

Электронный архив геолого-геофизических данных

А.А. Прохоров
(Корпорация «Электронный Архив»)

The geological-geophysical data electronic archive

A.A. Prokhorov
(Electronic Archive Corporation)

В настоящее время процессы добычи полезных ископаемых, в частности нефти, существенно усложнились, для их нормального функционирования требуются постоянный сбор и быстрый анализ большого объема разнообразной, часто разнородной информации. В добывающих компаниях уже повсеместно внедряются информационные системы разведки, разработки и эксплуатации месторождений (ИС РРЭМ), а также системы контроля и мониторинга добычи (СКМД). Однако, по-прежнему, важной задачей с точки зрения специалистов управлений геологии и разработки месторождений остается надежное хранение и возможность быстрого доступа к эксплуатационной документации объектов добычи, например, нефтяных скважин.

В настоящее время на каждую нефтяную скважину заводится специальное дело, состоящее, как минимум, из нескольких сотен бумажных документов: бумажные отчеты; проекты на бурение и разработку; карты; технологические схемы; обоснования; заключения и др. В процессе эксплуатации скважины число документов в деле увеличивается. При этом дело осложняется тем, что документы дел скважин пребывают в постоянной работе и должны находиться на промысле. Кроме того, следует помнить, что скважина – объект потенциальной опасности. Пожар на нефтепромысле – это одновременно и техногенная, и экологическая катастрофа. При чрезвычайных ситуациях необходимо, чтобы в распоряжении соответствующих служб всегда находилась полная и достоверная информация об объекте. При этом к ней должен быть организован удобный и быстрый круглосуточный доступ с одновременным соблюдением правил секретности и конфиденциальности. Для того, чтобы соблюсти указанные условия (надежное хранение и оперативный доступ к информации), необходимо создание информационной системы, ядром которой является электронный архив.

Электронный архив – это особый вид информационной системы, которая способна надежно и долговременно хранить сверхбольшие объемы электронной информации любого вида, а также обеспечивает быстрый, удобный и санкционированный доступ к информации, гарантируя ее сохранность и эффективность использования. Обеспечить надежную сохранность документов и гарантированный доступ специалистов предприятия к необходимой информации можно только созданием цифровых копий документов дел скважин и их сохранением в соответствующей информационной системе поиска и хранения электронных документов – в электронном архиве. Построение электронного архива включает следующие этапы.

Process of digitization of paper documents collections is considered. Advantages of the electronic data bursting input method are considered. Functionalities of the specialized program platform SAPERION are given.

Оцифровка фондов бумажных документов

Перевод бумажных документов в электронный вид осуществляется с помощью сканеров. Поскольку объемы документов при создании подобных архивов часто исчисляются многими сотнями тысяч листов, а разнообразие документов с трудом поддается описанию, для сканирования используются специализированные промышленные сканеры, обладающие повышенной производительностью и специальными возможностями улучшения качества получаемых электронных графических образов бумажных документов. Эта функция на практике весьма востребована, так как нередко качество части документов, особенно большеформатных, низкое.

При переводе в электронный вид схем, карт, специальных планов и подобных документов нередко возникает необходимость получения цветных графических образов. Даже если речь идет об обычной технической, технологической или ремонтной документации, которая часто бывает в работе, то и здесь могут понадобиться ее сканирование в специальных режимах и сохранение в особых форматах, например JPEG. Все указанные операции в полной мере могут обеспечить специализированные промышленные сканеры. Однако часто на практике этого недостаточно. Если документы старые, ветхие, низкого качества, выполнены на плохой бумаге, испорчены в результате интенсивной эксплуатации или небрежного хранения, то после сканирования требуется еще и дополнительная обработка полученных электронных графических образов документов.

Для обеспечения возможности поиска электронного образа документа должна быть создана соответствующая поисковая карточка. Между этой карточкой и графическим файлом,

содержащим образ документа, должна быть установлена однозначная связь. Каким образом будет представлен конкретный документ – решается на этапе проектирования системы электронного архива (СЭА). Например, несколько сотен электронных образов отсканированных страниц технического руководства могут быть представлены в виде одного PDF-файла с закладками. С ним будет связана одна поисковая карточка. Здесь необходимо пояснить несколько специальных терминов и основополагающих понятий, которыми оперируют во время выполнения подобных работ.

Электронный документ – это файл, содержащий электронный графический образ документа с привязанной к нему поисковой карточкой, например, многостраничный PDF-файл с закладками; файл формата JPEG, содержащий цветной электронный образ карт; файл, изготовленный с помощью текстового редактора MS Word.

Поисковые атрибуты – конкретные значения, занесенные в соответствующие поля поисковой карточки. Именно по ним в дальнейшем проводится поиск нужного документа в электронном архиве. Храниться поисковые карточки могут в любой промышленной СУБД. Процесс создания поисковой карточки и привязки ее к электронному образу документа называется индексированием.

Возможны два способа наполнения СЭА информацией: 1) самостоятельное проведение сканирования бумажных документов; 2) применение так называемый метода «залпового ввода».

При выполнении работ своими силами компании по сути придется организовать у себя производственную линию, прежде всего создать участок ввода и оцифровки бумажных документов, а для этого оснастить внушительным парком сканирующего оборудования. Поскольку бумажные документы, содержащие геолого-физические данные, отличаются большим многообразием, и электронный архив изначально должен быть наполнен большим объемом информации, для обработки и введения данных необходимо минимум три сканера:

- поточный – для сканирования расшитых документов форматом не более А3 (ремонтно-эксплуатационная, технологическая, проектно-техническая документация и др.);
- планетарный – для сканирования документов, которые невозможно подвергнуть расшивке;
- широкоформатный – для сканирования большеформатных документов (чертежей и схем, планов, карт и др.).

Производительность работы каждого из этих сканеров существенно различается. Если для поточного сканера она может составлять от 3 тыс. до 5 тыс. листов бумажных документов формата А4 за 8-часовую рабочую смену, то производительность работы планетарного сканера за такое же время составляет не более 800 страниц. При этом для протяжных сканеров максимальная производительность варьирует в пределах 300 листов. Указанные цифры не отражают чистую скорость работы сканера. Процесс перевода в электронный вид бумажных документов содержит множество рутинных операций, которые с трудом поддаются автоматизации. Прежде всего надо иметь в виду, что документация, как правило, сброшюрована в книги (тома, папки, скоросшиватели и др.), которые можно расшивать, причем в 99,9 % случаев это неоднородный массив. В нем встречаются вставки документов большего формата, цветные врезки, чертежи, схемы и др. Поэтому непосредственно операции сканирования обязательно предшествует ряд других обязательных операций: расшивка, сепарация, прокладка технологическими картами, распределение документов по сканерам и др. После сканирования выполняется обратная сборка и сшивка документов. Иногда часть

ветхих бумажных документов может быть даже отреставрирована. Данные о производительности сканеров собраны на основе более чем десятилетнего опыта практической работы, в частности, на основе информации, полученной в результате создания СЭА в ОАО «Самотлорнефтегаз», ОАО «ТНК-Нефтьартовск», ОАО «Пурнефтегазгеология», ООО «Кавказтрансгаз» и др. Для того чтобы перевести в электронный вид бумажный архив, состоящий всего из нескольких десятков или сотен тысяч листов, потребуются значительные людские и материальные ресурсы. Кроме того, для сохранения в работоспособном состоянии сканеры потребуют специального технического обслуживания, а после окончания работ эта дорогостоящая техника останется на балансе предприятия.

Метод залпового ввода

При этом методе заказчик находит специализированную компанию-исполнителя, которая проводит весь спектр работ и выдает готовый электронный информационный ресурс (ЭИР), состоящий из электронных графических образов документов и связанных с ними поисковых карточек. ЭИР потом может быть загружен в соответствующую информационную систему.

Преимущества подобного подхода состоят в том, что решенные задачи можно целиком возложить на компанию-исполнителя. Последняя должна иметь не менее нескольких десятков единиц сканирующего оборудования для обработки бумажных документов различного вида, большой штат операторов-сканировщиков, операторов индексирования, а также сотрудников, занятых на вспомогательных работах (сшивка, расшивка, сепарация и др.). Благодаря этому работы можно провести в короткие сроки и через несколько месяцев получить готовый результат. При этом сразу исключается проблема, как использовать закупленную технику и набранный штат специалистов после окончания работ.

В ходе выполнения работ заказчик получает следующие дополнительные преимущества.

1. Оцифровка фондов бумажных документов любых качества и формата.

Работы могут выполняться выездными бригадами на территории заказчика, в том числе, например, на нефтяных и газовых промыслах, в предельно сжатые сроки. Средняя скорость обработки может составлять от 1500 дел скважин в месяц.

2. Обработка цифровых копий, индексирование и подгонка данных для наполнения электронного архива.

Работы выполняются специалистами Центра ретроспективной конверсии компании-исполнителя в сжатые сроки с обеспечением заданных критериев качества. В ходе работ выделяются этапы эксплуатации скважины и проводится их индексирование по соответствующим реквизитам. Для большего удобства поиска в электронном архиве специфические названия документов унифицируются. Например, многие дела скважин ведутся с двадцатого века. За этот период точные наименования документов менялись не один раз. Поэтому во избежание путаницы документы одного типа приводятся к единой согласованной с заказчиком форме наименования. Данная проблема характерна для бумажных архивов геолого-геофизических данных. Часто их источниками являются документы, различающиеся и по времени разработки, виду и по материалу, на котором они выполнены, по тому, кто собирал и систематизировал представленные данные. На этапе проектирования структуры ЭИР заказчик может определить не только число полей индексирования и правила их заполнения, но и единую форму наименования, что помогает избе-

жать путаницы документов, содержащих информацию одного типа.

3. Обработка цифровых копий, индексирование и подготовка данных для наполнения электронного архива.

Работы выполняются сотрудниками Центра ретроспективной конверсии компании-исполнителя в сжатые сроки с обеспечением заданных критериев качества. Требуемый состав и наименования полей для каждого типа документов определяются, исходя из требований заказчика. Для повышения качества работ желательно применять технологии параллельного ввода и верификации информации согласно стандарту ISO 9001:2000.

4. Загрузка электронных данных (графических образов документов и базы данных) в систему архивного хранения.

Система архивного хранения обеспечивает сохранность документов электронного архива и быстрый, удобный и одновременно защищенный доступ к ним. Эти цели достигаются разными способами в зависимости от того, на какой программной платформе строится СЭА. Например, корпорация «Электронный Архив» предлагает специализированную программную платформу САПЕРИОН, на которой были построены многие из вышеуказанных решений. Для платформы САПЕРИОН характерна следующая особенность: все данные хранятся в виде специального закодированного файла собственного формата (мета-файла) для исключения возможности несанкционированного копирования и использования всей информации или ее части. Это обеспечивает надежную защиту и хранение информации, так как каждая СЭА обладает своим собственным, уникальным ключом защиты, поэтому даже владелец другой системы САПЕРИОН не сможет получить доступ к данным.

Благодаря функциональным возможностям СПО САПЕРИОН при работе с документами электронного архива обеспечиваются:

- быстрый поиск по заданным параметрам;
- оперативный доступ к данным архива;
- многопользовательский доступ к документу, в том числе из удаленных подразделений;
- ограничение доступа к документу/фрагменту документа;
- управление маршрутами движения документа;
- передача документов на удаленные участки;
- внесение пометок в электронные образы документов.

Важной особенностью системы СПО САПЕРИОН является возможность интеграции со специализированными приложениями СКМД, ИС РРЭМ и др. В результате новые создаваемые файловые документы могут сразу помещаться в электронный архив с сохранением ссылки на документ из интерфейса любого специализированного приложения.

Сохранность документов в СЭА обеспечивается также на аппаратном уровне путем использования так называемых архивных накопителей или роботизированных библиотек на базе DVD/BD/UDO-дисков, с которыми САПЕРИОН может работать напрямую. Такие архивные накопители способны хранить существенные объемы данных (в одном накопителе

может храниться до 34 ТБ информации). При этом данные, расположенные на оптических дисках, энергонезависимы, могут быть продублированы в любом числе экземпляров, причем копии могут храниться в специальном сейфе, удаленном на много километров от промышленного объекта. Это обеспечивает абсолютную катастрофоустойчивость системы архивного хранения документов при возникновении чрезвычайных ситуаций.

Электронный архив обладает еще рядом преимуществ.

Эффективность:

- доступность электронной копии любому числу сотрудников, включая сотрудников территориально удаленных подразделений;

- сокращение времени и трудоемкости процессов передачи документации производственным, эксплуатационным и ремонтным подразделениям;

- получение неограниченного числа копий документов в бумажном и электронном видах без обращения к бумажному оригиналу.

Надежность:

- централизованное хранение в едином электронном фонде (архиве) с единым управлением: документы не дублируются, упрощается их поиск и др.;

- гарантированный срок хранения информации электронного фонда на специализированных накопителях несколько десятков лет;

- отсутствие необходимости резервного копирования, быстрое восстановление работы электронного фонда после сбоя.

Безопасность:

- гарантированная защита как от случайного, так и от попыток умышленного уничтожения страхового фонда (архива) пользователями;

- регламентация прав доступа к делам, документам, отдельным страницам или аннотациям;

- возможность просмотра архивных копий документов, которые были удалены.

Правильно разработанный и наполненный электронный архив должен стать ядром информационной системы всего предприятия, тем необходимым звеном, которое объединит многочисленные программные приложения и разнородную геолого-геофизическую информацию, имеющуюся всегда в избытке на предприятиях. Кроме того, электронный архив может избавить организацию от «лоскутного» хранения данных, когда каждое специализированное программное приложение хранит свои данные в своем собственном хранилище и испытывает большие проблемы как из-за роста их объема, так и из-за невозможности обмена этими данными с другими приложениями. В результате повышаются общий уровень информационного обеспечения предприятия, эффективность обработки и использования имеющихся данных.