Data Mapping: Data Types and Locations

This document sets out locations and types of data that are more likely to hold responsive data of evidential value. This is in contrast to the de-NISTing process which involves the removal of known, standard system files from collected data which are unlikely to hold any evidential value.

Whether or not a data type is likely to hold evidential value depends on the facts of your matter. The "unknown unknown" should also be considered; it may not be expected that some data types contain evidential value (e.g. graphics) but if there is a risk that they do consider whether they need to be collected and reviewed.

Consider whether you must follow any regulatory or court-based requirements on which data sources or locations to consider and report against. By way of example, the list below uses to a large extent (but not exactly) the same wording and order as that used in the Disclosure Review Document under the Disclosure Pilot Scheme under the Civil Procedure Rules in England & Wales.

Possible Data locations

(1) Document repositories and/or geographical locations

(2) Computer systems or electronic storage devices

(3) Mobile phones, tablets and other handheld devices

(4) Document management systems

(5) Email servers

(6) Network Servers

(7) Cloud based data storage (e.g. Google Drive, Dropbox, Microsoft OneDrive, Salesforce)

(8) Webmail accounts e.g. Gmail, Hotmail etc

(9) Back-up systems Archival systems (e.g. Mimecast, Splunk, Sumologic)

(10)Social media accounts

(11)Chat accounts

(12) Third parties who may have relevant documents (e.g. agents or advisers).

(13) Hard copy documents & files.

Questions which may be relevant to one or more of the above include:

(1) Which system and version is in use (e.g. Office 365)

(2) What make or models are in use (e.g. of the email servers, or of any devices);

(3) Where are the servers located;

(4) How long have the systems, servers, devices etc. been in use;

(5) To what extent is the data stored on the systems, servers, devices stored “live” as opposed to being accessible via back-up media;

(6) How is the data organised and structured and who has access rights (e.g. by department or personal);

(7) Are local copies permissible (e.g. PST files in Outlook, or local copies in a laptop) or are they always held on servers;

(8) Are there restrictions on how and when data is transferred (e.g. can data be copied or transferred to external devices such as USB keys, external hard drives or CDs/DVDs, or sent to personal email addresses);

(9) Who is responsible for each of the data types (e.g. who owns the data on the front-end versus who from IT ensures retention is maintained on the back-end);

(10)What are the policies for each of the data types;

(11)Are chat logs actively stored and backed-up by the organization;

(12) Are litigation holds in place and is IT aware of the obligations set out in the hold to ensure back-end preservation is taking place;

(13) When and how are back-ups taken of the data types (e.g. full back-ups versus monthly or incremental);

(14) Have you discussed with custodians about how the data sources were used? Does that information align with what else you know, for example, from the IT policies and those responsible for IT?

(15) Where hard copy documents are concerned, are these ancillary copies of documents that are otherwise duplicative of content that would have been disseminated via email (and otherwise be accessible via an email collection).

We are grateful to BDO and TCDI who provided assistance with identifying potential data types.

The following are common file extensions which are more likely to be of evidential value.

Electronic Documents[1]	Email	Containers	Forensic Containers (Possible Matches)	Graphics
accdb	dbx	7z	001	art
asd	edb	bkf	ad1	bmp
csv	eml	gz	e01	gif
doc	emlx	rar	ex01	jpe
docm	idx	tar	l01	jpeg
docx	imb	tgz	lx01	jpg
dotm	mbox	zip	s01	png
dotx	mbx		DD	tif
dw	msf			tiff

htm	msg			wmf	
html	nsf				
mdb	olk14MsgSource				
odm	olm				
ods	ost				
odt	pab				
one	pfc				
otp	pst				
ots	rge				
ott	wab				
pdf	olk15MsgSource				

pot					
potm					
potx					
ppam					
pps					
ppsm					
ppsx					
ppt					
pptm					
pptx					
pub					

rtf					
sdc					
sdw					
stc					
stw					
sxc					
sxg					
sxi					
sxw					
txt					
vor					

wg1					
wg2					
wg3					
wk1					
wk3					
wks					
wpd					
wps					
xlam					
xls					
xlsb					

xlsm					
xlsx					
xlt					
xltm					
xltx					
xlw					

[1] TCDI table courtesy of David York