



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

Data Collection, File Types and Costs Checklists	
1. Electronic Data Collection Checklist	
Types of Data Stores	Places to Search for Relevant Electronic Data
Shared Servers/ or Network	All network nodes
Computers	Desktop(s) Laptop(s)/ Notebooks At work and at home
Handheld devices and PDA (Personal Digital Assistant)	TREO Blackberries Mobile Phones iPods or MP3 players
Memory Devices	Thumb Drives Memory Stick Memory cards (from digital cameras, phones, other devices)
Removable Media	DVDs CDs Floppy drives Tapes Other removable media cards Detached hard drives Zip drives Security Camera tapes
Archive or Backup media	Backup Tapes Storage Locations: <ul style="list-style-type: none"> - ~ File shares - ~email devices - ~archival tapes - ~hosted emails, - ~ Attachments Archives Document Managements systems
Databases	Credit card/debit card Databases VOIP/Phone logs IM databases



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	Business application databases (Quicken, Calendar, Address Book, etc.) Other Data Entry systems
Online	Audit Logs Access Logs Web Pages Blogs Printer memory caches Copier/Scanner memory caches
Deleted/Unused space	Slack Space Deleted/Recovered Files
Other	Desk Drawers File Cabinets Closets Warehouses Obsolete Computer Equipment 3 rd Parties Partners/Licensors Escrow placeholders Be creative in your thinking here



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

2. File Type Checklist

Overview of File Type	Probable Extensions	
Documents	.doc .pdf .pps .ppt	.rtf .txt .wpd .xls
Images	.gif .jpeg .jpg .png	
Web Pages	.htm .html .mht .mhtml .shtml	
Media	.adf .avi .mov .mp3 .mpeg .mpg .ra	.ram .rm .swf .wav .wma .wmv
Databases	.mdb	
Applications	.qdf	
Others	www.filext.com	



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

Specific File Type	Summary	Typical File Extensions
ARCHIVE CONTAINER FILES		
ARCHIVE	Unknown compressed archive. This file type can contain one file or many files in multiple directories.	varies
ARCHIVE.ACE	A compressed archive. This file type can contain one file or many files in multiple directories. ACE is a proprietary data compression archive file format with superior compression compared to the ZIP file format , but at the cost of a lower compression speed.	.ace
ARCHIVE.ARJ	A compressed archive. This file type can contain one file or many files in multiple directories. ARJ is an archiving program created by Robert Jung for IBM-compatible computers. The letters stand for "Archive Robert Jung."	.arj
ARCHIVE.BZ2	A compressed archive. This file type can contain one file or many files in multiple directories. Bzip 2 UNIX Compressed File	.bz2
ARCHIVE.CAB	A compressed archive. This file type can contain only one file or many files in multiple directories.	.cab



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

ARCHIVE.EXE.ZIP	A self extractable zip archive or any executable that may contain zip archive.	.exe
ARCHIVE.GZ	A compressed archive which is generated using the GNU zip program. This program is used to compress the size of a single file. If more than one file has to be compressed, they must be tarred together first.	.gz
ARCHIVE.ISH	A compressed archive. This is not very commonly used.	.ish
ARCHIVE.ISO	An optical disk image or a Universal Disk Format (UDF) image. It contains all the data files and the filesystem metadata in a single file.	.iso
ARCHIVE.JAR	A compressed archive. JAR (for "Java ARchive") is a file format that bundles multiple files into a single archive file. Typically a JAR file contains the class files and auxiliary resources associated with applets and applications.	.jar
ARCHIVE.LZH	A compressed archive which uses the Lempel-Ziv and Haruyasu compression algorithm. It is used mostly in Japan.	.lzh
ARCHIVE.RAR	A compressed archive. RAR is a compressed file like ZIP, but uses a different encoding format	.rar
ARCHIVE.TAR	An archive. This file type can contain only one file or many files in multiple directories. Tar (for "Tape ARchive") is a Unix shell command that creates a single file called an " archive " from a number of specified files or extracts the files from such an archive. A tar archive has the file suffix ".tar". The files in a tar archive are not compressed,	.tar .tar.gz .tar.z .z



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	just gathered together in one file.	
ARCHIVE.UUE	<p>A compressed archive.</p> <p>This file type can contain only one file or many files in multiple directories.</p> <p>Compressed ASCII file archive created by UUENCODE (uuexe515.exe)</p>	.uue
ARCHIVE.Z	<p>A compressed archive.</p> <p>This file type can contain only one file or many files in multiple directories.</p> <p>UNIX Compressed Archive File</p>	.Z
ARCHIVE.ZIP	<p>A compressed archive.</p> <p>This file type can contain only one file or many files in multiple directories.</p> <p>Zip files compress data and therefore save time and space and make downloading software and transferring e-mail attachments faster</p>	.zip
EMAIL CONTAINER FILES		
EUDORA	<p>An email archive.</p> <p>May contain one or more email messages, calendar items, etc.</p>	.mbx
LOTUSNOTES	<p>A Lotus Notes database file.</p> <p>This type contains email folders, calendar items, tasks, custom forms, contact information, or other files that are stored in the database. Lotus Notes and Domino are a fundamental component of IBM Workplace.</p>	.nsf
MBOX	<p>An email archive.</p> <p>RFC-822 compliant format that may contain one or more email messages, calendar items, etc. This file type is a standardized</p>	<p>.mbx</p> <p>.eml</p>



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	email format.	Nescape Mozilla
OE	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	
OE.DBX	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	.dbx
OE.IDX	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	.idx
TEXT EXTRACTABLE FILES		
ASCII	A text file with pure ASCII text and no formatting.	.txt, .asc, .prn
CODE.C	C/C++ Program File	.cpp, .c, .h
CODE.JAVA	Java Source Code	.java
CODE.JS	JavaScript Source Code, This extension can be captured by a number of applications including: Windows Script Host, Dreamweaver MX, Notepad, Netscape Navigator, PavScrip, UltraEdit. The files are generally text files. Note: This file type can become infected and should be carefully scanned if someone sends you a file with this extension.	.js
CODE.SQL	Structured Query Language Data	.sql
CONCORDANCE	Single page tiff formats bounded by summation meta file documents.	.dat/.dii
EMAIL	Individual email messages or components (calendar items, tasks, etc.).	



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	Culled from Exchange or Notes.	
HTML	A Hypertext Markup Language file. Any text data can be encoded in HTML.	.htm .html
MSOFFICE	Desktop application files that are part of Microsoft Office other than Excel, Word or PowerPoint**	unknown
MSOFFICE.DOC	Individual desktop application files. Files that are identified as part of Microsoft Word family.	.doc
MSOFFICE.MPP	Microsoft Project	.mpp
MSOFFICE.PPT	Individual desktop application files. Files that are identified as part of Microsoft PowerPoint family.	.ppt
MSOFFICE.XLS	Individual desktop application files. Files that are identified as part of Microsoft Excel family.	.xls
PDF	An Adobe <u>Acrobat portable document format</u> . The PDF format has become a standard for document transfer between computer system architectures.	.pdf
RTF	A Rich Text Format File	.rtf
VSD	A Microsoft Visio file. A visual/graphics program that helps in creating business and technical diagrams.	.vsd
WPD	A <u>WordPerfect document</u> . WordPerfect was once a substantial word processing application.	.wpd
XML	An Extensible Markup Language File. Any text data can be encoded in XML.	.xml



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

IMAGE FILES		
IMAGE.BMP	A Windows OS/2 bitmap graphic. Common image file type.	.bmp
IMAGE.GIF	A Graphic Interchange Format. Common image file type.	.gif
IMAGE.JPG	JPEG/JIFF Image. Common image file type.	.jpg .jpeg .jif
IMAGE.PCX	A PC Paintbrush bitmap graphic. Common image file type.	.pcx
IMAGE.PNG	PNG (for Portable Network Graphics) is a file format for image compression	.png
IMAGE.TIF	A Tagged Image Format File (tiff). TIFF images are a common format when converting electronic/paper materials into a standardized storage or review mechanism.	.tif
AUDIO FILES		
AUDIO	<p>An audio file format is a container format for storing audio data on a computer system.</p> <p>The general approach towards storing digital audio formats is to sample the audio voltage in regular. Therefore sample rate, resolution and number of channels are key parameters in audio file formats.</p>	.669,.6cm,.8cm,.8med,.8svx,.ahx, .aifc,.aiff,.al,.alaw,.amr,.ape,.au,.avr, .cda,.cel,.cmf,.cwp,.d00,.d01,.dss,.far, .flac,.iff,.it,.kar,.kin,.ktp,.m3u,.m4a,.med, .mmf,.mod,.mp2,.mpc,.mpu,.mp_, .mtm,.nap,.ogg,.pca,.pcm,.pvc,.rmf, .rmi,.rmx,.s3m,.saf,.sd2,.sdx,.sf, .sfk,.sib,.smf,.snd,.sng,.svx, .sw,.swa,.tfmx,.thx,.tsp,.uax,.ub,.vmf,.voc, .voi,.vox,.wave,.wax,



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

		.wrk,.xm, .xmiax,.wrk, .xm,.xmi
AUDIO.AAC	Highly compressed audio file, similar to MP3, that maintains quality nearly indistinguishable from the original audio source	.aac
AUDIO.AIF	CD-quality audio file, similar to a .wav file	.aif
AUDIO.MID	Musical Instrument Digital Interface standard file; contains music data such as what notes are played, when they are played, how hard the key is pressed, how long it is held, etc; does not contain actual audio	.mid
AUDIO.MIDI	Musical Instrument Digital Interface file; contains music data such as what notes are played, when they are played, how hard the key is pressed, how long it is held, etc; does not contain actual audio	.midi
AUDIO.MP3	Compressed audio format developed by the Moving Picture Experts Group; near-CD quality sound (stereo, 16-bit)	.mp3
AUDIO.MPA	Compressed audio file using MPEG Layer I, II or III compression	.mpa
AUDIO.RA	Audio file created by RealPlayer	.ra
AUDIO.RAM	Audio file created by RealPlayer	.ram
AUDIO.WAV	CD-quality audio file, similar to AIFF	.wav
AUDIO.WMA	Audio file compressed with Windows Media compression	.wma
VIDEO FILES		
VIDEO	An video file format is a container format for storing video data on a computer system.	.3g2,.3gp, .3gp2,.3gpp, .3mm,.avs,.by, .dir,.dv, .dvrms,.dxr, .fla,.flc,.fli,.flv, .flx,.gl,.grasp, .ifo,.ivf,.ivs, .lsx,.m1v, .m4e,.m4u, .mkv,.moov, .movie,.mpe, .mpeg,.mpv2, .mvc,.omf,.prx,.qtch,.rmvb,



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

		.rp,.rts,.scm, .smil,.smv,.spl, .svi,.vfw,.vid, .viewlet,.viv, .vivo,.vob, .vro,.wm, .wmd,.wmx, .wvx
VIDEO.ASF	Media file developed by Microsoft; can be compressed using a variety of codecs	.asf
VIDEO.ASX	Shortcut to an audio/video file (typically in .ASF format); references a URL for the location of the actual media	.asx
VIDEO.AVI	High-quality video format; uses less compression than similar formats such as .mpeg and .mov	.avi
VIDEO.MOV	Common multimedia format (mainly used for video); cross-platform, but developed by Apple	.mov
VIDEO.MP4	Video clip or movie that uses MPEG-4 compression	.mp4
OTHERS		
CLASS	Java Bytecode	.class
DLL	A dynamic link library (DLL) is a collection of small programs, any of which can be called when needed by a larger program that is running in the computer. The small program that lets the larger program communicate with a specific device such as a printer or scanner is often packaged as a DLL program (usually referred to as a DLL file).	.dll
EXE	A program file. Virtually all programs that run under Windows or DOS are in the .EXE format.	.exe
KNOWN_OS	File belongs to a collection of digital signatures of known, traceable software applications. For reference see http://www.nsr1.nist.gov	Various



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

UNKNOWN	Unable to identify type of the file.	Various
XSL	XSL (Extensible Stylesheet Language), formerly called Extensible Style Language, is a language for creating a style sheet that describes how data sent over the Web using the Extensible Markup Language (XML) is to be presented to the user.	.xsl
MSACCESS	Microsoft Access Database	.mdb
	Microsoft Active Server Page	.asp
	Macintosh Binary File	.bin
	Cascading style sheet for web pages	.css
	Comma Separated Values; sometimes used as a spreadsheet format or for import/export of data	.csv
	dBASE	.dbf
	MS Word Template format	.dot
	Autocad Drawing	.dwg, .dxf
	Macintosh BinHex archive	.hqx
	A compressed file archive	.lha, .lhz
	Lotus WordPro	.lwp
	Microsoft Access queries and reports	.maq, .mar
	MacWrite	.mw
	Open Office Spreadsheet	.ods, .sxc
	Open Office Documents	.odt, .ott, .sxw



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	Aportis (Palm) document	.pdb
	Postscript files	.ps, .eps
	Microsoft Publisher	.pub
	Pocket Excel	.pxl
	StarWriter	.sdw
	Standard Generalized Markup Language	.sgml
	Corel Presentations format	.shw
	Macintosh Stuffit archive	.sit, .sea
	Netscape Mail	.snm
	Shockwave Flash	.swf
	Corel Quattro Pro spreadsheet format	.wb3
	Lotus 1-2-3	.wk1 or any#, .wks, .wku
	MS Works spreadsheet	.wks
	MS Works document	.wps
	MS Write	.wri
	MS Excel spreadsheet template	.xlt
	Yahoo Messenger	.ymg



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

3. Electronic Data Cost Consideration Checklist

Single vs. Multiple Vendors	<ul style="list-style-type: none"> - Using different vendors for various aspects of the projects might decrease the cost. - More efficient to use vendors that can integrate with in-house applications to reduce loading and hosting costs.
General Fees	<ul style="list-style-type: none"> - Set up Costs <ul style="list-style-type: none"> o Fixed fee o Hourly - Consider a cap on fees - User license fees <ul style="list-style-type: none"> o (month vs. months) - License subscription fees
Training Costs	<p>Live Training</p> <ul style="list-style-type: none"> • Day rates vs. hourly rates • Travel costs • Vendor's waiver for some of training sessions <p>Web based training</p> <ul style="list-style-type: none"> • Session fee vs. hourly fee <p>Free web-based training by vendor</p>
Bundling	<p>Identify what terminology vendor is using and what activities are included in price.</p> <ul style="list-style-type: none"> • Are services bundled together for one fee, or assessed separately? • What is included? Be Specific
Data Loading and Processing Costs	<p>Loading Costs</p> <ul style="list-style-type: none"> • Native • TIFF/PDF • Other file types- <p>Costs Assessed</p> <p>Other processing/loading issues</p> <p>Common for volume discounts.</p> <ul style="list-style-type: none"> • Negotiate reduced rates as volume increases
Hosting Costs	<p>Data hosting charged per volume vs. monthly basis.</p> <p>Negotiate</p>
Best Practice to Reduce Hosting Cost	<p>Reduce volume of data as much as possible</p> <ul style="list-style-type: none"> • Custodians of Interest • Date Range • Deduplication methods <ul style="list-style-type: none"> o Horizontal



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	<ul style="list-style-type: none"> ○ Vertical ○ Across data set Review sources of data prior to loading to exclude duplicate data
Hosting Shared Databases	Use vendor to hold unique copy of document
Costs of Archiving Data	Ability of vendor to archive data for a dormant case for a reduced hosting fee Cost for re-activating database
Costs of Production	Dependent on format Reduced rate to reproduce earlier production Additional Costs Paper production Costs (per page)
Professional Services Fees	Charged per hour. <ul style="list-style-type: none"> • Unpredictable costs • Dedicated Resources • Pre-pay for a bulk set of hours (bulk purchase for lower rate)
Cost Savings in Professional Service Fees	Other persons (in-house counsel, outside counsel, litigation support personnel or consultant) can perform certain services. Sliding fee/level of experience Blocks of hours for reduced fee Build in level of support, or hours of support, into loading charges.
Termination Costs	Costs for closing down database or transferring the data and work product to another vendor
Service Level Agreements	Service Level Agreements provide for certain level of service guaranteed by financial penalties.
Other Cost Considerations	See Sedona RFP document http://www.thesedonaconference.org/content/miscFiles/RFP_plus_july05ver.pdf



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

4. Other Checklists

Search Types	<p>Basic Search Types</p> <ul style="list-style-type: none"> • Phrase • Boolean • Proximity • Directed Proximity • Phonic • Stemming • Numeric • Macro • Wildcard • Fuzzy • Concept/Synonym/Thesaurus • Combining Search Types <p>Advanced Search/Information Retrieval Methods</p> <ul style="list-style-type: none"> • Ontologies • Clustering (Concept Searching) • Combination of these methodologies (Context searching)
Report Types	<ul style="list-style-type: none"> • Exception Reporting: Report what files and locations could not be indexed • Network Report: Report what locations are found, what are currently attached, what devices and locations are no longer attached • Search Report: Report the results of a query including metadata for found documents and the location of "hits" within the document. Reports can be generated in the background. • Location of Interest: Show all locations and files that have been selected as having documents of interest. Show the approximate size and contents of each location • Audit Report: List of all queries that were run and who ran them • Privilege Log – all Priv markings, made by whom, when, and why • Redaction Log – all redactions, made by whom, when, and why • Foreign Language Report – a list of all foreign languages found in the data set • Analytics Reports – various reports about the data and what is found in it from an analytical standpoint