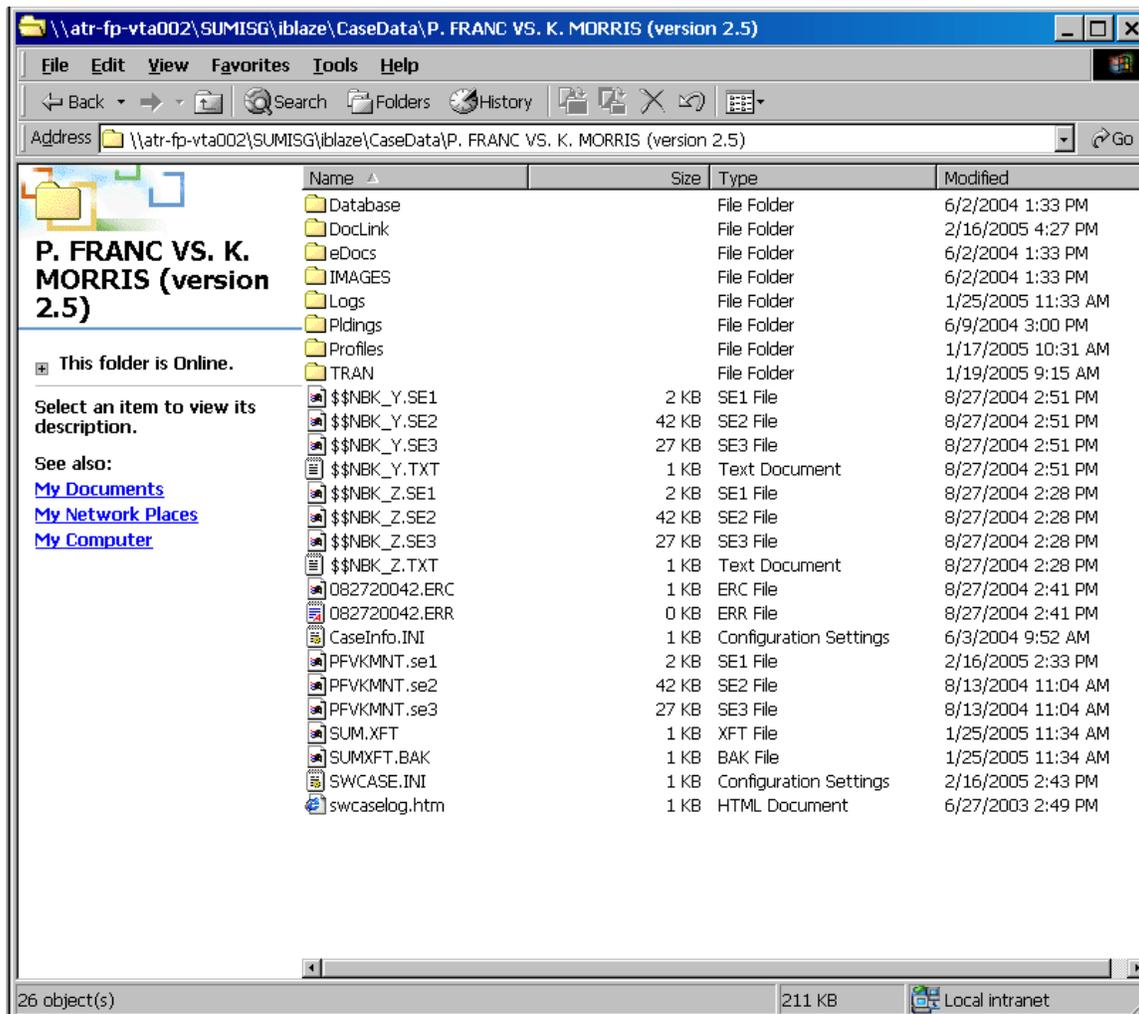### U.S. DEPARTMENT OF JUSTICE, ANTITRUST DIVISION
Summation Database Specifications

- DOJ to provide empty Summation database shell. The Division currently uses Summation version 2.6.3.

- DOJ will accept loaded Summation databases with the following conditions:
    - Each database has no more that 5-6 GB OCRBase.
    - If the database will have more than 120,000 records, it must be addressed with Division staff prior to production.
    - Custodians do not cross databases, except under limited circumstances.
    - Metadata fields must be populated
    - Data should be structured in the standard Summation format, with images stored inside the Images subdirectory of the case folder.
    - When records include a Doclink field to a native file, the native file should also reside inside the case folder in a folder called DocLink.

See Example of Case Directory Structure, below.

U.S. DEPARTMENT OF JUSTICE, ANTITRUST DIVISION
Summation Submission Requirements

Via e-mail or on CD-ROM, the DOJ has provided a DOJShell Summation database directory. Please categorize your submissions by placing the *DOJShell* folder under sequential Database folder names. The folder naming scheme should be 2 to 3 letters (indicating your company) followed by 3 numbers. For example:

For the first 3 databases from ABC Co., the root of the piece of media (External, CD, or DVD) should display the following folders: ABC001, ABC002, and ABC003. Each of these folders should contain the **loaded** DOJShell Summation Case Directory.

The cover letter for each submission of loaded Summation databases should include information about the loaded Summation database(s) included on each External Hard Drive or other piece of media submitted, preferably in spreadsheet format.

Include the following for each submission:

A. For each piece of media:

1. Assign a unique identifier for each piece of media that is also readily identifiable *on* the piece of media (i.e. Submission #; Serial number is also acceptable), and

2. Identify the Databases on the piece of media.

B. For each Database:

1. The Custodians included;

2. The total number of records;

3. The number of records for each Custodian (e.g., ABC001 contains 183,000 records: Jones - 150,000 records, Smith - 13,000 records, Doe - 20,000 records);

4. The Bates number ranges (and any gaps therein) for each Custodian;

5. The total number of native files in each Database;

6. The number of native files for each Custodian (e.g., ABC001 contains 15,980 native files: Jones - 1,500; Smith - 5,250; Doe - 9,230);

7. The total number of images included in each database; and

8. The total number of images included for each Custodian (e.g., ABC001 contains 15,980 images: Jones - 1,500; Smith - 5,250; Doe - 9,230).

U.S. DEPARTMENT OF JUSTICE, ANTITRUST DIVISION
METADATA & FAMILY RECORD SPECIFICATIONS

Default File Layout (.txt)	Field Name has 8 character limit.						
Field Name	**Field Description**	**Field Type**	**Hard Copy**	**E-mail**	**Spreadsheets**	**Presentations**	**Other Elec. Docs.**
Company	Company submitting data	Note Text	X	X	X	X	X
Box#	Submission/Volume #	Note Text	X	X	X	X	X
Custdian	Custodian(s)/Source(s) -- formatted Last, First	Multi-Entry	X	X	X	X	X
Begdoc#	Start Bates (including Prefix) - No spaces	Note Text	X	X	X	X	X
Enddoc#	End Bates (including Prefix) - No spaces	Note Text	X	X	X*	X	X
DocID	Populate with exact same value as Start Bates	Note Text	X	X	X	X	X
PgCount	Page count	Integer	X	X	X*	X	X
ParentID	Parent Bates, including Prefix (ONLY IN CHILD RECORDS)	Note Text	X	X	X	X	X
Attchids	Child document list - Start Bates of each Child (ONLY IN PARENT RECS)	Multi-Entry	X	X	X	X	X
FamlyRng	Family Start and End Bates (including Prefix) (i.e. ABC-001 - ABC-003	Note Text	X	X	X	X	X
Prprties	Record Type -> (File, E-mail, Attachment, Hard Copy); Privilege Notations -> (Redacted, Document Withheld Based On Privilege)	Multi-Entry	X	X	X	X	X
From	Author -- formatted Last, First	Multi-Entry		X	X	X	X
To	Recipient -- formatted Last, First	Multi-Entry		X	X	X	X
Cc	Cc field -- formatted Last, First	Multi-Entry		X	X	X	X
Bcc	Bcc field -- formatted Last, First	Multi-Entry		X	X	X	X
Subject	Subject / Document Title	Note Text		X	X	X	X
DocDate	Document Date / Date Sent - MM/DD/YYYY	Date Keyed		X			
Timesent	Time email was sent	Time		X			
Datecrtd	Date Created	Date			X	X	X
Datesvd	Date Modified	Date			X	X	X
Datercvd	Date Accessed / Received	Date		X	X	X	X
Filesize	File size	Note Text			X	X	X
Attitle	File name - Name of file as appeared in original location	Note Text			X	X	X
Applicat	Application used to create native file (e.g., Excel, Word)	Note Text			X	X	X
FilePath	Data's source filepath information	Note Text		X	X	X	X
Docllnk	Current filepath location to the native file	Note Text			X	X	X
FolderID	Email folder path (sample: Inbox\active) or Hard Copy Folder Information	Note Text	X	X			
Paragrph	Paragraph # to which the document is responsive	Note Text	X	X	X	X	X
Hash	Hash Value (used for deduplication or other processing	Note Text		X	X	X	X
Srchtrms	List of Terms used to identify record as responsive (if search terms used)	Multi-Entry			X	X	X

*Indicates field may be empty if only native files produced

Parent IDs, Attachment IDs, and Family Range Details			
Customer Notes:			
Y	Confirm Family Range definition for attached files		
Y	Confirm Field names and types.		
Y	Each member of the Family is its own record		
Family Range Definition:			
	All records will have a family range when the file or email has a parent or children		
	Family Range will start with the first page of the top most parent and go until the last child's last page		
Example:			
Description:	Top most Email	Attachment to Doc1	Attachment to Doc1
	Doc No. 1	Doc No. 2	Doc No. 3
Begin Bates	ABC-001	ABC-011	ABC-016
End Bates	ABC-010	ABC-015	ABC-020
ParentID	{empty}	ABC-001	ABC-001
Attchids	ABC-011; ABC-015	{empty}	{empty}
Family Range	ABC-001 - ABC-020	ABC-001 - ABC-020	ABC-001 - ABC-020