U.S. Securities and Exchange Commission

Data Delivery Standards

The following outlines the technical requirements for producing scanned paper collections, email and electronic document/ native file collections to the Securities and Exchange Commission. The SEC uses *Concordance®* 2007 v9.58 and *Concordance Image®* v4.53 software to search, review and retrieve documents produced to us in electronic format. Any proposed production in a format other than those identified below, the proposed use of *Predictive Coding*, *computer-assisted review* or *technology-assisted review* (TAR), or the use of de-duplication during the processing of documents, must be discussed with and approved by the legal and technical staff of the Division of Enforcement (ENF) and the methodology must be disclosed in the cover letter. We appreciate your efforts in assisting us by preparing data in a format that will enable our staff to use the data efficiently.

General Instructions

1. A cover letter should be included with each production. *This letter MUST be imaged and provided as the first record in the load file.*
   The following information should be included in the letter:
   a. List of each piece of media (hard drive, thumb drive, DVD or CD) included in the production by the unique number assigned to it, and readily apparent on the physical media.
   b. List of custodians, identifying:
      1) The Bates range (and any gaps therein) for each custodian
      2) Total number of records for each custodian
      3) Total number of images for each custodian
      4) Total number of native files for each custodian
   c. List of fields in the order in which they are listed in the data file.
   d. Time zone in which emails were standardized during conversion (email collections only).
2. Documents created or stored electronically MUST be produced in their original electronic format, not printed to paper or PDF.
3. Data can be produced on CD, DVD or hard drive; *use the media requiring the least number of deliverables.*
4. Label all media with the following:
   a. Case number
   b. Production date
   c. Bates range
   d. Disk number (1 of X), if applicable

5. Organize productions by custodian, unless otherwise instructed. All documents from an individual custodian should be confined to a single load file.
6. All productions should be checked and produced free of computer viruses.
7. All produced media should be encrypted.
8. Passwords for documents, files, compressed archives and encrypted media should be provided separately either via email or in a separate cover letter from the data.

Delivery Formats

I. *Concordance®* **Production**

All scanned paper, email and native file collections should be converted/processed to TIFF files, Bates numbered, and include fully searchable text. Additionally, email and native file collections should include linked native files.

Bates numbering documents:
The Bates number must be a unique, consistently formatted identifier, i.e., an alpha prefix along with a fixed length number for EACH custodian., i.e., ABC0000001. This format MUST remain consistent across all production numbers for each custodian. The number of digits in the numeric portion of the format should not change in subsequent productions, nor should spaces, hyphens, or other separators be added or deleted.

The following describes the specifications for producing image-based productions to the SEC and the load files required for *Concordance®* and *Concordance Image®*.

1. **Images**

   a. Images should be single-page, Group IV TIFF files, scanned at 300 dpi.
   b. File names cannot contain embedded spaces.
   c. Bates numbers should be endorsed on the lower right corner of all images.
   d. The number of TIFF files per folder should not exceed 500 files.
   e. Rendering to images PowerPoint, AUTOCAD/ photographs and Excel files:
      1) PowerPoint: All pages of the file should be scanned in full slide image format, with any speaker notes following the appropriate slide image.
      2) AUTOCAD/ photographs: If possible, files should be scanned to single page JPEG (.JPG) file format.
      3) Excel: TIFF images of spreadsheets are not useful for review purposes; because the imaging process can often generate thousands of pages per file, a placeholder image, named by the *IMAGEID* of the file, may be used instead.

2. *Concordance Image®* **Cross-Reference File**

The image cross-reference file is needed to link the images to the database. It is a comma-delimited file consisting of seven fields per line. There must be a line in the cross-reference file for every image in the database.

The format is as follows:
ImageID,VolumeLabel,ImageFilePath,DocumentBreak,FolderBreak,BoxBreak,PageCount

ImageID: The unique designation that *Concordance®* and *Concordance Image®* use to identify an image.
Note: *This imageID key **must** be a unique and fixed length number. This number will be used in the .DAT file as the ImageID field that links the database to the images. The format of this image key must be consistent across all productions. We recommend that the format be a 7 digit number to allow for the possible increase in the size of a production.*

VolumeLabel: Optional

ImageFilePath: The full path to the image file.

DocumentBreak: The letter "Y" denotes the first page of a document. If this field is blank, then the page is not the first page of a document.

FolderBreak: Leave empty

BoxBreak: Leave empty

PageCount: Optional

Sample

```
IMG0000001,,E:\001\IMG0000001.TIF,Y,,,
IMG0000002,,E:\001\IMG0000002.TIF,,,,
IMG0000003,,E:\001\IMG0000003.TIF,,,,
IMG0000004,,E:\001\IMG0000003.TIF,Y,,,
IMG0000005,,E:\001\IMG0000003.TIF,Y,,,
IMG0000006,,E:\001\IMG0000003.TIF,,,,
```

3. *Concordance®* **Data File**
   The data file (.DAT) contains all of the fielded information that will be loaded into the *Concordance®* database.

   a. The first line of the .DAT file must be a header row identifying the field names.
   b. The .DAT file must use the following *Concordance®* default delimiters:
      Comma   ¶   ASCII character (020)
      Quote   þ   ASCII character (254)
      Newline   ®   ASCII character (174)

   c. Date fields should be provided in the format: mm/dd/yyyy
   d. All attachments should sequentially follow the parent document/email.
   e. All metadata associated with email, audio files, and native electronic document collections must be produced (see pages 4-5).

   f. The .DAT file for scanned paper collections must contain, at a minimum, the following fields:
      1) FIRSTBATES: Beginning Bates number
      2) LASTBATES: Ending Bates number
      3) IMAGEID: Image Key field
      4) CUSTODIAN: Individual from whom the document originated
      5) OCRTEXT: Optical Character Recognition (file path, or text)

   Sample of .DAT file (when text files are provided separately)

```
þFIRSTBATESþ¶þLASTBATESþ¶þIMAGEIDþ¶þCUSTODIANþ¶þOCRTEXTþ
þPC00000001þ¶þPC00000002þ¶þIMG0000001þ¶þSmith, Johnþ¶þE:\TEXT\PC00000001.TXTþ
þPC00000003þ¶þPC00000003þ¶þIMG0000003þ¶þSmith, Johnþ¶þE:\TEXT\PC00000003.TXTþ
þPC00000004þ¶þPC00000005þ¶þIMG0000004þ¶þSmith, Johnþ¶þE:\TEXT\PC00000004.TXTþ
```

   Sample of .DAT file (with text)

```
þFIRSTBATESþ¶þLASTBATESþ¶þIMAGEIDþ¶þCUSTODIANþ¶þOCRTEXTþ
þPC00000001þ¶þPC00000002þ¶þIMG0000001þ¶þSmith, Johnþ¶þ*** IMG0000001 ***®®The world of
investing is fascinating and complex, and it can be very fruitful. But unlike the banking
world, where deposits are guaranteed by the federal government, stocks, bonds and other
securities can lose value. There are no guarantees. That's why investing is not a spectator
sport. By far the best way for investors to protect the money they put into the securities
markets is to do research and ask questions.®® *** IMG0000002 ***®®The laws and rules that
govern the securities industry in the United States derive from a simple and
straightforward concept: all investors, whether large institutions or private individuals,
should have access to certain basic facts about an investment prior to buying it, and so
long as they hold it. To achieve this, the SEC requires public companies to disclose
meaningful financial and other information to the public. This provides a common pool of
knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a
particular security. Only through the steady flow of timely, comprehensive, and accurate
information can people make sound investment decisions.þ
þPC00000003þ¶þPC00000003þ¶þIMG0000003þ¶þSmith, Johnþ¶þ***IMG0000003 ***®®The result of this
information flow is a far more active, efficient, and transparent capital market that
facilitates the capital formation so important to our nation's economy.þ
þPC00000004þ¶þPC00000005þ¶þIMG0000004þ¶þSmith, Johnþ¶þ *** IMG0000004 ***®®To insure that
this objective is always being met, the SEC continually works with all major market
participants, including especially the investors in our securities markets, to listen to
their concerns and to learn from their experience.®® *** IMG0000005 ***®®The SEC oversees
the key participants in the securities world, including securities exchanges, securities
brokers and dealers, investment advisors, and mutual funds. Here the SEC is concerned
primarily with promoting the disclosure of important market-related information,
maintaining fair dealing, and protecting against fraud.þ
```

The text and metadata of Email and the attachments, and native file document collections should be extracted and provided in a .DAT file using the field definition and formatting described below:

Field Name	Sample Data	Description
FIRSTBATES	EDC0000001	First Bates number of native file document/email
LASTBATES	EDC0000001	Last Bates number of native file document/email **The LASTBATES field should be populated for single page documents/emails.
ATTACHRANGE	EDC0000001 - EDC0000015	Bates number of the first page of the parent document to the Bates number of the last page of the last attachment "child" document
BEGATTACH	EDC0000001	First Bates number of attachment range
ENDATTACH	EDC0000015	Last Bates number of attachment range
PARENT_BATES	EDC0000001	First Bates number of parent document/Email **This PARENT_BATES field should be populated in each record representing an attachment "child" document
CHILD_BATES	EDC0000002; EDC0000014	First Bates number of "child" attachment(s); can be more than one Bates number listed depending on the number of attachments **The CHILD_BATES field should be populated in each record representing a "parent" document
CUSTODIAN	Smith, John	Email: mailbox where the email resided Native: Individual from whom the document originated
FROM	John Smith	Email: Sender Native: Author(s) of document **semi-colon should be used to separate multiple entries
TO	Coffman, Janice; LeeW [mailto:LeeW@MSN.com]	Recipient(s) **semi-colon should be used to separate multiple entries
CC	Frank Thompson [mailto: frank_Thompson@cdt.com]	Carbon copy recipient(s) **semi-colon should be used to separate multiple entries
BCC	John Cain	Blind carbon copy recipient(s) **semi-colon should be used to separate multiple entries
SUBJECT	Board Meeting Minutes	Email: Subject line of the email Native: Title of document (if available)
DATE_SENT	10/12/2010	Email: Date the email was sent Native: (empty)
TIME_SENT	07:05 PM	Email: Time the email was sent Native: (empty) **This data must be a separate field and cannot be combined with the DATE_SENT field
LINK	D:\001\ EDC0000001.msg	Hyperlink to the email or native file document **The linked file must be named per the FIRSTBATES number
MIME_TYPE	MSG	The content type of an Email or native file document as identified/extracted from the header
FILE_EXTEN	MSG	The file type extension representing the Email or native file document; will vary depending on the email format
AUTHOR	John Smith	Email: (empty) Native: Author of the document
DATE_CREATED	10/10/2010	Email: (empty) Native: Date the document was created

TIME_CREATED	10:25 AM	Email: (empty) Native: Time the document was created **This data must be a separate field and cannot be combined with the DATE_CREATED field
DATE_MOD	10/12/2010	Email: (empty) Native: Date the document was last modified
TIME_MOD	07:00 PM	Email: (empty) Native: Time the document was last modified **This data must be a separate field and cannot be combined with the DATE_MOD field
DATE_ACCESSD	10/12/2010	Email: (empty) Native: Date the document was last accessed
TIME_ACCESSD	07:00 PM	Email: (empty) Native: Time the document was last accessed **This data must be a separate field and cannot be combined with the DATE_ACCESSD field
PRINTED_DATE	10/12/2010	Email: (empty) Native: Date the document was last printed
FILE_SIZE	5,952	Size of native file document/email in KB
PGCOUNT	1	Number of pages in native file document/email
PATH	J:\Shared\SmithJ\October Agenda.doc	Email: (empty) Native: Path where native file document was stored including original file name.
INTFILEPATH	Personal Folders\Deleted Items\Board Meeting Minutes.msg	Email: original location of email including original file name. Native: (empty)
INTMSGID	<000805c2c71b75977050cb8306d1@MSN>	Email: Unique Message ID Native: (empty)
MD5HASH	d131dd02c5e6eec4693d9a0698aff95c 2fcab58712467eab4004583eb 8fb7f89	MD5 Hash value of the document.
TEXT	From: Smith, John Sent: Tuesday, October 12, 2010 07:05 PM To: Coffman, Janice Subject: Board Meeting Minutes Janice; Attached is a copy of the September Board Meeting Minutes and the draft agenda for October. Please let me know if you have any questions. John Smith Assistant Director Information Technology Phone: (202) 555-1111 Fax: (202) 555-1112 Email: jsmith@xyz.com	Extracted text of the native file document/email

4. Text
Searchable text of the entire document must be provided for every record, at the document level.

a. Extracted text must be provided for all documents that originated in electronic format. The text files should include page breaks that correspond to the 'pagination' of the image files. Note: Any document in which text cannot be extracted must be OCR'd, particularly in the case of PDFs without embedded text.

b. OCR text must be provided for all documents that originated in hard copy format. A page marker should be placed at the beginning, or end, of each page of text, e.g. *** IMG0000001 *** whenever possible. The data surrounded by asterisks is the *Concordance*® ImageID .

Sample page markers with OCR text:

> *** IMG0000001 ***
>
> The world of investing is fascinating and complex, and it can be very fruitful. But unlike the banking world, where deposits are guaranteed by the federal government, stocks, bonds and other securities can lose value. There are no guarantees. That's why investing is not a spectator sport. By far the best way for investors to protect the money they put into the securities markets is to do research and ask questions.
>
> *** IMG0000002 ***
>
> The laws and rules that govern the securities industry in the United States derive from a simple and straightforward concept: all investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it, and so long as they hold it. To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a particular security. Only through the steady flow of timely, comprehensive, and accurate information can people make sound investment decisions.

c. For redacted documents, provide the full text for the redacted version.

d. Delivery
The text can be delivered two ways:
1) As multi-page ASCII text files with the files named the same as the ImageID field. Text files can be placed in a separate folder or included with the .TIF files. The number of files per folder should be limited to 500 files.
2) Included in the .DAT file.

5. Linked Native Files
Copies of original email and native file documents/attachments must be included for all electronic productions.
a. Native file documents must be named per the FIRSTBATES number.
b. The full path of the native file must be provided in the .DAT file for the LINK field.
c. The number of native files per folder should not exceed 500 files.

II. Audio Files
Audio files from telephone recording systems must be produced in a format that is playable using Microsoft Windows Media Player™. Additionally, the call information (metadata) related to each audio recording MUST be provided. The metadata file must be produced in a delimited text format. Field names must be included in the first row of the text file.

The metadata must include, at a minimum, the following fields:

1) Caller Name: Caller's name or account/identification number
2) Originating Number: Caller's phone number
3) Called Party Name: Called party's name
4) Terminating Number: Called party's phone number
5) Date: Date of call
6) Time: Time of call
7) Filename: Filename of audio file

III.   Video Files

Video files must be produced in a format that is playable using Microsoft Windows Media Player™.

IV.   Electronic Trade and Bank Records

When producing electronic trade and bank records, provide the files in one of the following formats:

1.   MS Excel spreadsheet with header information detailing the field structure. If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details.

2.   Delimited text file with header information detailing the field structure. The preferred delimiter is a vertical bar "|". If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details.

V.   Electronic Phone Records

When producing electronic phone records, provide the files in one of the following formats:

1.   MS Excel spreadsheet with header information detailing the field structure. If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details. Data must be formatted in its native format (i.e. dates in a date format, numbers in an appropriate numerical format, and numbers with leading zeros as text).

2.   Delimited text file with header information detailing the field structure. The preferred delimiter is a vertical bar "|". If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details.

The metadata must include, at a minimum, the following fields in separate columns:

1)   Account Number: Caller's telephone account number
2)   Originating Number: Caller's phone number
3)   Terminating Number: Called party's phone number
4)   Connection Date: Date of call
5)   Connection Time: Start time of call
6)   End Time: End time of call
7)   Elapsed Time: Duration in minutes of the call

Each field of data must be loaded into a separate column. For example, Connection Date and Connection Time must be produced in separate columns and not combined into a single column containing both pieces of information. Any fields of data that are provided in addition to those listed here must also be loaded into separate columns.

VI.   Email Native File Production

When approved, Outlook (.PST) and Lotus Notes (.NSF) email files may be produced in native file format. A separate folder should be provided for each custodian.