge service platform in detail. We draw conclusions in Section 5.

2 Background and Related Work

To address the problem of heterogeneous data sources with different data models, a proper data exchange standard is urgently needed. Ontology, which provides a common understanding of information in a field among individuals or organizations, is considered to be the very effective way to solve the problems. How to realize fast ontology construction is a hot research focus. At present some process-based knowledge management systems are available, can they be applied to the oil field? This chapter reviews some main data exchange technologies, automatic or semi-automatic ontology construction technologies and approaches of recent process-based knowledge management.

2.1 Data exchange technologies

The management and application of massive heterogeneous data greatly hinder the construction of dataset information system in the oil field. Facing the continuous improvement of data models and business processes, a standard data model is urgently required to realize the heterogeneous data exchange and sharing of oil field. Many traditional methods are available for data exchange, such as Federated database system [14], Mediator system [15] and data warehouse [16]. With the development of information technology, various new technologies have been applied in data integration, like XML [17], CORBA [18], Web service [19], Ontology [3, 20], etc. XML is one of major data exchange standards in the internet which integrates data in different data sources to realize data transmission and sharing in heterogeneous platforms. POSC [12] is a petro-technical open standards consortium which is set up to realize the information sharing and cooperation between different companies and professions. The technical goal is providing a set of standard specifications for exploration and development applications. POSC published some xml-based projects to support intelligent data exchange and display, such as GeophysicalML, WellLogML, LogGraphicsML and ProductionML. However, the xml-based data exchange standard has its limitations on semantics and the unity of concepts. Ontology, as someone mentioned is the "explicit formal specification of shared conceptualization" [21], which is applied in order to provide a common shared comprehension of an information in a field among individuals or organizations with the benefit of formally defined and description logics supported. OWL is used to describe ontology. Because owl has formal semantics, automated reasoning is possible. RDF, of which OWL is an extension, also has a dedicated query language namely SPARQL. Ontology is considered to be the very effective way to solve the problems of semantic heterogeneity and interoperability [3]. It does not only define the syntax of data sharing, but also reflects the semantics of the data by using a shared and extendable domain-wide model. The comparison of different data exchange technologies are shown in Appendix 1. In this article, we use ontology to address the problem of heterogeneous data sources with different data models.

2.2 Ontology construction

Ontology can be built artificially using ontology editors such as protégé or by leveraging some data model specifications for (semi-) automatic ontology construction. Manual ontology construction is an expensive and time-consuming task, which needs a complete engineering, systematic approach to support and a certain number of domain experts to participate in. Fast ontology construction is the key issue for the development of knowledge-based applications. Thus, automatic or semi-automatic ontology construction has become a hot research focus. Using existing resources to realize (semi-) automatic ontology construction has been paid more attention. The resources may be relational database, text documents, web content, and etc.

Much work has been done in extracting ontology from relation databases. The main techniques are based on reverse engineering, schema mapping and data mining. Reverse engineering used conceptual data model, i.e., the Entity-Relationship (ER) as source, and built ontology based on ER transformation by using the graph transformation operations. In the schema mapping technique, it converts relational database schemas to ontology by some predefined mapping rules. Data mining approaches are used in RTXON which exploited the content of the databases to identify categorization patterns, and combined a classical schema analysis with hierarchy mining in the data [22].

Automatic ontology construction from text or literature mainly includes three stages, i.e., text preprocessing, ontology extraction, and taxonomic relationships acquisition. Natural language processing and information extraction techniques are applied to acquire and classify ontology instances. Statistic patterns are used to extract words [23] and semantic similarity for ontology clustering [24].

Using Wikipedia for semi-automatic ontology construction is adopted in some literatures [25, 26, 27 and 28]. Wikipedia is a large and valuable source of semantic information; it can serve various tasks including information extraction, information retrieval, question answering systems, and ontology building. The ontology learning procedure from Wikipedia is carried out as follows: (1). obtain the Wikipedia in the specific domain, (2). acquire the concepts and instances based on Network Category Structure, Information Box or Definition Sentence, (3). obtain ontology relations through pattern matching or statistical learning. However, the concepts and instances are difficult to distinguish in the process of automatic extracting concept, since overlapping relationships widely exist.

The techniques of automatic or semi-automatic ontology construction mentioned above own the following weakness as follows: (a) the various and heterogeneous data models have a direct impact on ontology quality, and data duplication and redundancy still occur in most of legacy databases [4]. Moreover, the ontology extracted from RDB must be maintained manually by domain experts, (b) precision and recall are key actors in ontology evaluation, however, the ontology constructed from the text must be revised according to the errors in the aspects of natural language processing and related statistics techniques, (c) lack of the domain knowledge in the process of extracting knowledge in the database or web.

2.3 Process-based knowledge management

Knowledge is generated and utilized during the execution of business processes. Vacuum doesn't necessary help creation, sharing and reuse of knowledge. It can only work well under combination with specific business flow and reuse of knowledge. To combine business process with process knowledge, it is necessary to find a proper modeling language to represent the business processes. At present, the representation of business processes has already been a major subject in several research domains, including business process reengineering, workflow management, and software engineering [29, 30, and 31] , and many proven products and technology are used to model the business process such as Unified Modeling Language, Business Process Modeling Language, Petri-nets, Rational Rose, and Java Business Process Management. Nevertheless, despite the fact that many of these tools are useful for understanding the business processes, they only support diagrammatic and mathematical modeling [6], and no systematic and consistent approach is currently available for representing the business processes in the oil filed. While workflow is an important technology for business process modeling, reengineering and execution, it puts more emphasis on business process and does not support the general knowledge management [8]. Moreover, process-related data are simply defined by the entity type and the relations between entities are seldom mentioned.

Integrating knowledge management (KM) with business process management (BPM) has been investigated for various disciplines such as product design, project management and software development.

The MILOS system is an internet-based process-centered knowledge management environment [32, 33]. It integrates project planning and workflow technologies over the Internet, but put more emphasis on them. Jung, Choi, and Song proposed an integrated architecture for KMS and BPMS based on a comprehensive framework [7]. They integrated the extended functionalities of existing KMSs and BPMSs according to the lifecycle requirements of both knowledge and business processes.

Han and Park developed a PCKMS which is extended from the PPMF (Process-based Performance Measurement Framework) [34]. They proposed the framework for process-centered knowledge model and enterprise ontology for the context-rich and networked knowledge storage and retrieval required during task execution. Savvas and Bassiliades proposed a process oriented approach, and build a web-based knowledge management system [10]. The system employed ontology in OWL for representing the public administration structure. Kwan Hee Han and Jun Woo Park proposed a framework for process-centered knowledge model and enterprise ontology for the context-rich and networked knowledge storage and retrieval required during task execution, a process-centered KMS [9]. Ri Hai and Manfred Theißen presented a generic and extensible modeling language which was defined in form of ontology for different types of work processes [36].

Daiyi Li introduced a representation method which combined domain ontology and task ontology based on crop cultivation standards (CCS) [37]. Bi long Wen proposed a method of building Petro-Onto (petroleum exploration and development domain ontology). The top-level ontology is designed， and an approach is proposed to automatically capture concepts and the relationships among concepts from business model and data model [38].

The earlier work mentioned above still have the limitations as follows. First, the traditional business modeling methods only support diagrammatic and mathematical modeling [6], and no systematic and consistent approach is currently available for representing the business processes in the oil filed. Moreover, process-related data are simply defined by the entity type and the relations between entities are seldom mentioned. Second, the existing information systems for business and knowledge management could not realize the integration of automatic data acquisition, domain-specific business processes and process-related knowledge. Especially, in these systems no graphical process representation tools are provided for users to design the business process. Therefore, these systems could not be applied to the oil exploration and development filed because of their limitations and domain-specific requirements. Please refer to Appendix 2 for the summary of limitations of state-of-the-art and state-of-the-practice systems.

3 Our proposed approach

Fig. 1 gives an overview of our approach, three parts are included: ontology construction, ontology-based process modeling and process-oriented knowledge management. First, we use the epicentre data model as the standard for data exchange and data sharing among different applications in oil exploration and development, and then create the mapping rules between epicentre data model and OWL ontology to automatically extract concepts, attributes, and relations for the ontology construction. In the process modeling, we use ontology to describe the basic entities related to each activity, such as input and output entities. Although the static relations are defined in the ontology, the dynamic relations between different entities can be generated by the business activities. When the business process is instantiated, the process case is monitored and the process-related knowledge is transferred and stored.

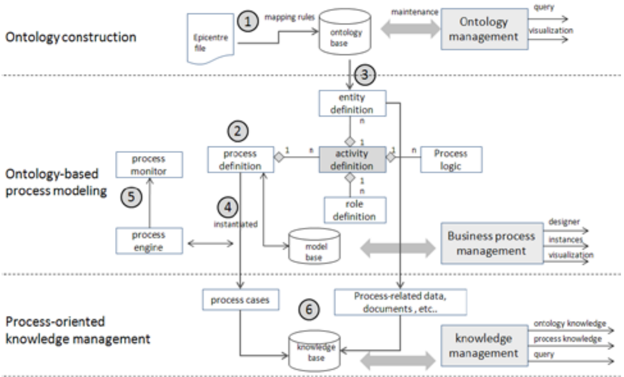


Fig. 1 The overview of our research approach

3.1 Ontology construction from Epicentre Data Model

Ontology is domain-specific. In this section, we build exploration and development ontology from the epicentre data model, which is published by POSC. Epicentre is a global petroleum industry data model, and defines all the logic data models in express language (an international standard information description language) [13]. More than twenty internationally renowned companies directly participate the model building. The epicentre data model covers all the objects, activities and properties at each stage in the oilfield lifecycle. Transforming the epicentre data model into ontology is significant because it is domain-specific, normalized and formalized. In our approach, we create schema mapping rules between Epicentre data model and owl formal language.

The epicentre data model covers all the objects of exploration and development which are organized by the objective relationship rather than the discipline. Epicentre is an object-oriented, hierarchical and structured data model composed of activity, property, object_of_interest, association, geological_process, ref_data, etc. The top entities are described in Fig.2.

Separating the properties from an object is a success in the epicentre data model. An activity acts on object and produces properties. The relations are shown in Fig. 3. The property in the epicentre data model starts with prefix pty.

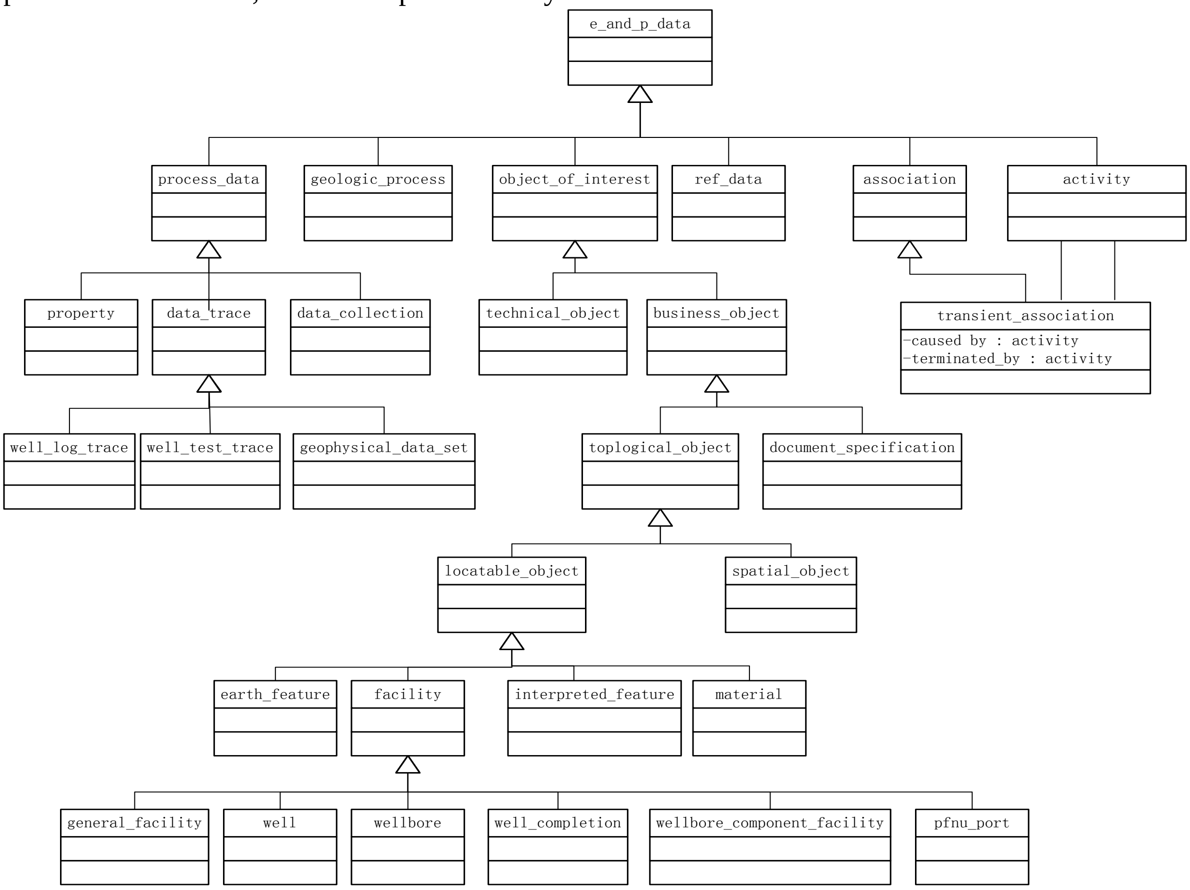


Fig. 2 The top entities in an epicentre data model.

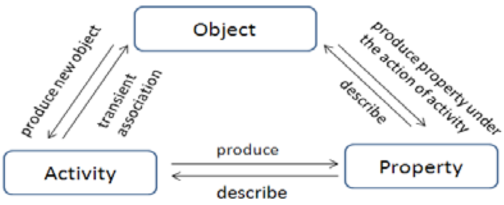


Fig. 3 Relations between Activity, Object and Property.

The epicentre data model is defined in the express language. The structure of entity and enumeration in the express language are shown in Fig. 4 and Fig. 5, respectively. The express grammatical rules must be taken into account in constructing OWL ontology. We provide the translation rules between the epicentre data model and the OWL formal language, on which OWL ontology can be built in a more efficient way than built from scratch.

```
ENTITY pty_porosity
SUBTYPE OF
(
 property
);
 data_value_VALUE        : REAL(6);
 data_value_UNIT         : REF_UNIT_OF_MEASURE;
 distribution            : OPTIONAL line;
 distribution_type       : OPTIONAL
                               ndt_probability_distribution_type;
 maximum_value           : OPTIONAL ndt_real4;
 minimum_value           : OPTIONAL ndt_real4;
 rock_feature            : OPTIONAL rock_feature;
 UNIQUE
 si: rock_ feature,
     activity;
WHERE
 attrib_valid_1: quantity_valid ('porosity', data_value_VALUE,
    data_value_UNIT);
END_ENTITY;
```

Fig. 4 The pty_porosity entity in the epicentre data model.

```
TYPE ndt_probability_distribution_type = ENUMERATION OF
(
 uniform,
 triangular,
 normal,
 lognormal,
);
END_TYPE;
```

Fig. 5 The definition of enumeration in the epicentre data model.

(1). Entities and constraints

Epicentre entities and constraints correspond to OWL class, and propertity restrictions are shown in Table 1.

Table 1 Epicentre entity and OWL component

Epicentre entity	OWL component
Entity Head	Class
Attribute	Data properties, Object properties
Unique	InverseFunctionalProperty
Inverse	AllValuesFrom
Where	OWL property restriction

(2). Data type

Table 2 shows the corresponding data types in the Epicentre data model with respect to the OWL data types. OWL uses XSD (XML schema definition) which is provided by W3C.

Table 2 Data type of Epicentre and XSD

Data Type	Epicentre Data Type	XSD
Number	number integer real	xsd:double xsd:int xsd:double
Text	string	xsd:string
Logic	boolean,logical	xsd:boolean
Data Time	date time,timestamp	xsd:date xsd:time
Binary	binary	xsd:base64Binary
Enu m	Enumeration of	xsd:class,xsd:individua l
set	array,list,set,bag	user-defined data type

(3). Mapping rules

The main elements of a mapping rule between epicentre entity and OWL ontology are the mapping rules of *entity, enumeration, property,* and *activity*. Given E as a collection of Entity, $E=\{e_1, e_2,\dots, e_i\}$, p_i is the property set of e_i, $p_i=\{p_{i1}, p_{i2}, \dots, p_{in}\}$. Given M as a collection of enumeration m, $M=\{m_1, m_2, m_3, \dots, m_i\}$, where m_i contains elements l_i, $l_i=\{l_{i1}, l_{i2}, \dots, l_{in}\}$.

Rule 1: An entity is mapped to an OWL class with the ID name is taken from the entity name.

Notation: $\forall e_i \in E \rightarrow class(id(e_i))$

In the epicentre data model any entity can be converted to an OWL class. The key words SUBTYPE stand for the inheritance relationships among entities. For instance, the OWL formal language for the classes associated with the entity in Fig.3 are:

```
<owl:Class rdf:ID="pty_porosity">
     <rdfs:subClassOf rdf:resource="property"/>
</owl:Class>
```

Rule 2: The name of enumeration is mapped to an OWL class and the elements are mapped to individuals.

Notation: $\forall m_i \in M \rightarrow class(id(m_i))$

$$\forall m_i \in M \wedge l_{ij} \in l_i = 1 \rightarrow individual(id(m_i))$$

For instance, the OWL formal language for the classes associated with the entity in Fig.4 are:

```
<owl:NamedIndividual rdf:ID="uniform">
   <rdf:type rdf:resource="ndt_probability_distribution_type"/>
</owl:NamedIndividual>
```

Rule 3: The properties of the entity are mapped to data properties or object properties. Function dp() is used to distinguish data properties with object properties, and it returns false when the type of property is an entity, otherwise returns true.

$$\forall e_i \in E \wedge p_{ij} \in p_i \wedge dp(p_{ij}) \rightarrow DataTypeProperty(p_{ij})$$

$$\forall e_i \in E \wedge p_{ij} \in p_i \wedge \neg dp(p_{ij}) \rightarrow ObjectProperty(p_{ij})$$

With the translation rules, the epicentre data model can be mapped to OWL ontology. The lithology ontology and the hierarchical structure are shown in Appendix 3 after importing the owl file into ontology editor protégé. The exploration and development ontology provides a set of terms, relations, and constraints that facilitate sharing a common understanding of the structure of information among stakeholders in a domain. They have been used to define the input and output entities for activity nodes in the process model.

3.2 Ontology-based business process modeling

We use ontology to define the data entities in the business process modeling, so the relations between entities can be defined with the benefits of well-defined relations in the domain ontology. Based on the ontology-based business process framework, as shown in Fig. 6, we realized the integration of automatical data acquisition, domain-specific business process, and process-related knowledge.

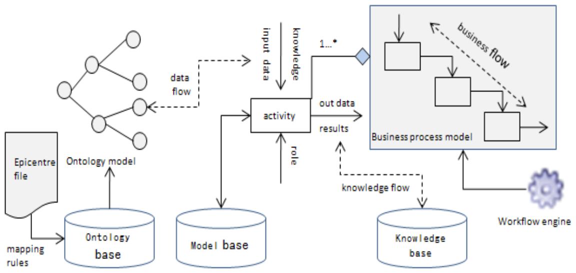


Fig. 6 An ontology based business process framework.

A business process is presented as a collection of activities, which are implemented to accomplish a specific business goal. The processes can be indefinitely refined and recursively decomposed into so-called second processes, third processes and so on. The atomic unit is called an activity, which includes inputs, outputs, roles, and a set of operations. When an activity is initialized, it is necessary to obtain the desired business data from a variety of professional databases. Due to the heterogeneity of data sources and inconsistency of data models, it is difficult to uniformly map data to the activity. In order to achieve the versatility and universality of input and output entities, it is need to establish a standard normalized logical model which can convert heterogeneous data sources into a unified one. Ontology, a well-recognized conceptual model of standardization, is independent with any applications and physical storage implementation. It can deal with the interoperability between heterogeneous data sources by integrating data from different expertise in different fields. The ontology is applied to the process model to achieve the consistency and standardization between the input and output entities. Neither does the ontology depend on the process model nor is affected by modification of process model in its stability. The formalized structure of process and activity are given using BNF, which is shown in Fig.7.

```
<business process> ::= <activity> {, <activity> } | <business process>
<activity> ::= <input> <output> <role> < process logic>
<input> ::= { <data entity> } | {<document entity>}
<output> ::= { <data entity> } | {<document entity>}
<data entity> ::= <ontology>
<document entity> ::= <ontology>
<ontology> ::= <concept-name> <attribute-list> <relations> <individuals>
<process logic> ::= <service> {, <service> }
<service> :: = <service-name> <operation> {, <operation> }
<operation> ::= <operation-name> [{ <parameter> }, <output >]
<parameter> ::= <key> <value>
```

Fig. 7 The BNF of an ontology-based business process framework.

(1). Two categories of inputs and outputs of activity: data entity and document entity. The domain ontology is used to define the input and output entities to guarantee the consistency in the logic models. When it comes to the definition of input and output entities, the ontology could be selected immediately from an ontology base. When a business process is instantiated, data is converted to ontology instances through the mapping rules between RDB and OWL ontology [4] and provides required data to the selected activity.

The same ontology can be utilized by different activities. In order to distinguish ontology instances belonging to different activities, an attribute key ID is added to each ontology definition, recording instance-related activities. The dynamic relations between different entities can be generated automatic by the business activities, as shown in Fig.8, the green diamonds stand for the activities, and the yellow rectangles stand for the entities.

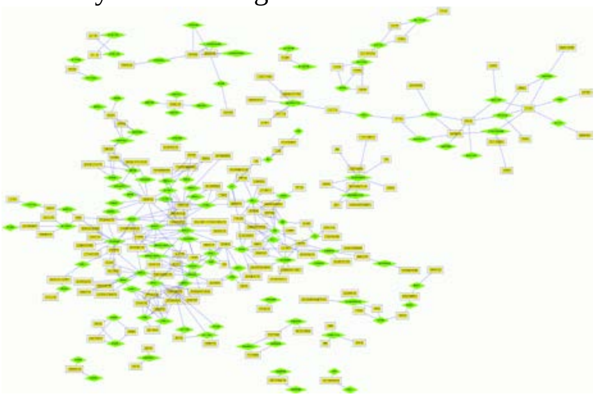


Fig. 8 The entity relationships between different activities.

(2). Processing logic of an activity is consisted with a series of services. We encapsulate some core algorithms or frequently-used charts as services and provide the service names and operation names for calling.

(3). Roles define the classifications of activity operators, for instance, analyst or interpreter. During the instantiation of a business process, roles are assigned to specific people, with the activity shown in the task list of assigned people. When the task is completed, the user uploads the output results to the system and notifies the workflow engine and then moves the flow forward. The role plays an important part in cooperation between departments so that data and knowledge stream can be transferred and shared in different activities.

The instantiation of a business process includes three parts: (1). instantiate the input ontology of each activity (2). pass parameters to related operation sets (3). assign the role to specified business people. Here is an example in lithology recognition of reservoir---an important sub-flow of reservoir evaluation using well logging. Lithology recognition of reservoir mainly contains: rock components analysis, well data processing, core data processing, producing lithology recognition template using cross-plot and data mining techniques. In well data processing, both input and output entities are defined by the well log trace ontology, and the role is well logging interpreter. Operation sets include environmental calibration, standardization, and normalization. When initializing well data processing, the role, well logging interpreter, is assigned to each corresponding people, and input entity is instantiated by the raw well logging curve. The instantiation of an activity called well logging processing in the business process of lithology recognition is shown in Appendix 4.

4 Implementation Details

4.1 Framework

A POKMS is developed based on the proposed process-oriented and ontology-based knowledge model, and is applied in the well deployment filed at the early stage. Currently, this system has been deployed at Shengli oilfield in China.

The architecture of the knowledge service platform is J2EE, through the composite frame mode of Spring and Hibernate, Oracle database, and Flex language to research the system. The development environment is detailed as follows:

- Web Server: Apache Tomcat 6.0
- Programming Language: Java, Flex
- Development Platform: Eclipse 3.5, Adobe Flash builder 4
- RPC Service: BlazeDS
- Database: Oracle 10g
- Library: RDF, Jena
- Graph Visualization Software: Prefuse, BirdEye

We use Flex language to design the graphical process representation tool, and the core classes are ExcutionUnit, Activity, ActivityRect, Opparameter, ProConnect and Port. The class diagram of the business process is shown in Appendix 5. When the design of business process is completed, it is stored in the XML file. Fig. 9 shows a simple process defined in the XML.

```xml
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process expanded="true" height="402" width="710">
        <activity activated="true" class="read_excel" expanded="true" height="60"
              name="read the excel file" width="90" x="45" y="75">
          <parameter key="excel file" value="C:\test.xls"/>
          <parameter key="row offset" value="0"/>
          <parameter key="column offset" value="0"/>
        </activity>
        <activity activated="true" class="set_role" expanded="true" height="76"
             name="number setting" width="90" x="179" y="75">
          <parameter key="name" value="number"/>
        </activity>
        <activity   activated="true"   class="set_role"   expanded="true"   height="76"
               name="class identifier setting" width="90" x="313" y="75">
          <parameter key="name" value="testing result"/>
        </activity>
        <activity activated="true" class="decision_tree" expanded="true" height="76"
               name="decision tree" width="90" x="447" y="75">
          <parameter key="standard" value="gain ratio"/>
          <parameter key="min split size" value="4"/>
          <parameter key="min leaf size" value="2"/>
          <parameter key="min information gain" value="0.1"/>
          <parameter key="max depth" value="20"/>
          <parameter key="confidence" value="0.25"/>
        </activity>
        <connect from_op="read the excel file"    to_op="number setting"   />
        <connect from_op="number setting"    to_op="class identifier setting" />
        <connect from_op="class identifier setting"   to_op="decision_tree" />
</process>
```

Fig. 9 A simple process defined in the XML.

4.2 Modules

The POKMS system consists of four sub-systems: ontology management, data service, process management and knowledge management. The general architecture of POKMS is shown in Appendix 6.

Ontology management consists of ontology extraction, ontology maintenance, and ontology visualization. The ontology extraction module aims at processing the epicentre data model file and generating the OWL ontology of exploration and development. Based on the extensible framework prefuse, which is a toolkit for interactive information visualization, the system provides tree view and graph diagrams for ontology visualization, as shown in Appendix 7. When clicking one ontology tree node, all instances, activities, processes or resources related to the selected ontology are pushed in a new window.

The data service sub-system has one module named business subject management. This module creates the association between business subject and domain ontology, and immediately pushes the related data to business experts to avoid the tedious work for collecting and organizing data manually.

Process management sub-system consists of process representation tool, process model management, workflow engine, and process instance management. The ontology base is the foundation and provides the standard input or output entities for activities. Process representation tool provides a friendly visible interface for process modeling. A business process graph can be drawn by users through various graphics control in this system. Workflow engine is used to control the execution order through the traverse algorithm of business process graph. Some core algorithms and frequently-used charts are encapsulated as services, which are maintained in the process model management module. The process model management module lists all the services for different activities to call. Appendix 8 shows the reservoir evaluation business process and the related process knowledge.

Knowledge management sub-system contains knowledge query and knowledge visualization. Process-oriented ontology-based knowledge model helps us realize the integrated management among data acquisition, domain-specific business process, and process knowledge, ontology knowledge and process knowledge are generated during the execution of business processes.

5 Conclusions and Future Work

In this paper, we first automatically build the exploration and development ontology based on the mapping rules between OWL ontology and epicentre data model, which is a global petroleum industry data model. The exploration and development ontology can provide a common data interpretation for all the data in different sources, and promote the data exchange and data sharing between the collaborating applications. In the business process modeling, we use ontology to define the process-related entities, so that the relations between entities can be expressed. With the benefits of well-defined relations in the domain ontology, a process-oriented knowledge net is formed. In addition, we build a web-based knowledge service platform with a graphical process representation tools for users to visually design the business processes. Enabled by this system, decision makers could make progress on each activity in time, transfer, and share results produced by each activity among departments, experts, and non-experts. Furthermore, end users, especially non-experts, can reuse the domain knowledge, and monitor the processes of the ongoing projects, playing an important guidance role to the new exploration and development. Currently, the ontology describes the knowledge framework in the field of exploration and development with formal semantics. Therefore, reasoning mechanism should be researched to infer new knowledge from the existing information in the ontology. In the future, analyzing the business processes and identifying possible deadlocks or bottlenecks to optimize business operation processes will be considered further.

6 Acknowledgement

This research is supported by National Major Scientific and Technological Special Project of China during the Twelfth Five-year Plan Period under grant 2011ZX05020-009 and the Step Plan of Beijing Municipal Science and Technology Commission under grant Z121109009212008.

References:

[1] Becerra-Fernandez I, Sabherwal R. Organizational knowledge management: A contingency perspective[J]. *Journal of management information systems*, 2001, 18(1): 23-56.

[2] Alan H. Smith. Putting a value on data management[J], *E&P*, 2002(9):69-71.

[3] U. Visser, H. Stuckenschmidt, G. Schuster, T. Vögele.Ontologies for geographic information processing[J]. *Computers & Geosciences*, 2002,28(1): 103-117.

[4] Santoso H A, Haw S C, Abdul-Mehdi Z T. Ontology extraction from relational database: Concept hierarchy as background knowledge[J]. *Knowledge-Based Systems*, 2011, 24(3): 457-464.

[5] Jia CX, Hu W, Bai WY. SMap:Semantically mapping relational database schemas to OWL ontologies. *Journal of Computer Research and Development*, 2012,49(10):2241-2250.

[6] K. Vergidis, A. Tiwari, B. Majeed, Business Process Analysis and Optimization:Beyond Reengineering, *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews* 38 (1) (2008) 69–81.

[7] Jung J, Choi I, Song M. An integration architecture for knowledge management systems and business process management systems[J]. *Computers in Industry*, 2007, 58(1): 21-34.

[8] Zhang XG, Li MS. Workflow-Based knowledge flow modeling and control. *Journal of Software*, 2005,16(2):184−193.

[9] Han K H, Park J W. Process-centered knowledge model and enterprise ontology for the development of knowledge management system[J]. *Expert Systems with Applications*, 2009, 36(4): 7441-7447.

[10] Ioannis Savvas, Nick Bassiliades. A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration. *Expert Systems with Applications*, 2009, 36(3): 4467-4478.

[11] The POSC Epicentre Data Model[EB/OL]. http://w3.energistics.org/archive/Epicentre/Epicentre_v3.0/index.html.

[12] The Energy Standards Resource Center[EB/OL]. http://www.energistics.org/.

[13] LI Xiaolin ,WU Xiying ,YI Hong. A methodlogy for the development of the EXPRESS data schema. *Computer Integrated Manufacturing Systems*,1998 ,4(5) :32 – 35.

[14] J.Leon Zhao. Schema coordination in federated database management: a comparison with schema integration[J].*Decision Support Systems*,1997,20(3): 243-257.

[15] Rafi Ahmed, Philippe De Smedt, Weimin Du, et al. The Pegasus Heterogeneous Multidatabase system. *IEEE Computer*[J], 1991, 24(12):19-27.

[16] Daniel L. Rubin, Terry S. Desser. A Data Warehouse for Integrating Radiologic and Pathologic Data[J]. *Journal of the American College of Radiology*, 2008,5(3):210-217.

[17] Sanjay Madria, Kalpdrum Passi, Sourav Bhowmick. An XML Schema integration and query mechanism system[J].*Data & Knowledge Engineering*,2008,65(2):266-303.

[18] Kai-Chih Liang, Daphne Chyan, etc. Integration of CORBA and object relational databases[J].*Computer Standards & Interfaces*, 2003,25(4):373-389.

[19] Byung Chul Kim, Hans Teijgeler, Duhwan Mun, Soonhung Han.Integration of distributed plant lifecycle data using ISO 15926 and Web services[J].*Annals of Nuclear Energy*, 2011, 38(11): 2309-2318.

[20] Rajpathak, Dnyanesh, Chougule, Rahul. A generic ontology development framework for data integration and decision support in a distributed environment[J]. *International Journal of Computer Integrated Manufacturing*, 2011,24(2):154-170.

[21] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition*, 1993, 5 (2): 199–221.

[22] F. Cerbah, Mining the Content of Relational Database to Learn Ontology with Deeper Taxonomies, *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2008.

[23] Li S L, Liao L J, Cao Y D, Cao S G. Ontology Learning for Chinese Documents Based on SVD and Concept Clustering. *Journal of Beijing Institute of Technology*, 2003, 12 (suppl).

[24] Cimiano P, Hotho A, Staab S. Comparing Conceptual, Divisive and Agglomerative Clustering for Learning Taxonomies from Text[C]//Ecai 2004: *Proceedings of the 16th European Conference on Artificial Intelligence. IOS Press*, 2004, 110: 435.

[25] Medelyan O, Milne D, Legg C, et al. Mining meaning from Wikipedia[J]. *International Journal of Human-Computer Studies*, 2009, 67(9): 716-754.

[26] Ferrández S, Toral A, Ferrández O, Ferrández A, Munoz R. Exploiting Wikipedia and EuroWordNet to solve cross-lingual question answering. *Information Sciences* , 2009;179(20):3473e88.

[27] Küçük D, Arslan Y. Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles[J]. *Renewable Energy*, 2014, 62: 484-489.

[28] Suchanek F M, Kasneci G, Weikum G. Yago: A large ontology from wikipedia and wordnet[J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2008, 6(3): 203-217.

[29] Boehm B W. Software engineering economics[J]. *Software Engineering, IEEE Transactions on*, 1984 (1): 4-21.

[30] Scheer A W, Nüttgens M. ARIS architecture and reference models for business process management[M]. *Springer Berlin Heidelberg*, 2000,pp.376-389.

[31] Sharp A, McDermott P. Workflow modeling: tools for process improvement and applications development[M]. Artech House, 2008,pp.24-26.

[32] Maurer F, Holz H. Process-oriented knowledge management for learning software organizations[C]//Proceedings of 12th Knowledge Acquisition For Knowledge-Based Systems Workshop. 1999.

[33] Maurer F, Goldmann S, Holz H, et al. Merging project planning and web-enabled dynamic workflow technologies[J]. *IEEE Internet Computing*, 2000, 4(3): 65-74.

[34] Han K H, Kang J G. A process-based performance measurement framework for continuous process improvement[J]. *INTERNATIONAL JOURNAL OF INDUSTRIAL ENGINEERING-THEORY APPLICATIONS AND PRACTICE*, 2007, 14(3): 220-228.
[36] Ri Hai, Manfred Theißen, Wolfgang Marquardt. An ontology based approach for operational process modeling. *Advanced Engineering Informatics*, 2011, 25: 748-759.
[37] Daiyi L, Li K, etc. An ontology-based knowledge representation and implement method for crop cultivation standard. *Mathematical and Computer Modelling*, 2013, 58: 466-473
[38] B.L. Wen, L. Zhang. Method of building petroleum exploration and development domain ontology. *Computer Engineering and Applications*, 2009, 45(34) : 1-3.

Appendix 1: The comparison of data exchange technologies

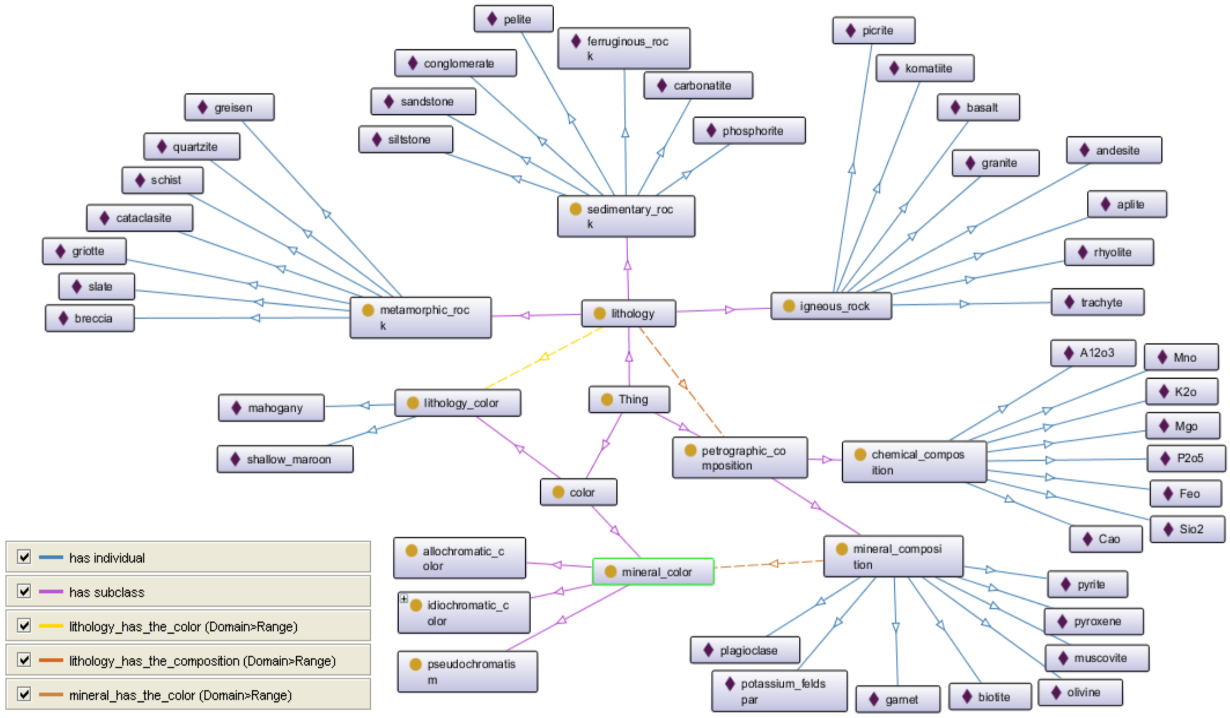
Technologies	Approaches	Limitations
Federated database system	Applications communicate with federated servers using any supported interface, e.g., jdbc, odbc, and federated server communicates with data sources by the wrapper.	(1)Large amount of interfaces and poor extensibility; (2)The task of creating mapping rules becomes tremendous when many databases are added in the federation. It needs to create n (n-1) mapping rules when n is very big.
ETL and data warehouse	It regularly extracts data from various data sources and loads the processed data into new data sources to guarantee the consistence of the data in different data sources. This method is appropriate for small-scale data and infrequently changed data sources.	(1)The ETL process will become very complex when facing big data and frequently changed data sources. (2)Because of its regular update, data warehouse has weak real-time performance. (3)It does not support the basic operations of addition, update and deletion. (4)This technology is mainly applied for data query and data analysis.
CORBA (Common Object Request Broker Architecture)	It is platform-independence and language-independence and provides the infrastructure to solve the heterogeneity of different data bases. The data model and query language OQL in the object-oriented database standard ODMG93 partly support the semantic heterogeneity.	(1)It is large and complex, and the update of technologies and standards is relatively slow.
Web service	It has advantages of wide universality, strong processing capacity for result information and perfect identifying functionality.	(1)The database manufactures need to invest more at technical level, and they will take a long time to realize it.
XML	It is the data exchange standard with the benefits of self-describing, expansibility and platform-independence and.	(1)The mapping rules of xml schema are very complex. Meanwhile, the realization of semantic mapping is not easy.
UML	UML is a combination of different model types, e.g., class diagrams or sequence diagrams. Object constraint language (OCL) can express some rules, but OCL has no formal semantics.	(1)As there is no reasoning, the user has to define the whole classification tree himself. (2)Additionally, no dedicated query language exists for UML.

Appendix 2: The approaches and limitations about the combination of business process and knowledge management

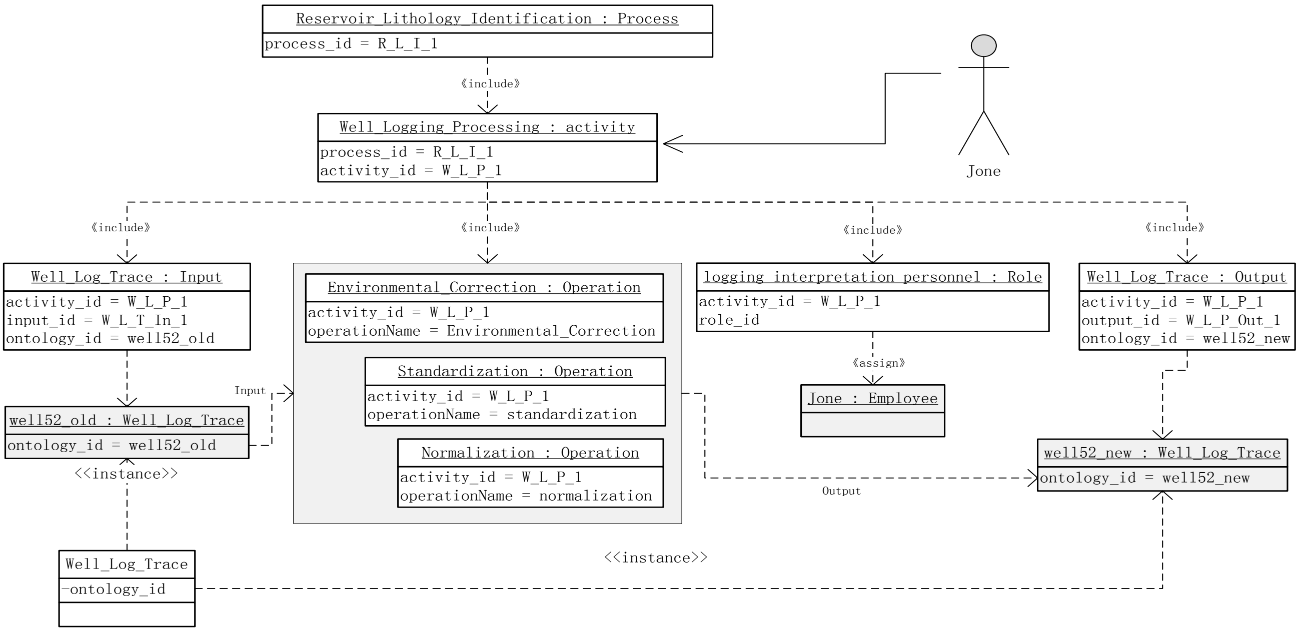
Article	Issues	Approaches	Limitations
An integration architecture for knowledge management systems and business process management systems[7]	In existing KMSs, knowledge is managed statically without explaining when it is used by whom.	propose an architecture for integrating (KMSs) and (BPMSs) to combine the advantages of the two paradigms; Define the concept of process knowledge and classify it into three types: process template knowledge, process instance knowledge; process related knowledge.	(1)Meta information was used to describe basic concepts or attributes of process template knowledge, however, relationships between concepts were not shown. (2) Only develop a prototype system to realize integration architecture, but it was not applied in any domain.
A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration domains[10]	Due to the great overhead of legal norms that exist in Greece, there is a true need for help to civil servants who use legislation, in order to maintain an updated version.	Adopt a process oriented approach through a web-based knowledge management system that provides this legal framework in an up-to-date and accurate manner. The system employs ontology in OWL for representing the public administration structure and any kind of documents.	(1)The ontology refers to administrative entities, procedures and documents rather than legal hierarchies. (2) Procedural aspects of the ontology are represented in a Semantic Web Service framework, namely OWL-S. (3)Public administration's structure is specified by laws and it is not a result of the consensus of professionals. The ontology and procedure are predefined in OWL and OWL-S. Hence, this system does not support a user to customize the process graphically.
Process-centered knowledge model and enterprise ontology for the development of knowledge management system[9]	Knowledge is separated from the business process context, and it does not lead to the ability to take the right action for target performance.	Propose the framework for process-centered knowledge model and enterprise ontology. The enterprise knowledge object in this model is classified into two types: process knowledge and task support knowledge. Process knowledge includes generic process and project instance.	(1)Enterprise ontology was built manually by using the generic concept template in the form of the UML. (2)The business process was created by the generic process template in the form of UML. The template was predefined, and it does not support designer to customize the process graphically. Moreover, the business process cannot be further cut down to smaller ones. (3)The task roles were not mentioned in the activity definition. (4)A process-centered KMS was developed to show the applicability of proposed framework. However, it was

			not equipped with the workflow engine function for control of the project progress.
Workflow-Based Knowledge Flow Modeling and Control[8]	(1) Knowledge management mechanisms could not be represented by current workflow process definition meta models. (2) Promote the knowledge sharing and cooperation between humans.	(1)Propose an innovative extended workflow process definition meta model to integrate workflow and KM. Knowledge requirement and knowledge engine were appended. (2)A knowledge flow modeling approach is proposed by using five kinds of knowledge flow components to represent knowledge distribution and reuse, cooperation and communication among participants.	(1)Focus on the theory of knowledge flow modeling and controlling. However, it is not suitable for business process management. (2)In the knowledge flow model, the meta information was not mentioned.

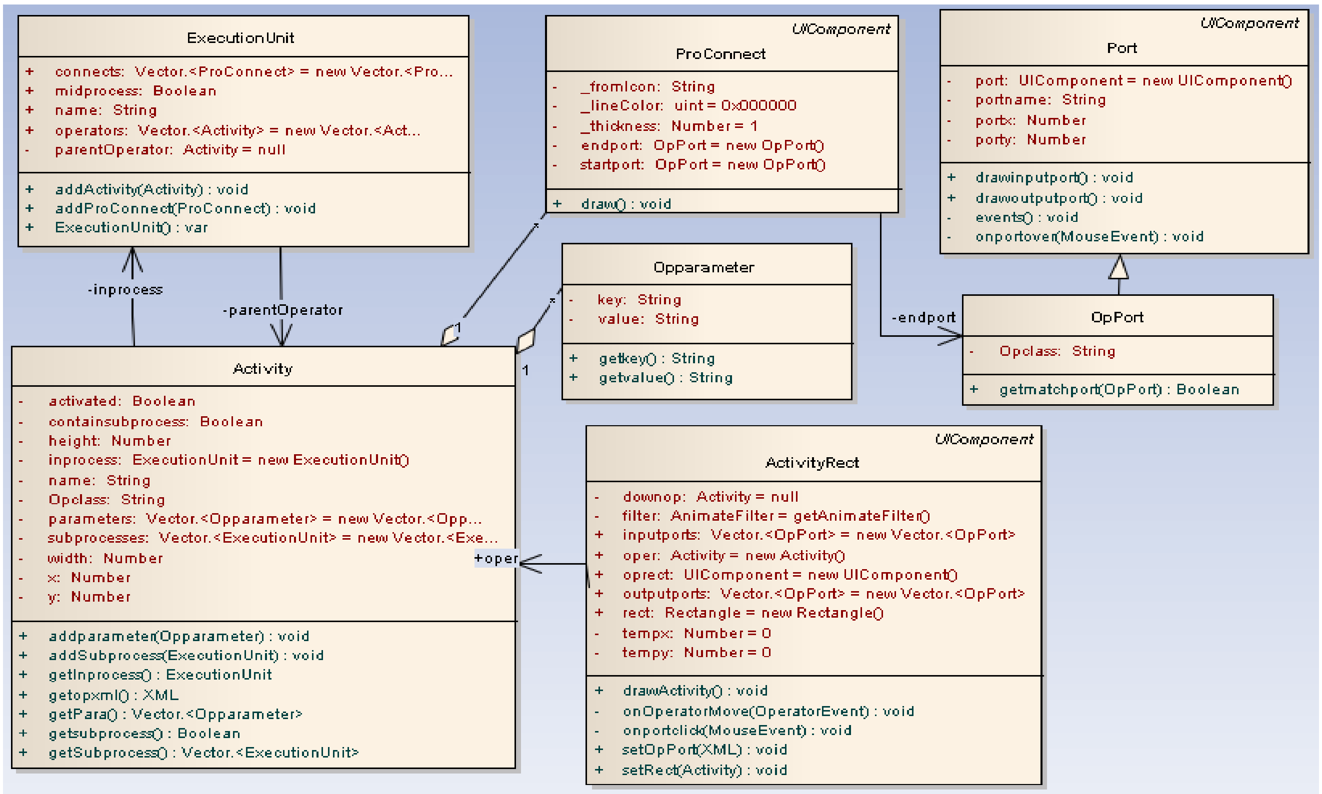
Appendix 3: The lithology ontology in protégé



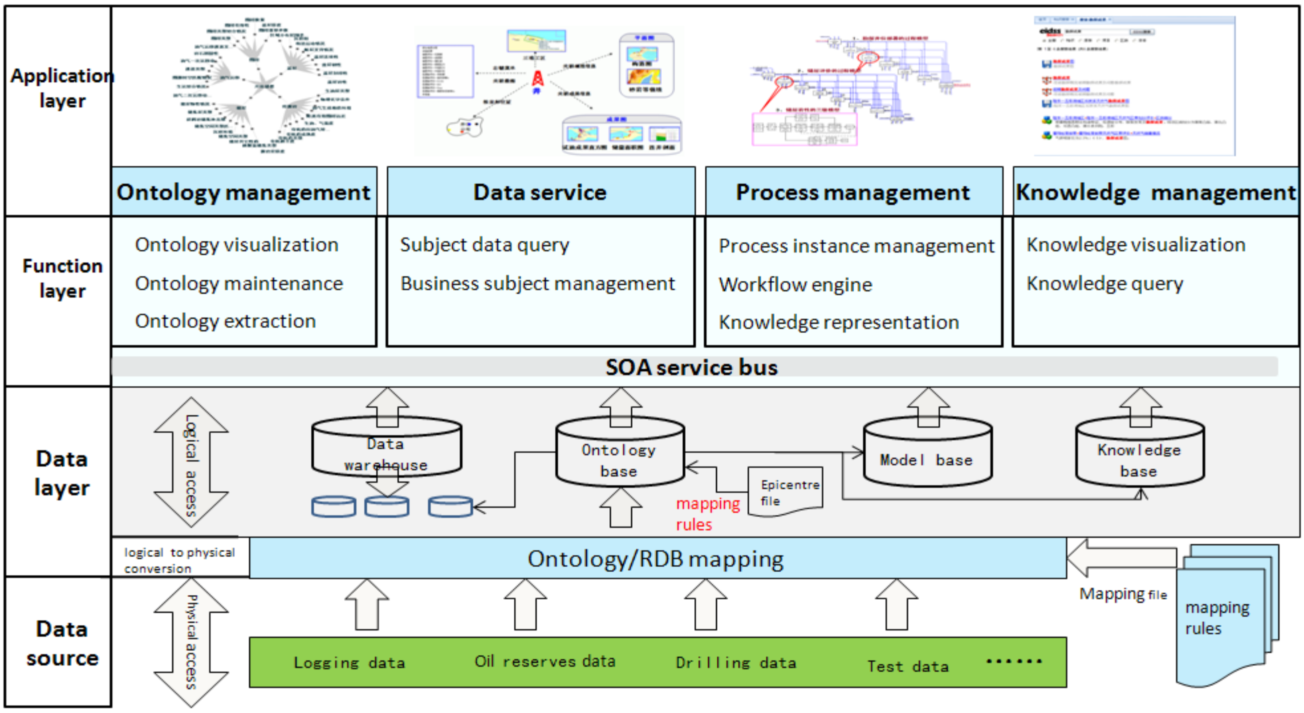
Appendix 4: The instantiation of well logging processing in the business process of lithology recognition



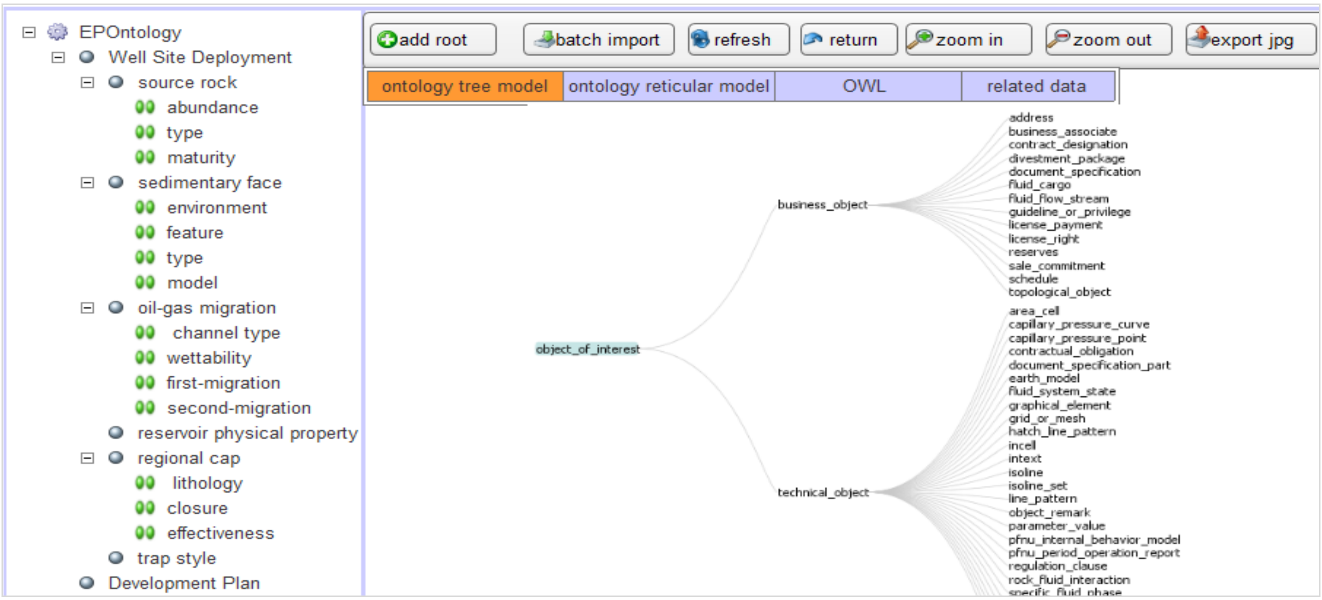
Appendix 5: The core class diagram of process in Flex



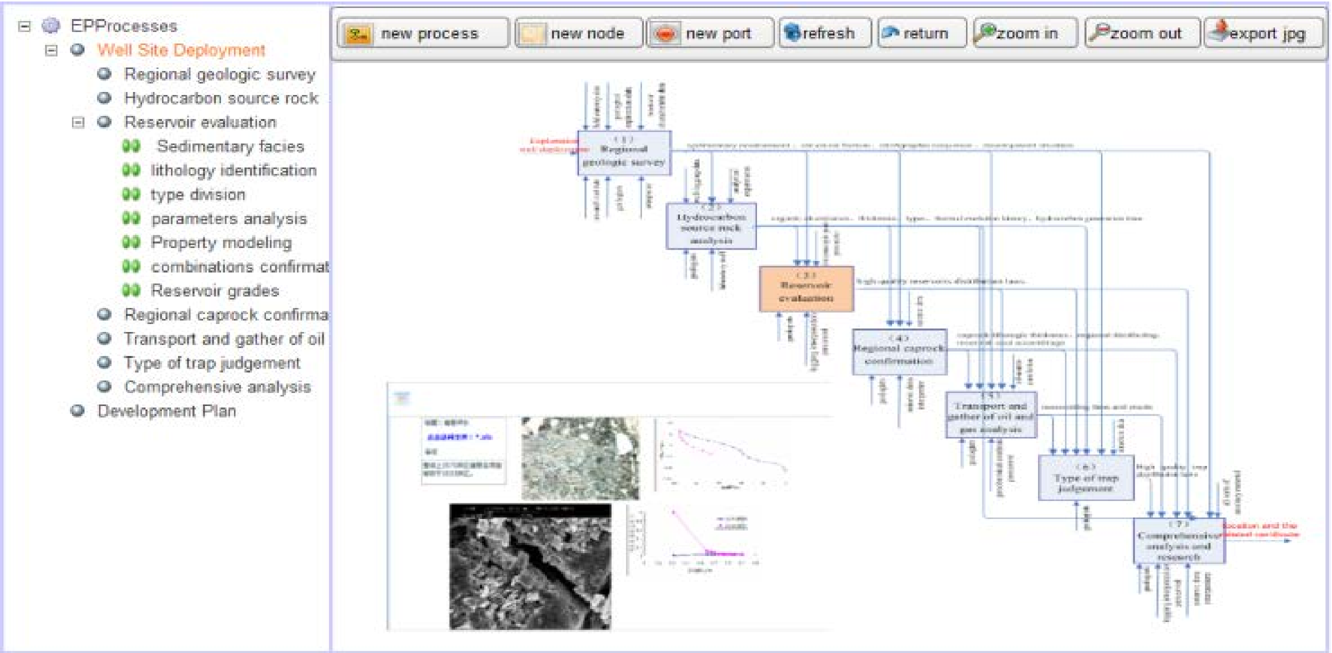
Appendix 6: The general architecture of POKMS



Appendix7: The ontology management sketch of POKMS



Appendix8: The reservoir evaluation business process in POKMS


EPProcesses
Well Site Deployment
Regional geologic survey
Hydrocarbon source rock
Reservoir evaluation
Sedimentary facies
lithology identification
type division
parameters analysis
Property modeling
combinations confirmat
Reservoir grades
Regional caprock confirma
Transport and gather of oil
Type of trap judgement
Comprehensive analysis
Development Plan
new process
new node
new port
refresh
return
zoom in
zoom out
export jpg