Appendix B

Guidelines for Requesting Data from Data Sources

The following guidelines save time and improve the chances for successfully obtaining and testing computer data.

I. Plan for the Request

Before requesting data from external systems or sources, you should have the following:

1. A basic understanding of the source system, including the purpose of the system, who uses the system, what data elements (or fields) are available, what reports are routinely generated, and what the data is used for.
2. An audit plan for reviewing or testing the data, including why you are testing the data, who will be testing it, and what other files will be required.
3. The contact information of: 1) the person responsible for maintaining the system; and 2) the person responsible for creating the data package in response to your request.

To help understand the data in an external system or source and identify exactly what data elements (fields) you will need for testing, you must obtain and review the appropriate **DATA DICTIONARY**. The dictionary should provide information such as the name, source, purpose, and a narrative explanation of each data element in the system. For example a Status field might have different values such as 1, 2, 3, and 4. You will need the Data Dictionary to understand the meaning of those values.

II. Request the Data in Writing

Once you have the above information, you are ready to make your data request. The request letter should include the name of the data elements requested as they are identified in the data dictionary. Request only those data elements that are relevant to your audit test. Depending on the size of the file requested, you might want to get all the fields in order to avoid possible omission and having to request the data again.

Your request letter should include:

- The date by which you need to have the data;
- The name and contact information of a person to contact if there are any questions regarding the request;
- A list of data parameters, such as specific transaction codes or a cut-off date for the data;
- The format in which you want the data; for example, MS-Access, MS-Excel, dBase, flat ASCII or EBCDIC files, ASCII Delimited;
- The media on which the data is to be put; such as, disk, tape, download, etc.
- The name and the contact information of the auditor requesting the data; and
- The name and address to which the data should be sent.

It is very important that the client provide, in writing, the total number of records in the database and the dollar amount (control totals) for all important numeric fields. This will allow you to reconcile the data and ensure that you have the right file and you imported the file correctly.

The section titled Technical Specifications for Computer Data contains lists of technical specifications and documentation requirements that the client should use when providing the data to you. You should provide a copy of the checklist to the client and request that they complete the list and forward it to you with the data. Failure to include these specifications may cause a delay in processing the data request.

III. Avoid Potential Problems

To reduce the probability of delays in processing your data request, you should be aware of the following general rules.

1. Many electronic data formats can be imported into IDEA. However, there are a wide variety of possible formats. Some formats can be troublesome. Microsoft Access or .dbf files (dBase format) files are the easiest file formats to import into IDEA. Spreadsheet files from Excel are easy to import, but Excel may handle certain exported data in unpredictable ways and can be quirky when handling date formats.
2. Print files are usually a copy of data listed on hard copy reports that is stored as a computer file. Accordingly, they often contain data such as headers, footers and subtotals that are shown on reports. If you do request a print file, you should also request some pages of the hard copy report. Also remember that the data in the report file has already been processed. Your test of the original data may not be complete if you limit yourself to just report files.
3. For ASCII and EBCDIC file formats, request fixed length files. Fixed length means that each record in the file has the same number of characters. If the client cannot provide fixed length files, you may have to perform additional steps to import the data into IDEA.
4. Verify that the client provided the required documentation. Incomplete documentation is often the cause of problems in processing computer data. Accordingly, it is recommended that you verify that the client has provided all the needed documentation and that the data is in the format you requested. If it is not, you should immediately contact the person responsible for providing the data.

Technical Specifications for Computer Data

1. Storage Medium:

	Storage Medium	Notes
☐	CD-ROM	Limited to 700Mb – More if the file is compressed*
☐	DVD	Limited to 4.7 Gb – Files can be compressed* You must have a DVD reader
☐	FTP Server	Can be slow depending on your internet connection Limited to the size of the hard drive of the server Requires an internet connection Security might be a concern – encrypting the data prior to uploading is recommended**
☐	Network Server	Can created network traffic depending on the size of the file Requires that the client allow you to access their network Limited to the size of the hard drive of the server
☐	USB Flash Drive	Limited to the size of the flash drive - which varies.
☐	Cloud Storage (DropBox, OneDrive, Google Drive, etc.)	Can be slow depending on your internet connection Size limitations depend on the particular Cloud Storage vendor Requires an internet connection Security might be a concern - encrypting the data prior to uploading is recommended**
☐	Other (please explain):	
*: You can compress file by using Winzip or another compression tool. **: You can use Winzip with a password or use other encryption tool.		

2. File Format (in order of preference):

	File Format	Notes
☐	MS Access	You might have to ask which table to import since Access can contains many tables
☐	Dbase	
☐	ASCII	All records must have equal length
☐	EBCDIC	All records must have equal length
☐	ASCII Delimited	Also known as Comma Separated You must know the Field Separator You must know the String Encapsulator
☐	MS Excel	MS Excel is easy to import but the file has to comply with some rules: • The data to be imported is in table format, such as a block of rows and columns where the first row of the table should be field names • All numeric columns are formatted with the required fixed number of decimal places • Any columns containing both text and numeric data are formatted as Text • All date fields are formatted as Date type • Blank lines, sub-total lines, and total lines are removed from the data area • The above steps are performed for all tables of data within the worksheet/spreadsheet
☐	Print Report File	
☐	ASCII Variable length*	All records must have equal length
☐	EBCDIC Variable length*	All records must have equal length
☐	Variable Length File	Field Separator _____ Record Delimiter(s) _____ String Encapsulator _____

	File Format	Notes
☐	Other (please explain):	
*: File Type: (For ASCII or EBCDIC Files)		

3. Required Documentation (for all types):

	For All Types
☐	Name and contact information of person(s) responsible for creating and providing the file.
☐	Total number of records in file.
☐	Control totals for important numeric fields.
☐	The date range expected in the file (earliest and latest date)

	For ASCII and EBCDIC	Notes
☐	Record layout that includes:	The beginning and ending position of each data element in the system. Each data element's width. Each data element's type, such as character, numeric with sign embedded, or alphanumeric, etc.
☐	A printout of the first records (to help you creating the record definition)	