

Трудности при вертикальном сейсмопрофилировании: проблемы архивации данных ВСП

Vertical seismic pitfalls: problems in archiving VSP data

Компании должны удостовериться, что хранимые ими данные ВСП заархивированы и находятся в хорошем состоянии, предупреждают Eleanor Jack*, главный геофизик отдела Information Management Practice компании Landmark. Для подтверждения своей точки зрения она приводит несколько примеров того, что зачастую все может выйти из строя.

Вертикальные сейсмические профили являются мостом между скважинными и сейсморазведочными данными, точкой, где пересекаются время и расстояние, где неопределенности в поверхностных сейсмических данных устраняются привязкой к скважинным данным. Также, данные ВСП должны быть надежными элементами архива, те в которых мы уверены больше всего. И все же при архивации данных ВСП, мы попадаем в мир конфликтующих форматов, процедур и технологий, где едва ли что-нибудь стандартизировано и все рискованно. Как это может быть?

Сводка проблем

Проблема частично заключается в том, что никто на самом деле не знает, куда поместить данные ВСП, которые плохо вписываются в базу скважинных данных, поскольку они являются, главным образом, сейсмическими, и не вписываются в сейсмическую базу данных, поскольку они относятся к скважине. Кроме того, есть, или скорее должно быть, разнообразие других типов данных, типа отчетов и привязанных коротажных кривых в дополнение к сейсмическим данным, все из которых имеют различную архивацию и требования к индексации.

Сейсмические данные ВСП также характеризуются свойственными проблемами, связанными, с тем, что имеются три компоненты: X, Y и Z. Многокомпонентные данные ВСП появились задолго до того, как кто-либо изобрел формат для обращения с ними, и поэтому такие данные подгоняли под формат SEG-Y множеством изобретательных путей. (Первая редакция формата SEG-Y наконец обозначила эту проблему в 2002 году, до которого накопился 20-летний или около того запас данных в первичном формате SEG-Y). Обработка этих данных на самом деле никогда не стандартизировалась, что привело к большому разному результатам обработки данных.

*ejack@lgc.com

Данные обрабатывались практически без учета того, что с этими данными могут работать и другие люди, кроме исполнителя, и это приводило к некоторым действительно странным восприятиям того, что на самом деле содержалось в формате SEG-Y.

Как будто этого было недостаточно, также имеет место фундаментальная разница в подходах формирования пакетов скважинных и сейсмических данных. Использование формата упаковки TIF (Tape Image Format) является стандартным для коротажных диаграмм, но, к сожалению, не для сейсмических данных. Архив, содержащий как коротажные диаграммы, так и сейсмические данные, может вполне быть смесью инкапсулированных и неинкапсулированных данных, однако, если архивариусы сделают попытку объединенного подхода и начнут инкапсулировать свои сейсмические данные, это приведет к тому, что данные станут нечитаемыми для большинства приложений. Этот вопрос будет подробно описан в одном из последующих примеров.

Ситуация, в итоге, является волнующей: в то время как очень сложно получать данные ВСП, их чрезвычайно легко потерять.

Типы данных ВСП

Полный набор данных ВСП должен включать:

- Сырые сейсмические данные (переведенные в формат SEG-Y)
- Обработанные сейсмические данные (в формате SEG-Y)
 - предварительные суммарные разрезы
 - результаты различных этапов обработки
 - разрез коридорного суммирования
 - синтетические сейсмограммы
- Привязанные коротажные диаграммы (LIS, LAS или DLIS)
 - обеспечивающие привязку время-глубина между данными ВСП и наземной сейсморазведки

- Отчеты (сканированные документы)
 - важные сведения, связанные с наблюдением и обработкой
 - навигация
 - таблицы скорость/время и вычисления
- Оцифрованные части отчета
 - (ASCII или электронные таблицы Excel, часто на гибких дисках)

К сожалению, с таким многообразием типов данных, на разных носителях, несейсмические данные часто исчезают в различных базах данных и библиотеках, и это обычно означает, что они исчезают полностью.

Но необходимо задаться вопросом, сколько примеров использования сейсмических данных без отчетов, говорящих нам, как и где была сделана съемка? И насколько разумно разделение сейсмических данных ВСП от некоторых результатов съемки ВСП (привязанные коротажные диаграммы и расчеты скоростей)?

Пробное технико-экономическое обоснование CDA

В 2004 г. Landmark выполнил технико-экономическое обоснование в оценке качества и загрузке данных ВСП в общую выборку данных (CDA — Common Data Access) DataStore. Мы подробно опишем это исследование, как пример реального мира ВСП в микромире. Данные получены по 12 скважинам, каждые три из которых принадлежат различным клиентам, с дальнейшим сбором данных по скважинам уже в CDA. Скважины были отобраны так, чтобы получить данные различных типов съемок, типов скважин и исполнителей. Результаты данного исследования сведены в табл. 1.

При первом обзоре таблицы сразу бросается в глаза сильная изменчивость объемов и типов данных по различным скважинам. По трем скважинам (2, 3 и 4) не имеются сырые сейсмические данные. По

скважине 3 вообще нет никаких сейсмических данных (что доказывает, что отбор скважин был честным!). Но есть и другие проблемы по скважине 3, кроме простого отсутствия всех данных, и это мы увидим в первом практическом примере.

Для большей части скважин отсутствуют привязанные каротажные диаграммы и таблицы скоростей, за исключением данных владельца D. Однако, в этом случае, мы должны кое в чем признаться. Данные владельца D были отобраны компанией Landmark из стороннего хранилища, используя критерии отбора, которые, возможно, отобрали в результате больше данных для этих скважин, чем по другим. Мы рассмотрели все имеющиеся данные по выбранным скважинам и отобрали все связанное с сейсмикой, все, что содержит номенклатурное обозначение VSP (BCП) или BSP (скважинная сейморазведка или его вариант реализации!), все каротажные кривые, обозначенные как «калибровано» и все, что включало слова «кривая скорости», «контрольный взрыв», «гидрофоны» и «геофоны».

Первый практический пример проблемы: скважина 3 — сырые данные

Мультиплексированные полевые сырые данные были получены по скважине 3 из стороннего хранилища данных, но это были копии, не исходные данные, и в процессе

копирования, длинные блоки исходных файлов SEG-B разделялись на две части. После склеивания блоков и демуплексирования данных, мы обнаружили аномалии в канале синхронизации, с сильной вероятностью того, что они были вызваны расщеплением длинных мультиплексных блоков. (Канал синхронизации дает время в миллисекундах любой развертки данных и если это не простой растущий ряд, например 0,1,2,3,4,5,6, ..., 1000, то вы действительно должны знать почему!) Поэтому мы запросили исходные ленты, однако получали вместо этого демуплексированную запись с ленты. К сожалению, эта запись содержала те же аномалии синхронизации, как и наша, из чего мы сделали вывод, что запись была сделана по тем же поврежденным копиям.

Без исходных лент, мы не смогли убедиться в том, что данные были привязаны к правильному времени, и неохотно должны были исключить эти данные из испытания. Здесь следует запомнить, что опасно копировать мультиплексированные данные и что записи должны быть сделаны с исходных носителей.

Второй практический пример проблемы: скважины 4, 5, 6 — инкапсуляция данных в формат TIF

Инкапсуляция в формат TIF компании Atlas является цифровым способом маркировки межблочных промежутков при чтении данных с

ленты или картриджа, данный формат является стандартом для каротажных диаграмм. TIF может использоваться для инкапсуляции содержания ленты в один файл, что является общей практикой при загрузке каротажных кривых в банк данных, поскольку это делает загрузку проще и обеспечивает целостность информации. Как указано ранее, это не является стандартной практикой для сейсмических данных, однако, если сейсмические данные были загружены в базу данных каротажных кривых, вполне можно обнаружить, что некоторые из них также были инкапсулированы в формате TIF.

Данные по скважинам 4, 5 и 6 были записаны на компакт-диски после выгрузки из базы данных. Шесть файлов являлись инкапсулированными образами лент, пять — неинкапсулированными. При распаковке шесть файлов-образов лент, было получено 26 файлов в формате SEG-Y, исходное содержание шести лент.

Очевидно, практика хранения SEG-Y файлов в различных состояниях инкапсуляции в одной базе данных нежелательна, как в ситуации, где некоторые SEG-Y файлы записаны в формате (кратко инкапсулированы), который не могут читать сейсмические приложения. Однако прежде, чем мы станем слишком строгими по поводу этого, мы должны узнать о том, что случилось потом, с тех пор как четыре из 26 файлов, оказалось, содержат серьезные ошибки, которые

Скважина	Владелец данных	SEG-Y		Привязанные каротажные кривые		Файлы в формате ASCII		Наличие сейсмических данных	
		Количество файлов	Ленты	Файлы	Ленты	Файл	Ленты	Сырые	Обработанные
1	A	12	2	4	2			Есть	Есть
2	A	2	1					Нет	Есть
3	A							Нет	Нет
4	B	20	4	1	1	1	1	Нет	Есть
5	B	5	5			1	1	Есть	Есть
6	B	6	2			3	2	Есть	Есть
7	C	12	2	1	1			Есть	Есть
8	C	7	5					Есть	Есть
9	C	9	2					Есть	Есть
10	D	20	5	1	1	3	2	Есть	Есть
11	D	18	3	5	4	3	2	Есть	Есть
12	D	17	2	1	1	11	4	Есть	Есть
Итого		128	33	13	10	22	12		

Табл. 1. Техничко-экономическое обоснование CDA

было бы почти невозможно устранить, если бы данные не были инкапсулированы.

Третий практический пример проблемы: скважина 4 — поврежденные заголовки файла

Каждый случай немного отличался, однако одного примера достаточно. Согласно формату SEG-Y начальный заголовок представляет собой EBCDIC, размером 3200 байт, за ним следует двоичный заголовок размером 400 байт, затем следуют трассы данных. В последующем примере пропущен двоичный заголовок.

Поскольку данные были инкапсулированы в формате TIF, мы могли проанализировать длины блоков, и, таким образом, увидеть, что было только 2 байта вместо 400-байтового двоичного заголовка, части, читаемой машиной, которая содержит важную информацию, такую как шаг дискретизации, код формата данных и т.д. Для устранения данной проблемы два дефектных байта были удалены и был вставлен новый двоичный заголовок, скопированный из другого файла. Этот новый двоичный заголовок был затем отредактирован, так, чтобы информация, которую он содержал, согласовывалась с данными.

Поврежденные трассы

В вышеописанном случае, ошибка была связана с заголовками файла, но ошибки могут легко содержаться и в трассах, и данные после распаковки будут выглядеть как нечто, похожее на рис. 3а, вместо исправленной картины, рис. 3б.

Тип блока	Размер блока	Описание
^ :неизв:	: 3200:	0: - заголовок EBCDIC
^ :неизв:	: 1:	0: - Остатки двоичного заголовка
^ :неизв:	: 1:	0: - Остатки двоичного заголовка
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса
* :ДАННЫЕ:	: 5240:	0: - Сейсмотрасса

Табл. 2. Сгенерированный компьютером список размеров блоков поврежденного файла SEG-Y. Первые две колонки показывают тип блока и его размер (в байтах). Последняя описательная колонка добавлена автором.

Идентификация данных

Идентификация данных может иногда быть хитрой, часто когда данные похожи на рис. 4, и заголовок EBCDIC, содержащий всю необходимую информацию, похож на рис. 5.

При записи данных с полевого формата, формируется заголовок EBCDIC, содержащий немного информации об устройстве воспроизведения, и весьма часто

заголовки EBCDIC после предыдущих записей оказываются полностью пустыми. Однако заголовки EBCDIC обработанных данных могут быть также короткими, с небольшими оговорками.

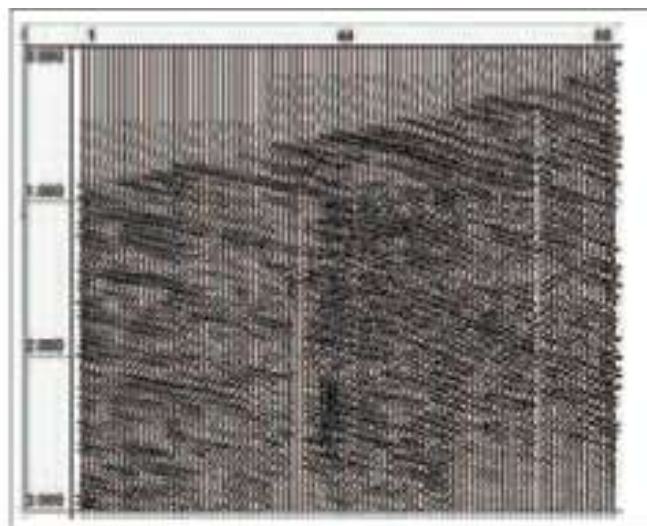
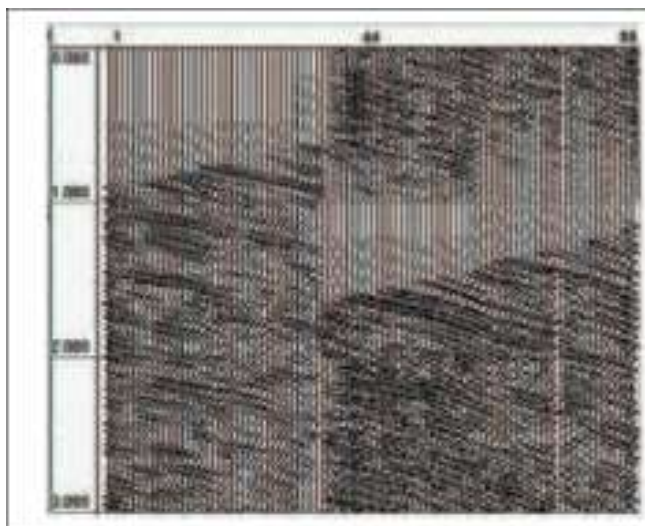


Рис. 3. а) короткая трасса приводит к нарушению синхронности данных

б) скорректированные данные

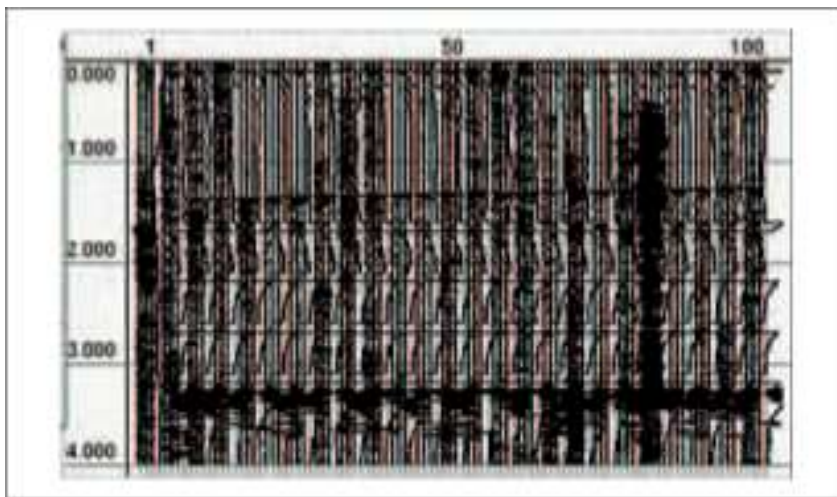


Рис. 4. Сырые 4-х компонентные данные VSP

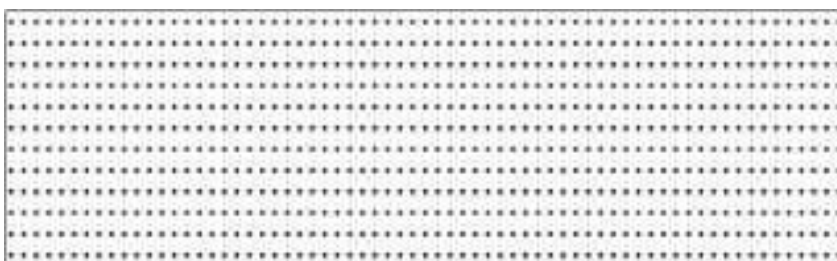


Рис. 5. Типичный заголовок EBCDIC — полностью пустой

Проблемы с двоичным заголовком

Проблемы с двоичным заголовком потенциально смертельны для данных, поскольку двоичный заголовок содержит такую информацию, как шаг дискретизации — основа декодирования данных. Однако согласно первому закону Мёрфи (то, что не невозможно, может произойти), мы видели двоичные заголовки, в которых весь существенный код формата, который определяет количество байтов в выборке и информацию о том записаны ли данные как числа с плавающей запятой или фиксированы, являлся нулевым, то есть не был определен. По-видимому, этот пользователь всегда использовал один и тот же формат и ожидал, что другие сделают то же самое.

Аналогично были случаи, в которых был нулевой шаг дискретизации — обработчики здесь, по-видимому, всегда использовали типовой интервал, заданный в заголовках трасс, игнорируя тот факт, что большинство приложений использует только двоичный заголовок. И для справедливости приведем пример, в котором шаг дискретизации был включен в двоичный заголовок, однако во всех

заголовках трасс был нулевым. Эти оба примера относятся к данным крупных подрядчиков.

Четвертый практический пример проблемы: скважина 12 — перепутаны компоненты X, Y, Z

До данного примера, рассмотренные проблемы влияли на считываемость или иначе на данные, но сейчас мы рассмотрим более сложную проблему, воздействующую на целостность информации. По скважине 12 были перепутаны компоненты X, Y и Z. Это случилось в процессе реорганизации данных подрядчиком, который взял исходные многокомпонентные файлы (и, поскольку так произошло, их дубликаты), разделил их на отдельные компоненты, а затем ошибочно идентифицировал некоторые из них, вероятно просто, внося их в список в алфавитном порядке вместо численного, до их маркировки.

Если бы мы имели дело только со скважиной 12, то эта проблема, вероятно, не была бы выявлена, однако скважина 12 входит в большую группу скважин, по которой мы сделали контроль качества. Приблизительно для 20 % этой группы данная проблема имела место.

Наличие дубликатов исходных данных стало основным элементом при контроле качества, поскольку некоторые файлы, которые должны были быть одинаковыми, ясно отличались, а другие файлы, которые должны были предположительно отличаться, были, фактически, одинаковыми. Кроме того, в исходных данных использовался определенный номер канала для каждой компоненты, и этот номер канала также сохранялся в заголовках трасс. Последним подарком стала сгенерированная компьютером информация, содержащаяся в заголовках EBCDIC, которая не согласовывалась с описаниями, добавленными вручную.

Решение проблемы было легким для необработанных данных; мы просто исправили содержание заголовков, однако проблема оставалась нерешенной для обработанных данных. Сгенерированная компьютером информация указывала на то, что ошибки «проползли» в обработку, и единственным решением проблемы, кроме переработки, являлось добавление предупреждающего сообщения в заголовки для обозначения того факта, что данные были обработаны с беспорядком в компонентах.

Закключение

Результаты данного предварительного исследования серьезны:

- по 25 % скважин потеряны все сырые сейсмические данные
- по 50 % скважин отсутствуют привязанные каротажные кривые (одна из основных целей выполнения съемки),
- по 45 % тех скважин, по которым действительно имелись данные сейсморазведки, имеются проблемы с этими данными.
- по 8 % скважин вообще нет данных

Если результаты этого исследования окажутся типичными (а с нашим выбором скважин мы попытались сделать их таковыми), то владельцы данных должны беспокоиться за сохранение своих данных, испытав значительные трудности и расходы для их получения. Решения просты — обозначать все важные данные как VSP, так, чтобы вы могли обнаружить их снова, сохранять все связанные данные (SEG-Y, таблицы, отчеты, привязанные каротажные диаграммы) вместе, и сохранять их на носителях, подходящих для различных типов данных, таких как CD/DVD или постоянные запоминающие

устройства. Прежде всего, важен полный и интеллектуальный контроль качества; иначе владелец может обнаружить себя в «раю дурака».

Данные ВСП дорого приобрести, однако их можно бесплатно потерять.

Благодарности

Благодарим Common Data Access, ConocoPhillips и BG.