Aspen Systems Corporation™
Culture of Commitment

2277 Research Boulevard • Rockville, MD 20850
www.aspensys.com

February 21st, 2006

Request ID: WMCU0356

Shipping Address:

Christina Nobelman
University of Maryland
Central Receiving Facility
Bldg 383
Paint Branch Pkwy
College Park MD 20742

Enclosed please find materials requested by you pertaining to the Federal Energy Regulatory Commission (FERC) Western Markets Investigation. The items on the list below are included in this package.

- Enron Email database
- Enron Email (re-released) database
- Enron Email (.pst) database
- Enron Email (.pst) (re-released) database
- Transcripts
- Scanned Documents database
- Scanned Documents (re-released) database

These names refer to the names applied to these databases on the FERC public access site, where they are available through the "Search iCONECT24/7" interface. The URL for this site is http://www.ferc.gov/industries/electric/indus-act/wem/pa02-2/info-release.asp.

Shipping Method: FEDEX First Overnight

For any issues, please call Giri Namasivayam at (301)519-6396.

The files on the included USB hard drive are organized as follows.

Enron Email Databases

The Enron Email databases are in the directory EnronEmails. In this directory, the file enron_email_PA_alldocs_(default_delim)120203.txt contains the "Enron Email" database; the file enemail_putbacks_120204(default_delim).txt contains the "Enron Email (re-released)" database; the file Email_Crossreference_Combined.txt contains a cross-reference between a record and its attachment file paths, as explained later in this document. The re-released database consists of records that have been redacted by the FERC and re-released to the public. All the attachment files are contained in the subdirectories within the EnronEmails directory. All attachment files are rendered as TIFF formatted images. The field names, which are included as the first line of the delimited text files, are: FIRSTBATES, LASTBATES, BEGIN_ATTACH, END_ATTACH, TITLE, SUBJECT, EDOC_CREATED, EDOC_CRDATE, EDOC_MODIFY, EDOC_MODATE, EDOCFILENAME, EDOCFILEPATH, EDOC_SOURCE, SENT, DATE_SENT, RECD, DATE_RECD, FOLDER_LOC, CUSTODIAN, ORIGIN, DOC_TITLE, LANGUAGE, BARCODE_NUM, SOURCE, CASE_NAME, TO, FROM, CC, BCC, UNK, OCR_TEXT. Not all fields will necessarily contain values.

Enron Email (.pst) Databases

The Enron Email (.pst) databases are in the directory EnronEmailsPST. In this directory, the file EnronEmailPst_120604(default_delim).txt contains the "Enron Email (.pst)" database. The file EnronEmailPst_putbacks_120204(default_delim).txt contains the "Enron Email (.pst) (re-released)" database. All the associated attachment files are contained within sub-directories inside the path "\achilles\data\ferc\enron_pst". The exported data contains the pathnames to these attachments in the 'Attachment' field. The directory structure matches the directory path used within the strings in the attachment field. Note that the attachment field might contain multiple strings, each of which is represents a separate file path. These strings are separated by the character ASCII DEC 174. Of course, not all records will have attachments.

To locate the attachment(s) associated with each record, parse the attachment field to obtain the individual filename and paths. Then follow the path to locate the file.

The field names are: SDOC_NO, BOX_NO, MEDIA_LABEL, FILENAME, FROM, ORG, TO, CC, BCC, DATE, TIME, ORIGIN, SUBJECT, FOLDER, ATTACHMENT, HEADER, MESSAGEID, BODY. Not all fields will necessarily contain values.

Scanned Documents Databases

The Scanned Documents databases are in the ScannedDocuments directory. The file scdocs_PA_alldocs_(default_delim)120203.txt contains the "Scanned Documents" database; the file scdocs_putbacks_120204(default_delim).txt contains the "Scanned Documents (re-released)" database; the file scandocs_crossreference_combined.txt contains a cross-reference between a record and its attachment file paths, as explained later in this document. The re-released database consists of records that have been redacted by the FERC and re-released to the public. All the attachment files are contained in the subdirectories. All attachment files are rendered as TIFF formatted images. The field names, which are included as the first line of the delimited text files, are: FIRSTBATES, LASTBATES, FIRSTATTACH, LASTATTACH, BOX_DESCRIPT, DATE,CONDITION, DOC_TYPE, OTHER_NO, TITLE, AUTHOR, ADDRESSEE, COPYEE, TRADER_NAME, OCR_TEXT. Not all fields will necessarily contain values.

Transcripts

The transcripts database is in the directory Transcripts. In this directory, the file transcripts(default_delim)121703.txt contains transcripts of depositions taken by the FERC.

Record Structure

The delimited data files use the following characters: ASCII Dec 20 for field separator, ASCII Dec 254 for text delimiter, and ASCII Dec 174 for string separator within a field. The last one is for fields, like the Attachment field, which contain multiple strings that are to be considered separately. These strings might, for example, have been originally separated by the CR-LF pairs and replaced during the export by the ASCII Dec 174 character.

Processing The Enron Emails and Scanned Documents Databases

Here is some information that will assist in processing the Enron Emails & Scanned Documents databases. Each record in these recordsets can have multiple images associated with it. For instance, if the message body is long, it might have been rendered into multiple image files. Similarly, any native file attachments might have been rendered into multiple image files. Note that ALL attachments are only available as image files. There are four columns of interest here from the delimited data. They are firstbates, lastbates, beginattach, endattach. The range represented by the firstbates and lastbates field values indicates the number of image files that are associated with that particular record. The range represented by the BeginAttach and EndAttach fields indicates the number of image files associated with the email message as a whole, including any attachments.

For example, consider the first two records in the listing below (not all the columns in the delimited record are shown here):

Record	FirstBates	LastBates	BeginAttach	EndAttach
1	ECD-000001903	ECd-000001903	ECd-000001903	ECd-000001909
2	ECD-000001904	ECd-000001909	ECd-000001903	ECd-000001909
3	ECD-000001875	ECd-000001877	ECd-000001875	ECd-000001890
4	ECD-000001878	ECd-000001881	ECd-000001875	ECd-000001890
5	ECD-000001882	ECd-000001886	ECd-000001875	ECd-000001890
6	ECD-000001887	ECd-000001889	ECd-000001875	ECd-000001890
7	ECD-000001890	ECd-000001890	ECd-000001875	ECd-000001890

The FirstBates and LastBates values are the same for the first record, indicating that there is only one image file associated with that record. The BeginAttach and EndAttach values on that record end with 1903 and 1909, respectively. This is interpreted to mean that there are other records in the dataset where the FirstBates and LastBates values fall in the range 1904 to 1909. This is indeed the case in the second record. Thus, the first two records are both parts of the same email message, including its attachment. Each unique BeginAttach, EndAttach range represents one complete email message with attachments. In the listing above, records 3 through 7 all represent parts of the same message.

The crossreference file will contain two columns that represent bates number, and file path. To discover the path to any image file, take the bates number for that file (from the ranges in the data record) and look up the path in the corresponding crossreference file.

\