A Digital Forensics Approach to Collecting Born Digital Archives at the British Library

Callum McKean, Digital Lead Curator,
Contemporary Archives and Manuscripts

BRITISH LIBRARY

Defining Digital Forensics

Digital forensics is a branch of forensic science that focuses on identifying, acquiring, processing, analysing, and reporting on data stored electronically.

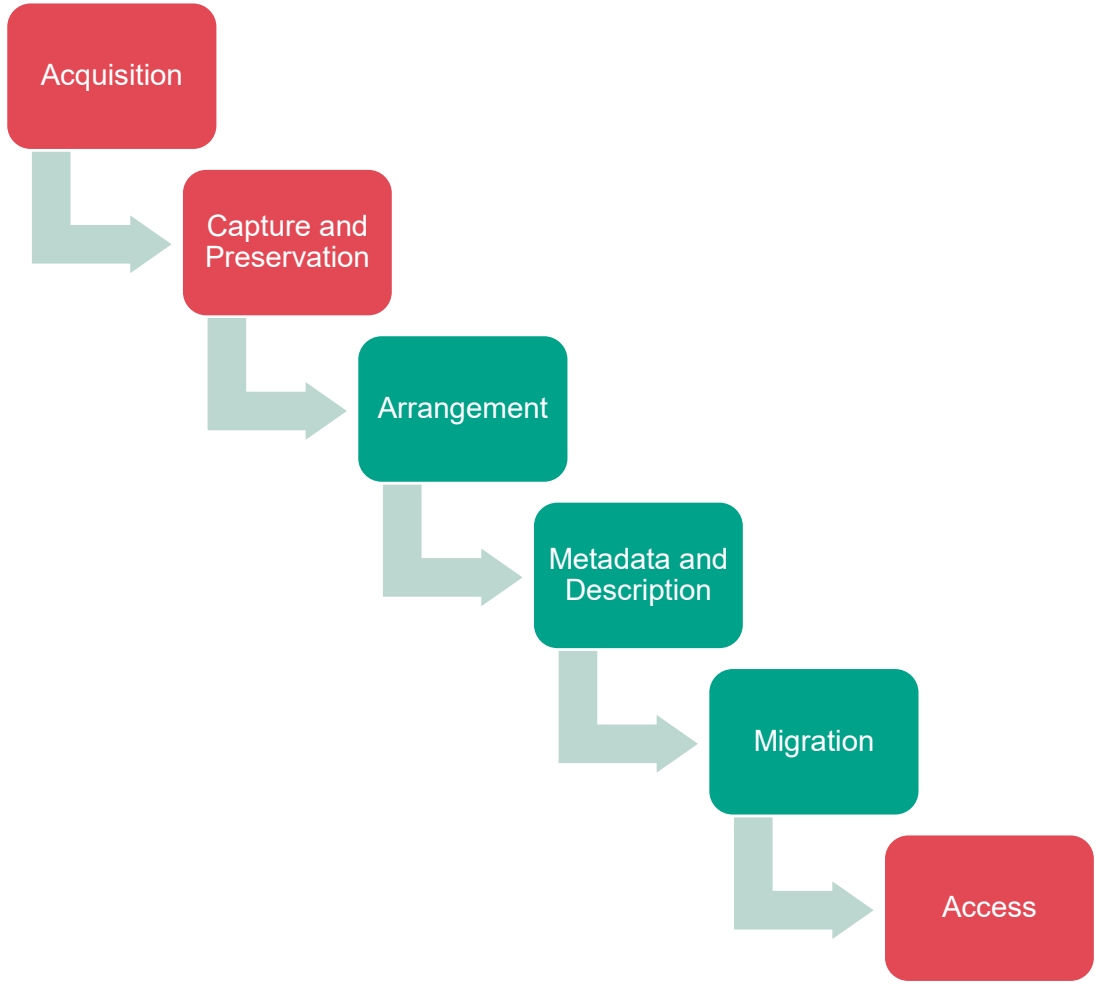
The main goal of digital forensics is to extract data from the electronic evidence, process it into actionable intelligence and present the findings for prosecution. All processes utilize sound forensic techniques to ensure the findings are admissible in court.

Interpol, 2024.

What do we gain?

- It allows us to store data independently of vulnerable carriers
- It allows us to make the most authentic copies possible
- It ensures that accurate metadata is retained
- It allows us to document and recreate our actions
- It does not foreclose future research possibilities.

How do we implement this approach?



Appraisal

- We conduct a rapid appraisal of hard-drives, USB sticks and computers using a SPEKTOR Ultra digital forensics machine.

- We inspect other forms of physical media with the Depositor, looking for labels which may give a clue to their content

- We ask Depositors to fill out a questionnaire in order to make a record of their hardware and software use

SPEKTOR Ultra Digital Forensics machine in Collector Mode for rapid appraisal

Amstrad Floppy Disks from the Will Self Archive (Add MS 89203) with handwritten labels

Capture and Preservation

- After acquisition, flash storage and computers are forensically captured as .e01 files using a SPEKTOR Ultra digital forensics machine

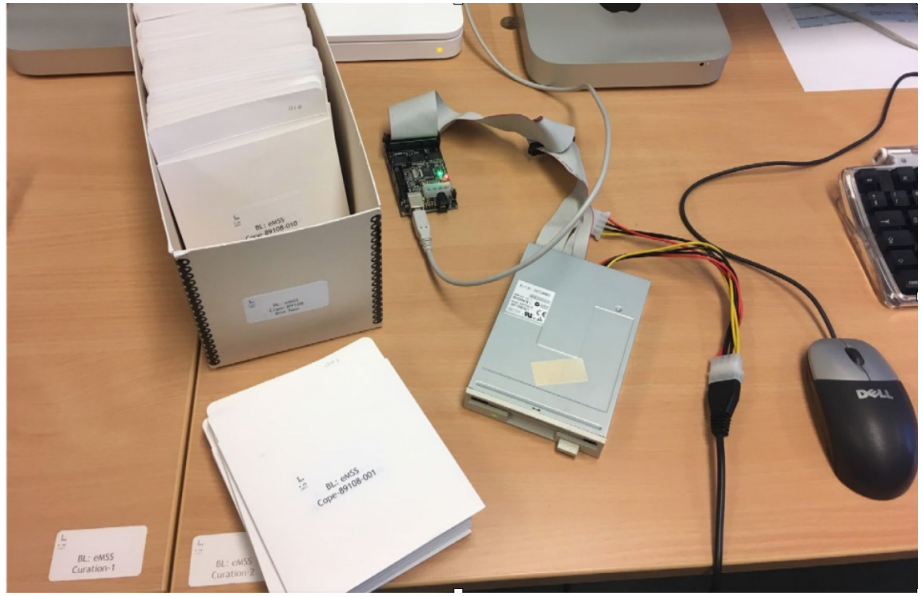
- Magnetic carriers (such as Floppy Disks) are captured at bit-level using Kryoflux

- For some carriers, we can not make forensic copies (zip disks, Amstrad disks etc.) For these we copy the files.

- These preservation copies are ingested into the Library's Minimum Preservation Tool (for backup and regular fixity checking)

- Carriers are photographed, labelled and moved to long-term storage.

BRITISH LIBRARY



Kryoflux creating bit-level captures of 3.5inch floppy disks from the Wendy Cope Archive (Add MS 89108)

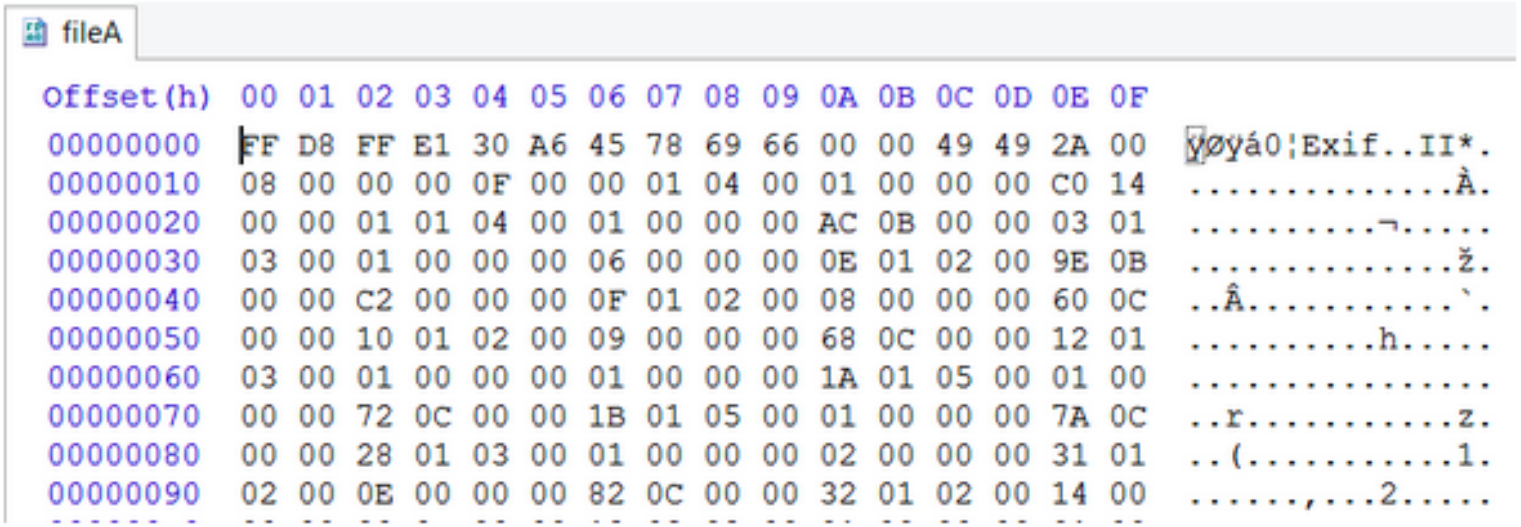
Intermedia and Windows XP machine used for non-forensic data extraction from other magnetic media

BRITISH LIBRARY

Why?

- We can not solve every problem immediately
- We do not know what research will look like in the future
- We do not know what the technological landscape will look like in the future
- We do not know what archival practice will look like in the future

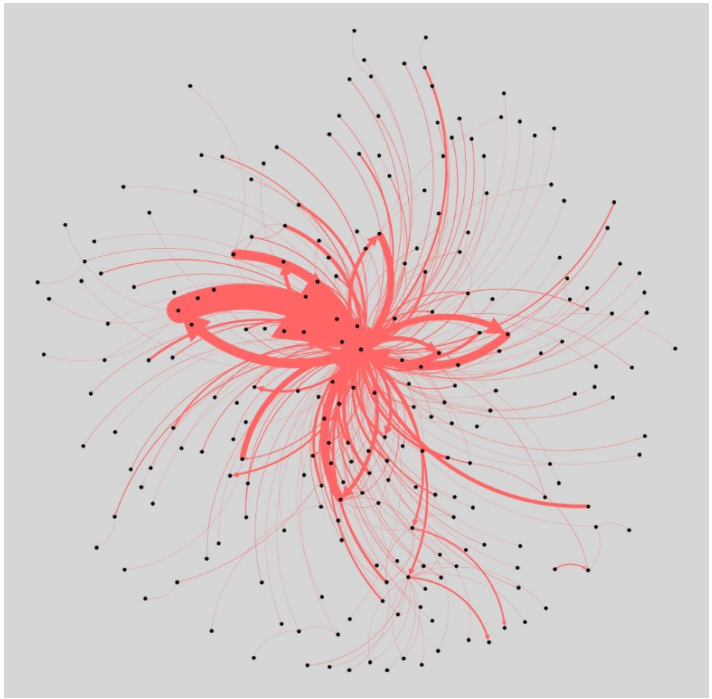
Access – Genetic Criticism



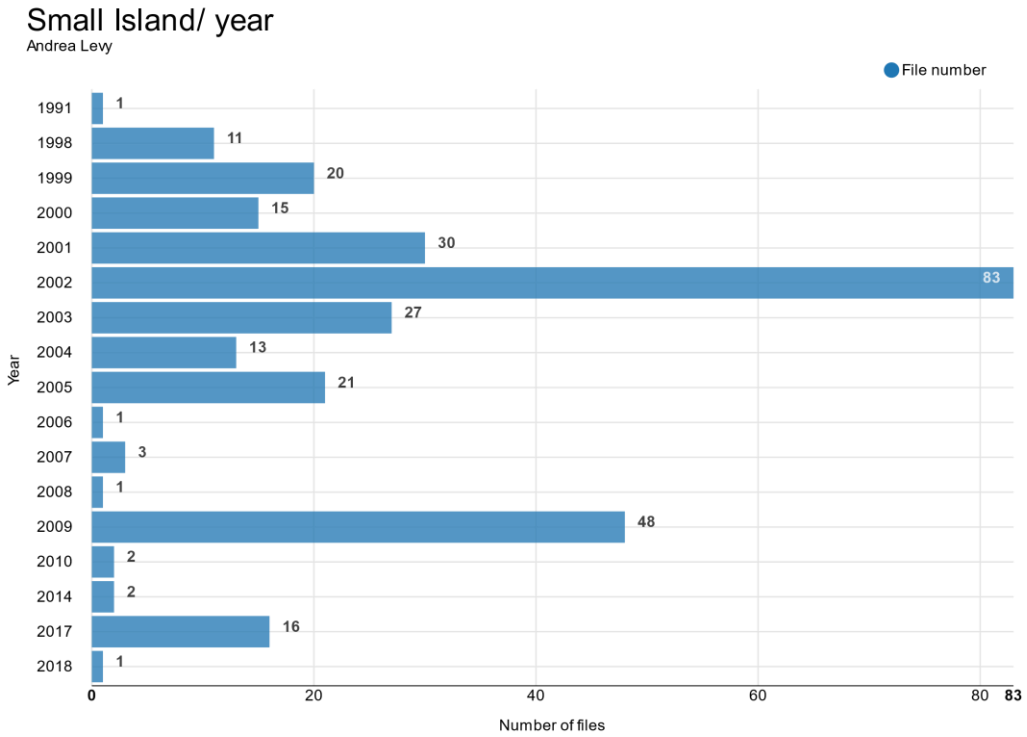
fileA

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	FF	D8	FF	E1	30	A6	45	78	69	66	00	00	49	49	2A	00	ÿØÿá0¦Exif..II*.
00000010	08	00	00	00	0F	00	00	01	04	00	01	00	00	00	C0	14	..............À.
00000020	00	00	01	01	04	00	01	00	00	00	AC	0B	00	00	03	01	..........¬.....
00000030	03	00	01	00	00	00	06	00	00	00	0E	01	02	00	9E	0B	..............ž.
00000040	00	00	C2	00	00	00	0F	01	02	00	08	00	00	00	60	0C	..Â..........`.
00000050	00	00	10	01	02	00	09	00	00	00	68	0C	00	00	12	01	..........h.....
00000060	03	00	01	00	00	00	01	00	00	00	1A	01	05	00	01	00	.....
00000070	00	00	72	0C	00	00	1B	01	05	00	01	00	00	00	7A	0C	..r..........z.
00000080	00	00	28	01	03	00	01	00	00	00	02	00	00	00	31	01	..(..........1.
00000090	02	00	0E	00	00	00	82	0C	00	00	32	01	02	00	14	00	......‚...2.....

Access – Data Analytics



Redacted network visualisation of the Harold Pinter e-mail archive created in Gephi as part of the project 'Data Analytics and Network Visualisation for Hybrid Correspondence Archives'. Latest re-usable code available via GitHub



Graph showing the time distribution of drafts of Andrea Levy's novel *Small Island* (2004) created as part of the Writers Digital Lives PhD Placement project by Ariel Li

Access - Emulation

Emulating the Windows 95 computing environment used by John Maynard Smith (Add MS 86569)

Thank you

BRITISH LIBRARY