

## Корпоративные репозитории для управления знаниями. Corporate repositories for knowledge management

Tarun Chandrasekhar,\* Neuralog рассматривают, что значит управление знаниями для нефтегазовой индустрии и зачем оно нужно, а так же обсуждают вопросы проектирования

Управление знаниями - последнее слово в индустрии разведки и добычи- оно свободно используется для определения любой деятельности от архивирования данных до управления информацией для разработки продуманного подхода к стратегии бизнеса.

Согласно Russell Ackoff<sup>1</sup>, специалисту в области теории систем и профессору в изменении организационной структуры (Ackoff, 1989), содержание человеческого разума может быть разделено на следующие категории:

**Данные:** символы, набор значений наблюдаемых, измеряемых или рассчитываемых признаков. Данные - сырье. Они просто существуют и не имеют никакого значения вне своего существования. Они могут существовать в любой форме, как годной, так и не годной к использованию. Сами по себе они ничего не значат. Примеры включают пункты взрыва от сейсморазведки или каротажные кривые (как кривые на бумаге, так и цифровые LIS/DLIS файлы).

**Информация:** данные, которые обработаны, чтобы быть полезными; обеспечивают ответы на вопросы типа 'кто', 'что', 'где', и 'когда'. Когда данные организованы, обработаны, и структурированы, они становятся информацией. Хорошим примером будет разделение сырых скважинных данных по типу на электрические, акустические, радиоактивные и электромагнитные. Когда бумажные кривые оцифрованы и сгруппированы, геолог может проанализировать эту информацию и извлечь знание. Эти кривые помогают ответить на вопрос "где" о формации, ее пористости и проницаемости.

**Знание:** применение данных и информации; ответы 'как' вопросы. Знание - соответствующий набор информации, сгруппированной чтобы быть полезной. Это - подмножество информации, которая была извлечена, отфильтрована, подвержена проверке и прошедшая проверку

Когда геолог коррелирует различные каротажные кривые, создает разрезы, и анализирует проницаемость и пористость структур, он или она создает знание из информации. Ключевой оператор, который преобразовывает информацию в знание, является сводом правил, которые утверждают эту информацию. Обычно, мы именуем их как деловые правила - руководящие принципы, которые определяют способы ведения бизнеса. Каждая организация, не зависимо от типа самоопределения, следует набору бизнес-правил,

которые помогают достигать цели. В последнее десятилетие процесс управления бизнесом стал наиболее важным аспектом управления знаниями. Правила верификации, помогающие создавать знание, так же важны, как и информация, преобразованная в знание.

### Понимание и мудрость: взвешенный анализ 'почему'

Понимание - когнитивный, аналитический и интерполяционный процесс, используя который можно создать новое знание из существующего. Понимание- это то, что геолог использует для корреляции знаний, полученных из разных источников для предсказания местоположения структуры или будущего положения скважины. Точно так же делается точная геологическая интерпретация, когда стратиграфия от буровых колонок коррелируется со временем пробега сейсмической волны.



Преобразование данных в информацию, знание и мудрость.

\* E-mail: tarun@neuralog.com.

Мудрость, с другой стороны, является экстраполяционным и недетерминированным братом понимания, который является критерием, с помощью которого мы различаем или судим о правильном и неправильном, хорошем и плохом. В нефтегазовой промышленности много сухих скважин были приписаны мудрости или нехватке ее! Было бы желательно для системы управления знаниями включить понимание или правила и процессы, которые управляют пониманием. Это выполнимая, но очень трудная задача. С искусственным интеллектом, робототехникой, и нейронными сетями, разработка такой сложной системы может быть в нашем ближайшем будущем.

Мудрость с другой стороны, уникальное человеческое состояние. Не предполагается, что оно в любой момент сделает скачок до компьютера. Организационная мудрость - суммирование мудрости личностей организации. Управление этой мудростью является сложным вопросом, как для индивидуума, так и для всей группы.

### Системы управления знаниями

Система управления знаниями для нефтяной индустрии (E&P) (в общем, основанные на информационных технологиях IT)-это система для создания, получения, хранения, управления и распространения данных, информации и знаний в пределах организации.

*Управление данными/информацией*  
Большинство организаций использует одну и более из следующих данных и/или информационных управляющих систем:

*Физические библиотеки документов* - Они часто используются, чтобы сохранить бумажные каротажные кривые, отчеты по скважинам, карты, аэрофотоснимки, и другую документацию с информацией о скважинах, полях, бассейнах и регионах. Общие недостатки таких систем включают трудность доступа, совместного использования и время предварительной обработки, требуемое для подготовки данных к анализу.

*Общедоступные сетевые диски* - Сканированные документы и отчеты так же как структурированная информация, которая была обработана, часто хранятся на диске на общедоступном сетевом хранилище. В то время как эти диски обеспечивают доступ различным пользователям к информации в цифровой форме, они испытывают недостаток в безопасности и надежности, предлагаемыми стандартными для отрасли базами данных.

*Прикладные базы данных* - Наиболее коммерчески доступные системы интерпретации используют базы данных, чтобы хранить данные бурения, добычи и окончания скважины. У этих систем может быть дополнительный компонент управления записями. Однако эти базы данных, прежде всего,

соответствуют потребностям приложения (которое является, а может и не является главное потребностью управления знаниями для организации). Функциональная совместимость данных с другими приложениями ограничена.

*Корпоративные репозитории* - Корпоративный репозиторий, обычно RDBMS (Relational database management system), хранит исходные данные или данные прошлых лет. Его главной целью чаще всего является накопление и хранение в безопасности ценных для компании данных. Репозитории все чаще и чаще помещаются в основе большинства систем управления данными, информационного управления, и систем управления знаниями. Типичные задачи включают функциональную совместимость и совместное использование данных с различными приложениями и системами внутри организации.

### Интегрирование управления бизнес-процессом

Нефтяные компании обозначили потребность в управлении их бизнес-процессами, чтобы продвинуть эффективность и прибыльность и теперь использовать как людей, так и технологии для изучения и улучшения этих процессов. Однако, корпоративные репозитории редко используются, чтобы управлять ими. Интегрированные корпоративные репозитории, которые хранят данные, информацию, и деловые правила, еще более редки. Но наличие таких репозиториях, которые хранят деловую документацию, обеспечивает возможность хранения знаний.

Знание здравого смысла - информация, которая была утверждена здравым опытом. Научное знание - информация (гипотезы и теории), утвержденная правилами и проведенными научным сообществом тестами. Организационное знание в этих терминах - информация, утвержденная по правилам и тестам организации, ищущей знание. Качество этого знания будет в значительной степени зависеть от способности тестов и правил верификации производить знание, которое улучшает организационную работу (организационная версия объективного знания). (Das, S. [2007] Различие между Данными, Информацией, Знанием и Мудростью, <http://knowmgt.blogspot.com/2007/07/difference-between-data-information.html>).

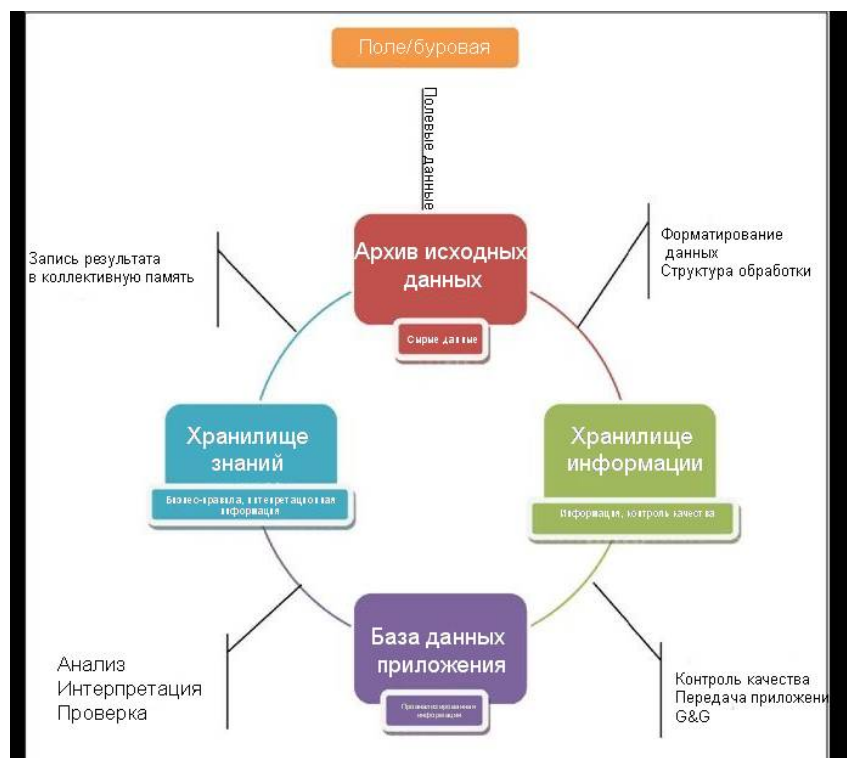
Одним из преимуществ проектирования систем управления знаниями, используя корпоративные репозитории является защита и сохранность данных, предоставляемый доступ пользователям, функциональная совместимость с другими системами.

Одним из первичных принципов планирования аварий и управления риском является построение защищенной от краха системы, которая защитит второй самый ценный актив компании - знание (первым являются люди). Хорошо спроектированные корпоративные репозитории имеют обширные возможности восстановления в аварийных ситуациях и полагаются на технологию базы данных (резервные копии, внешнее дублирование) а так же на логическую технологию (осуществляющую

администрирование и безопасность и предотвращающую случайное удаление или модификацию информации), чтобы удостовериться, что сохраненное знание безопасно. Репозиторий знания позволяет организации делать доступным все данные, информацию и деловые правила, которые были ранее рассеяны среди различных приложений, людей и систем. Пользователи, которым нужен доступ к определенной информации, теперь могут получить ее и все связанные с ней данные. Хорошо спроектированные системы обеспечивают доступ различным пользователям, приложениям, и базам данных в пределах организации, разрешая менеджерам данных управлять этой информацией с одной центральной точки. Пользователи могут использовать различные клиентские технологии - средства просмотра, сетевые порталы, ГИС- приложения, и различные продукты связующего ПО, чтобы непосредственно работать с информацией, хранившейся в системе.

#### Цикл жизни знания.

Корпоративные репозитории также включают различные модули, чтобы импортировать данные различных типов из различных



источников и обеспечить способность непосредственно доставить их приложению G&G. Так как такие репозитории не являются собственностью определенного разработчика, они имеют тенденцию быть основанными на стандартах открытого кода и часто совместимы со многими приложениями сторонних производителей

Такие организации как PPDM ([www.ppdm.org](http://www.ppdm.org)) и Energistics (<http://www.energistics.org/posc/Default.asp>) разработали доказанные модели данных и стандарты обмена данными, которые обеспечивают обоснование для озвученной реализации корпоративного репозитория.

Окончательная цель системы управления знаниями состоит в том, чтобы стать 'коллективной памятью' организации.

Вся информация, снабженная необработанными данными, обработкой, и деловыми правилами, используемыми для проверки правильности, сделана доступной, чтобы помочь принять лучшее решение. Применение деловых правил, которые делают корпоративный репозиторий центральным и единственным

источником информации осуществляет обмен информацией между людьми, отделами и системами.

#### Цикл жизни знания

Знания начинаются как необработанные данные. Сейсмические площади отстрелены, скважины пробурены, проведены скважинные исследования и тесты продуктивности, и собрано большое количество сырых данных. Как только эти данные получены, они сохраняются в хранилище исходных данных. Традиционно, когда полученные данные были на бумаге, они архивировались в комнатах, полных документов, таких как карты, сейсмические разрезы, каротажные кривые и отчеты. Сегодня, они продолжают существовать, в дополнение к цифровым хранилищам данных, все из которых вместе являются оригинальным хранилищем данных. Использование хранилищ, поддерживающих как растровые, так и векторные данные, позволяет организации объединить эти хранилища данных.

Жизненно важно не изменить данные на данном этапе. Оригинальное хранилище данных, созданное ли в физической или цифровой форме, чтобы быть точным и законченным представлением необработанных данных, которые получены в поле или из других источников. Важно сохранить эти данные и их необработанную природу так, чтобы пользователи могли вернуться назад в случае неверного анализа или повреждения данных.

Сырые данные из хранилища обрабатываются, форматируются, структурируются, и подготавливаются для анализа. В процессе этого они становятся информацией. Один из решающих процессов, которые должны пройти необработанные данные, является контроль качества (QC). У всех организаций есть QC, через который проходят необработанные данные. В некоторых компаниях эти процессы хорошо определены. Часто опыт оператора данных служит контролем качества. Качество хранящейся информации важно. Соответственно все больше организаций берут на вооружение процесс QC.

Последняя тенденция в информационном управлении сосредотачивается на том, чтобы сохранять эти процессы QC и проекты совместно с обработанными данными. Это служит ценной основой для пользователей в организации, чтобы создавать и выполнять контроль качества данных как деловое правило. Все большее число информационных управляющих систем сегодня имеют встроенные модули тестирования и управления контролем качества.

Информация готова к анализу, как только проведен контроль качества. Анализ осуществляется в прикладных системах G&G, таких как Petrel от Schlumberger, Kingdom от SMT, Geographix от Landmark, или платформе OpenWorks. Есть много прикладных систем, которые служат нефтегазовой

промышленности, и у большинства организаций есть, по крайней мере, одна из этих систем.

Во многих нефтяных и газовых базах данных для интерпретационных систем используются как системы управления данными. У этих баз данных есть несколько преимуществ - это главным образом зрелые системы с устойчивыми механизмами поставки к связанным с ними приложениям G&G. Они не требуют никаких дополнительных IT ресурсов (кроме IT ресурсов, ответственных за поддержание приложения G&G). Однако они не являются корпоративными репозиториями. Главной целью данных, хранящихся в прикладных базах данных, является анализ. Эти базы данных были спроектированы для реализации этой функции.

Дополнительно, у этих баз данных есть ограничение,- они не могут взаимодействовать с приложениями других разработчиков. Вообще трудно ожидать, что база данных одного разработчика будет в состоянии хранить данные, используемые в прикладной системе другого разработчика. Корпоративный репозиторий, типично, - независимое приложение и позволяет сохранять данные из различных источников полученными от различных приложений. Дополнительно,

Volant Solutions обеспечивают совместимость данных между корпоративными репозиториями и стандартными для промышленности прикладными базами данных.

Результат анализа, выполненного любой прикладной системой, интерпретируется квалифицированными специалистами, такими как геофизики, геологи или петрофизики, и заключительные результаты - карты, разрезы, оценки и отчеты - используются для обоснованного решения. Заключительная часть задачи состоит в хранении результатов исследований, конечных результатов, принятых решений так же как бизнес-процессов и правил, которые управляют этим процессом. Таким образом, знание не будет потеряно в случае увольнения человека в другую компанию. Оно будет сохранено, организовано и использовано совместно с начальными данными.

#### **Физическая разработка корпоративного репозитория**

Физический проект системы управления знанием должен позволить плавный поток данных и информации между ее различными компонентами. Реляционные базы данных лучше всего удовлетворяют потребностям, описанным выше. Кроме того, как упомянуто прежде, модель данных репозитория должна быть

В статье для Dow Jones Newswires Angel Gonzalez написал, что нефтяные и газовые компании, планирующие увеличить отдачу, приближаются к стене: 'через меньше чем десятилетие почти половина стареющей рабочей силы, занятой в разведке и добыче, будет уволена.' Таким образом их честолюбивые проекты могут опоздать или оказаться очень дорогими, 'только если не наймут и обучат достаточно людей для укомплектования.' В этой быстро стареющей промышленности важно записать и сохранить бизнес-процессы и правила, которые могут быть переданы недавно нанятым сотрудникам. Это - хороший способ сохранить интеллектуальную собственность и управлять 'коллективным разумом' организации.

промежуточные программные средства, такие как предлагаемые компаниями как OpenSpirit и



основана на открытых стандартах и поддерживать открытые стандарты обмена данными.

### Интегрированная модель

В интегрированной модели все компоненты репозитория - архив исходных данных, информационный банк, данные приложения или проекта и знание - все сохранены в одной централизованной базе данных. Это облегчает управление всей информацией с одного места. Однако если все компоненты постоянно находятся физически в одном месте базы данных, то там же должны быть деловые правила, которые сохраняют данные различных компонентов изолированными друг от друга. Например, если информация была изменена в прикладной системе, это не должно затронуть архив оригинальных данных или информационный банк. Другое ограничение для этой модели - то, что прикладные системы имеют тенденцию иметь их собственные базы данных и нет никакого смысла в дублировании этого компонента в интегрированной модели.

### Распределенная модель

С другой стороны, распределенная, денормализованная модель могла бы использоваться, когда у каждого из этих четырех компонентов есть своя собственная база данных, а так же база данных и деловые правила, которые синхронизируют и обновляют все базы данных, в случае изменения одной из них. Логически, это позволяет осуществить чистую реализацию с небольшим шансом поперечного вмешательства. Однако в этом случае для управления одним корпоративным репозиторием потребуется администрирование различных баз данных.

#### Гибридная модель

Гибридная модель использует

одну базу данных для управления архивом исходных данных, информационным банком и базой знаний. Данные приложения сохраняются в пределах прикладной системы, а различные программы ввода-вывода подключают и синхронизируют эти две базы данных. Имеет место некоторое дублирование, когда исходные данные или информация из корпоративного репозитория копируется на прикладную систему для анализа, и конечные результаты копируются назад в репозиторий. Критики утверждают, что это вызывает дублирование и расточительное использование дискового пространства. Однако, логически, этот проект действительно имеет смысл. Первичная цель корпоративного репозитория - защита исходных данных и информационных архивов. Напротив, данные, проанализированные в прикладной системе, изменчивы - они подвергаются многим изменениям, и конечный результат достигается после многочисленных итераций. Как только конечные результаты переданы обратно в репозиторий, данные из прикладной системы могут быть удалены. Гибридная модель сохраняет неприкосновенность данных, хранящихся в репозитории и знаний, полученных в прикладной системе за счет дискового пространства.

### Рассмотрение моделей

Преимущества систем управления знаниями довольно очевидны, однако есть многочисленные факторы, которые нужно рассмотреть, выбирая способ реализации системы управления знаниями. Помимо обычных соображений для любой RDBMS, есть несколько ключевых пунктов, которые нужно принять во внимание.

*Существующие системы управления данными* "Мы не нуждаемся в системе управления данными. Мы

нуждаемся в системе управления различными системами управления данными, которые мы имеем!". Большинство нефтяных и газовых компаний управляло данными и информацией, используя одну или более систем управления. Эти системы имеют тенденцию ограничиваться определенными отделами, разделами, или деловыми модулями и вообще не взаимодействовать друг другом. Когда организация решает установить корпоративный репозиторий, чтобы собрать знание *всей* компании, самая большая сложность заключается в проектировании, как эта система управления знаниями будет работать с существующими данными и информационными управляющими системами.

В некоторых случаях, компания может решить переместить все данные, хранившиеся в системе одного отдела, а новую систему управления знаниями (и таким образом вывести старую систему управления данными). Это хорошая стратегия, если пользователи были недовольны производительностью или функциональностью существующей системы. Однако если пользователи отдела зависят от существующей системы для повседневного бизнеса, то им будет необходимо изменить весь бизнес-процесс. Это может иметь смысл, а может и не иметь. У хорошей системы управления знаниями должна быть достаточная встроенная гибкость, чтобы объединиться, взаимодействовать, и/или заменить существующие системы управления данными. Хорошим примером этого было бы корпоративное решение управления документами, которое сохраняет все каротажные кривые, карты, сейсмический и геологические разрезы и связывается с финансовой управляющей системой, чтобы обеспечить доступ к документам по требованию.

*Производительность и доступ.* В то время как все системы управления знаниями, разработанные на базе RDBMS, имеют хорошую производительность, когда пользователи со всех концов континентов должны обратиться к данным, есть некоторые факторы, которые могут ограничить или препятствовать доступу к данным. Хорошая система управления знаниями должна быть обеспечена распределенной архитектурой со

спутниковыми базами данных, поддерживающими непрерывный доступ, синхронизацией по расписанию, и дублированием, таким образом, предоставляя быстрый доступ отдаленным офисам к относительно новым данным.

Для областей плохими сетями нужно рассмотреть оффлайновую синхронизацию.

*Совместимость с настоящим и будущим* Хорошая система управления знаниями должна иметь возможность расширения в

будущем - и в терминах типов данных, которые она поддерживает и прикладных систем, с которыми она работает. Модель открытых данных и стандарты обмена открытыми данными обеспечивают хорошую страховку от устаревания.

#### Ссылки

Ackoff, R. L. [1989] From Data to Wisdom. Journal of Applied Systems Analysis, Vol 16, p 3-9.