



Digital forensics and Bodleian Electronic Archives & Manuscripts (BEAM)

Susan Thomas
Digital Archivist, Bodleian Library

It all started when...

- Barbara Castle's archive accrued digital media in the middle of being catalogued
- While the Paradigm project was running
- Jeremy and the Digital Scriptorium assist with hard disk imaging
- Adopting imaging for digital media
- Enter the futureArch project to improve our capture capacity and extend our forensics work to analysis



BEAM processing manual

Last modified: 12. May 2010 08:08 PM

Processing Manual for Bodleian Electronic Archives and Manuscripts (BEAM). 2009-

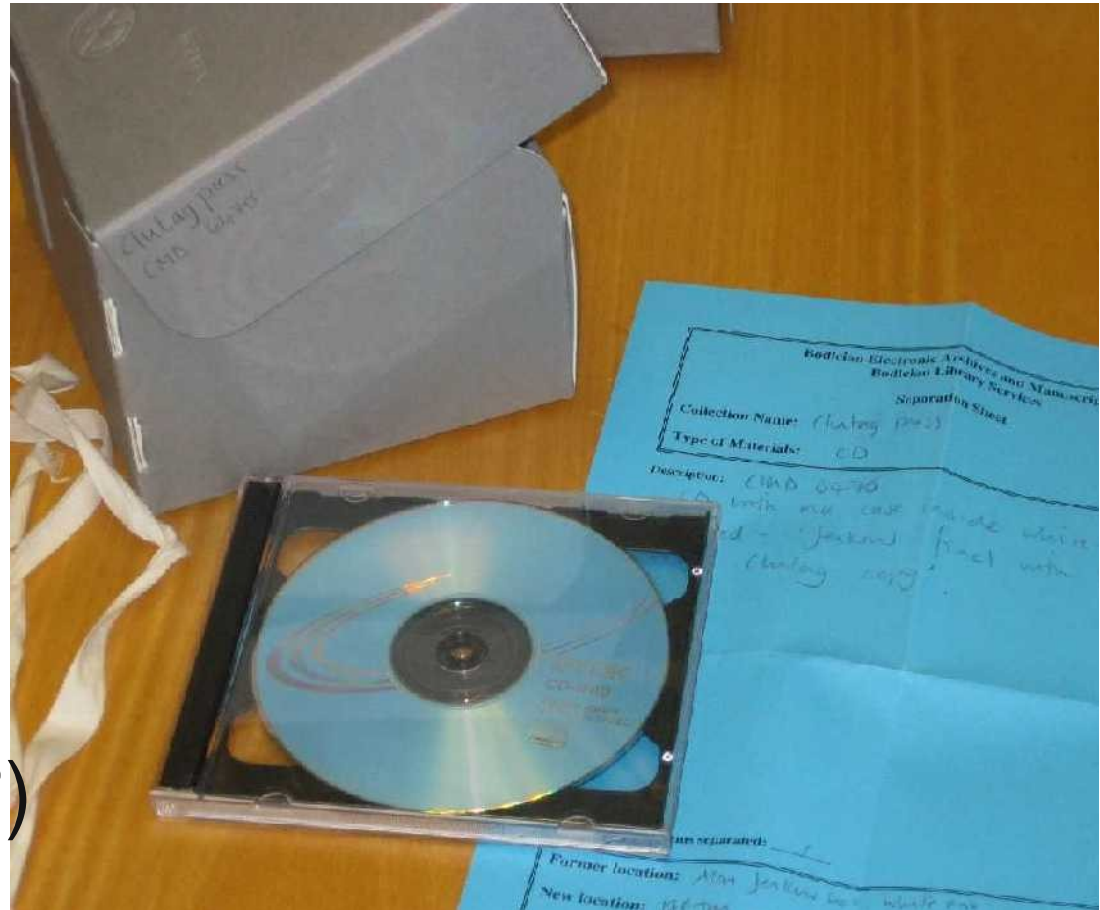
Table of Contents

Processing Manual for <u>Bodleian</u> Electronic Archives and Manuscripts (BEAM). 2009-.....	1
A) What to do when born-digital material forms part of a prospective accession.....	3
B) What to do when born-digital material is found in an accession.....	3
B1) When born-digital material is discovered on or after accession.....	3
B2) When born-digital material is discovered during/in advance of a <u>cataloguing</u> project.....	3
B3) Separating born-digital media from accessions.....	3
B4) Updating the Collections Management Database.....	4
C) On arrival at BEAM.....	5
C1) Update the Collections Management Database record.....	5
C2) Create a record in the BEAM accessions register.....	5
C3) Re-box and label material.....	5
D) Media Recognition Guide.....	6
E) Inventory and photo-documentation of media in individual collections or accessions.....	7
E1) Numbering.....	7
E2) Photographing.....	7
E3) Uploading Photos.....	8
F) Imaging software.....	9
F1) Access data FTK <u>Imager</u>	9
F2) <u>Infinadyne</u> CD/DVD Inspector.....	9
F3) Linux <u>dd</u>	9
G) Imaging media.....	10
G1) Hard disks.....	10
G2) <u>CDs and DVDs</u>	10
Imaging Introduction.....	10
Imaging Process for CD/DVD Inspector.....	10
File Systems.....	14
Post-Imaging.....	15
Imaging Optical Media With FTK <u>Imager</u>	15
G3) 3.5" floppy disks.....	17
G4) 5.25" floppy disks.....	17
G5) Zip disks.....	17
G6) 3" <u>Amstrad</u> disks.....	17
H) Carving Out Files.....	18
H1) Removable media.....	18
H2) Hard drives.....	19
I) BEAM Ingest.....	20
J) Analysis and appraisal.....	21
K) Preparation of use copies for <u>cataloguing</u> archivists.....	22
L) Arrangement and description processes.....	23
M) Publication process.....	24



Treating the media

- Record in CMD
- Separate/transfer to BEAM
- Repackage/label
- ID & inventory
- Photograph
- Retain (all? How long?)
- Cataloguing archivist does not have access to original media



BEAM processing room



Machines isolated from wider network; venue for capture, appraisal, cataloguing, access review; multiple users; multiple platforms

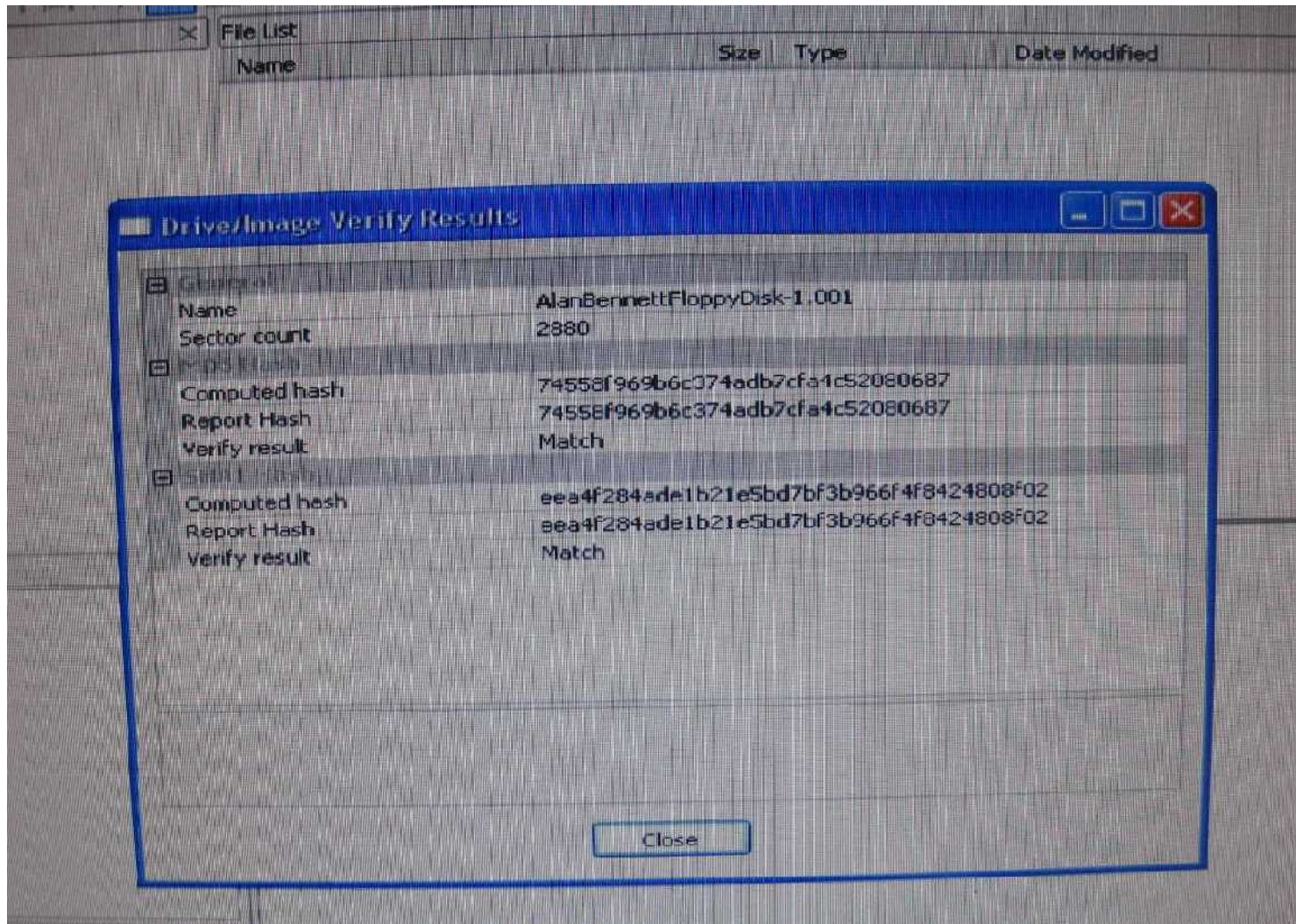


Imaging with FRED



- Used for most capture
- Supports wide range of media
- Most of our holdings are 3.5" floppies and optical media
- 5 HDDs to-date
- Use FTK Imager

FTK Imager



Imaging at BEAM (also used on-site with donors)

Mass imaging optical media



- Some colls with hundreds of optical media
- Not as straightforward as it seems.
- Used in conjunction with CD/DVD Inspector software

Imaging issue log

Issue ID
Nature of issue
Affected material
Severity of issue
Date identified
Identified by
Proposed resolution
Individual/organisation or process passed to for resolution
Expected resolution date
Actual resolution date
Manner of resolution
Does issue suggest change of procedure is required? If yes provide details

<u>Severity scale</u>	
1	Catastrophic. Error causing irrevocable loss of collection materials and/or damage to BEAM hardware or software.
2	Severe. A road block with no workaround. Performance is so poor that the system is universally regarded as 'pitiful'.
3	Moderate. No permanent loss of data, but wasted time. There is a workaround to the problem.
4	Minor but irritating problem. Slows work down slightly.
5	Minimal error.

Metadata for media/image

Digital media (physical)

Record ID

Accession ID (from CMD)

Date record created

Date record modified

Collection name

Disk number

Case or sleeve label

Physical label

Media type

Brand

Type

Capacity

Capacity used

Virtual disk label (as shown by the operating system once mounted)

file system

serial/batch number

Operating system requirements

Comments

Associated photos

Related disk image metadata

Disk image type

Imaging software used

Imaging hardware configuration used

Disk image filename/s

Disk image number of parts

Disk image MDS hash

Date imaged

Date ingested

Imaging comments

Disk contents

Covering dates

Scope and content



Legacy hardware/software



Guides – media/filesystems

Media Recognition

It is important to be able to recognise the media you come across in collections. Determining what type of data storage an object is can be easy, since floppy disks look very different from a CD. However, distinguishing between different types of floppy disk or CD can be more challenging. This guide aims to help you identify object characteristics, thereby allowing you to ascertain the specific type of data storage device in your possession.

1. **Floppy Disks:**

- 8 inch disks
- 3 inch disks
- 5.25 inch disks
- 3.5 inch disks

2. **Optical Media:**

- Compact Disk
- Audio and Data CD Comparison
- CD-ROM
- CD-R
- CD-RW
- DVD
- HD DVD
- Blu-Ray

3. **Hard Disk Drives:**

- AT Attachment Interface
- Serial ATA Interface
- SCSI Interface
- Serial Attached SCSI Interface
- External Hard Disk Drives

4. Iomega Zip Drive

5. **Flash Media:**

- USB flash drive
- FireWire flash drive

File Systems

1. **Media Types**
2. **File Systems**
3. **File System Extensions**
4. **File Systems in Image Files**
5. **File System Tables**

File systems allocate space on media and organise the data in an easily identifiable and accessible way. The file system, or systems used in different media is determined by the type of media, its content and the platform required. Whilst each file system has different abilities and restrictions some systems have extensions which expand their capabilities in specific ways.

1. Media Types

Hard Drives

Hard drives require a file format that can support the constant rewriting and updating of files at incremental times. The File Allocation Table (FAT) was originally the primary file system for several operating systems including MS-DOS and Microsoft Windows. Moreover, FAT is compatible with virtually all operating systems, allowing data to be easily transferred between platforms. However, FAT has to some extent been superseded by the New Technology File System (NTFS), particularly on later Windows platforms from Windows 2000 onwards. This is primarily due to the FAT file size limit of 4 GB and NTFS's greater reliability. However, NTFS is not as universal and while compatible with a number of operating systems, this does not include Mac OS.

Optical Media

The different types of optical media can be separated into two categories: write-once and re-writable media. Re-writable optical media is capable of being supported by FATs and NTFSes, but unlike files on hard drives, optical media files are read-only. Also, there are alternative file systems specifically designed for optical media, therefore FATs and NTFSes are not the standard file systems.

The standard system for DVDs is Universal Disk Format (UDF). This system supports re-writable media, therefore can also be used for CD-RW. Some file systems have size restrictions, but UDF can support files larger than 4 GB, therefore is suited to DVD media, which tends to contain large files. However, write-once DVDs with smaller files can write ISO9660 along with UDF.

The CD-ROM standard is currently ISO9660, although this is intended to change to the latest ISO version, ISO13490.

It is only optical media containing data files that has file systems; audio disks do not, they have tracks and track data instead. Originally the technology of the time was

Blog and document



John Barton



Recent stack discovery - Professor spec. in Roman Law at Oxford



Appraisal - Alan Bennett

The screenshot shows the AccessData Forensic Toolkit interface. The top window displays the 'File Content' of a selected file, showing the text: 'SATURDAY 3 JUNE 1972' followed by a paragraph starting 'As I search for paper to make notes I keep coming across lists of actors, odd names jotted down on postcards, remnants of attempts to address the play which closed last Saturday...'. The bottom window shows a 'File List' table with columns for Name, Item #, Extension, Path, Category, L-Size, MDS, Modified, and SHA.

Name	Item #	Extension	Path	Category	L-Size	MDS	Modified	SHA
BENNETT.72B	5300	72b	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72B	WordPerfect 6.0	33.02 KB	CAE2...	10/05/1995 08:59:34 ...	ABE
BENNETT.72C	5302	72c	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72C	WordPerfect 6.0	39.87 KB	9D68...	10/05/1995 09:45:46 ...	9B6
BENNETT.72D	5304	72d	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72D	WordPerfect 6.0	38.27 KB	5867...	01/04/1995 12:16:40 ...	082
BENNETT.72E	5306	72e	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72E	WordPerfect 6.0	21.91 KB	64E8...	02/04/1995 13:22:00 ...	A72
BENNETT.72F	5309	72f	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72F	WordPerfect 6.0	33.06 KB	8C6E...	11/01/1997 22:11:52 ...	984
BENNETT.72G	5310	72g	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72G	WordPerfect 6.0	21.21 KB	B073...	04/09/1996 00:38:04 ...	912
BENNETT.72H	5312	72h	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT.72H	WordPerfect 6.0	34.17 KB	5AB1...	16/04/1995 03:26:30 ...	847
BENNETT1	5314	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT1	WordStar 5.0	31.00 KB	C1F2...	01/01/1980 03:47:52 ...	8C4
BENNETT2	5316	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT2	WordStar 5.0	30.50 KB	1602...	01/01/1980 04:40:24 ...	1A1
BENNETT3	5318	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT3	WordStar 5.0	19.88 KB	E64A...	01/01/1980 02:15:30 ...	D43
BENNETT4	5320	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT4	WordStar 5.0	32.00 KB	493C...	01/01/1980 01:39:36 ...	766
BENNETT5	5322	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT5	WordStar 5.0	23.38 KB	3102...	01/01/1980 01:33:52 ...	0CE
BENNETT6	5323	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT6	WordStar 5.0	26.63 KB	E989...	01/01/1980 01:36:54 ...	85F
BENNETT7	5324	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT7	WordStar 5.0	21.63 KB	A1C7...	01/01/1980 06:54:56 ...	9DC
BENNETT8	5326	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/BENNETT8	WordStar 5.0	15.63 KB	7140...	01/01/1980 06:55:56 ...	B25
DAYOUT	5327	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/[root]/DAYOUT	WordPerfect 6.0	20.79 KB	CC24...	30/04/1995 08:06:22 ...	885
unallocated space	5121		bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space	Folder	0 B		28/08/2009 15:48:20 ...	
116	5227	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space/116	INSO Unknown ty...	14.00 KB	C696...	16/03/2009 15:59:19 ...	158
219	5229	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space/219	WordPerfect 6.0	13.00 KB	7618...	16/03/2009 15:59:19 ...	488
312	5231	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space/312	INSO Unknown ty...	6144 B	OCEA...	16/03/2009 15:59:19 ...	EBC
374	5233	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space/374	WordPerfect 6.0	33.00 KB	8764...	16/03/2009 15:59:19 ...	555
469	5235	<missing?>	bennettoriginals/alanbennettfloppydisk-bennettblue/files/unallocated space/469	INSO Unknown ty...	2048 B	D777...	16/03/2009 15:59:19 ...	AF3

Several legacy WP formats in use



Appraisal - Barbara Castle

AccessData Forensic Toolkit Version: 2.2.1.646 Database: localhost Case: Castle

File Edit View Evidence Filter Tools Help

Filter: -unfiltered- Define...

Explore Overview Email Graphics Bookmarks Live Search Index Search

Thumbnails

Loaded: 8,357 Filtered: 8,357 Total: 50,850 Highlighted: 1 Checked: 0 Show Tooltip

File List

<input type="checkbox"/>	Name	Label	Item #	Extension	Path	Category	P-Size	L-Size	MDS
<input type="checkbox"/>	Carved [1220608].emf		57289	emf	Quantum CY nonprop.0...	Windo...	n/a	13.80 KB	7AF7
<input type="checkbox"/>	Carved [1220608].gif		56808	gif	Quantum CY nonprop.0...	GIF	n/a	2929 B	9186
<input type="checkbox"/>	Carved [1224402].png		63657	png	Quantum CY nonprop.0...	PNG	n/a	33 B	DD81
<input type="checkbox"/>	Carved [1224704].emf		57174	emf	Quantum CY nonprop.0...	Windo...	n/a	948 B	5ECF
<input type="checkbox"/>	Carved [1225650].png		62197	png	Quantum CY nonprop.0...	PNG	n/a	45 B	6C38

Ready

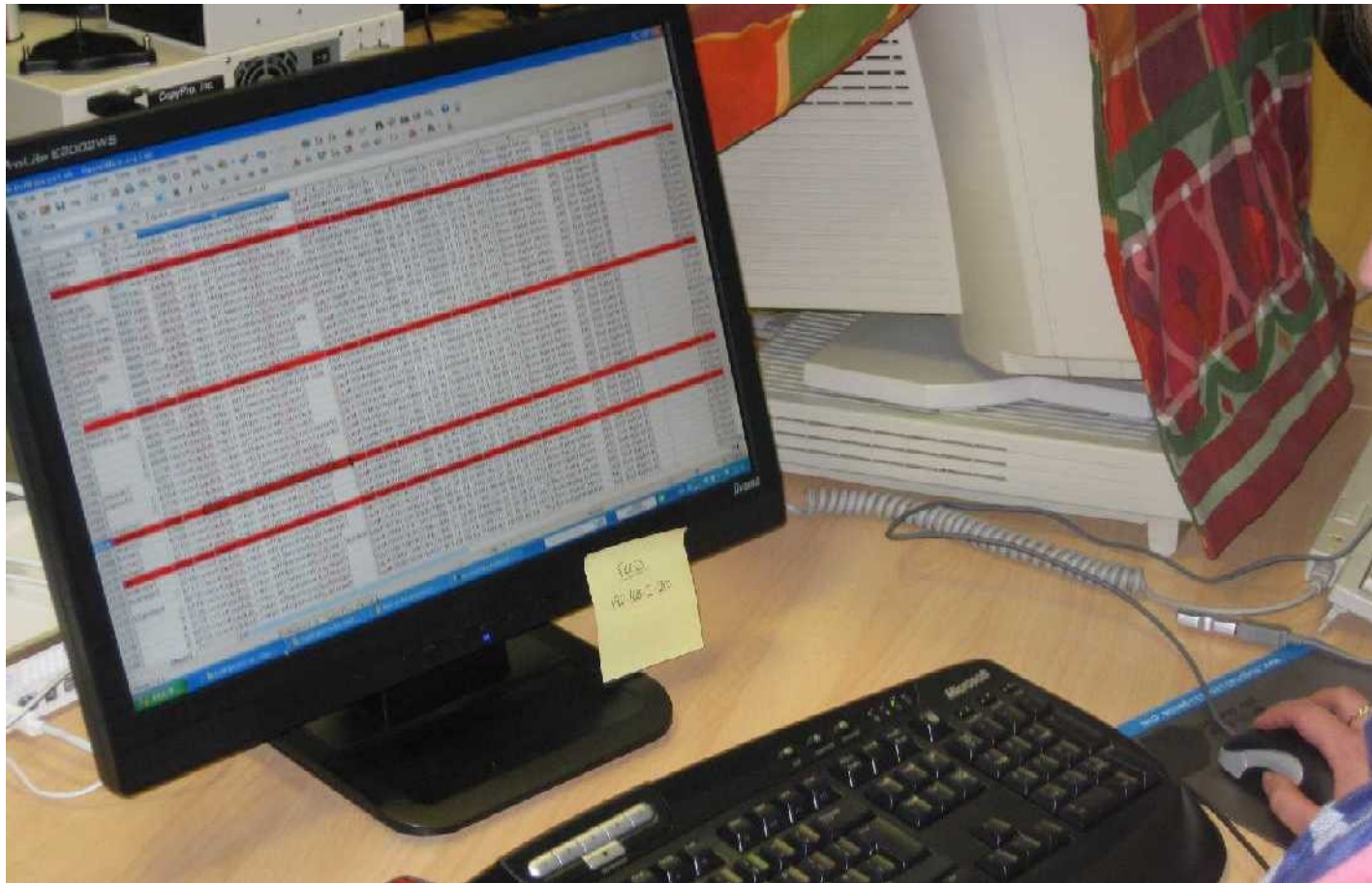
Graphics Tab Filter: Graphic Files

Quantum CY nonprop.001/Partition 1/WINDOWS 95 [FAT32]/[unallocated space]/000003/103747/Carved [1220608].emf

tex

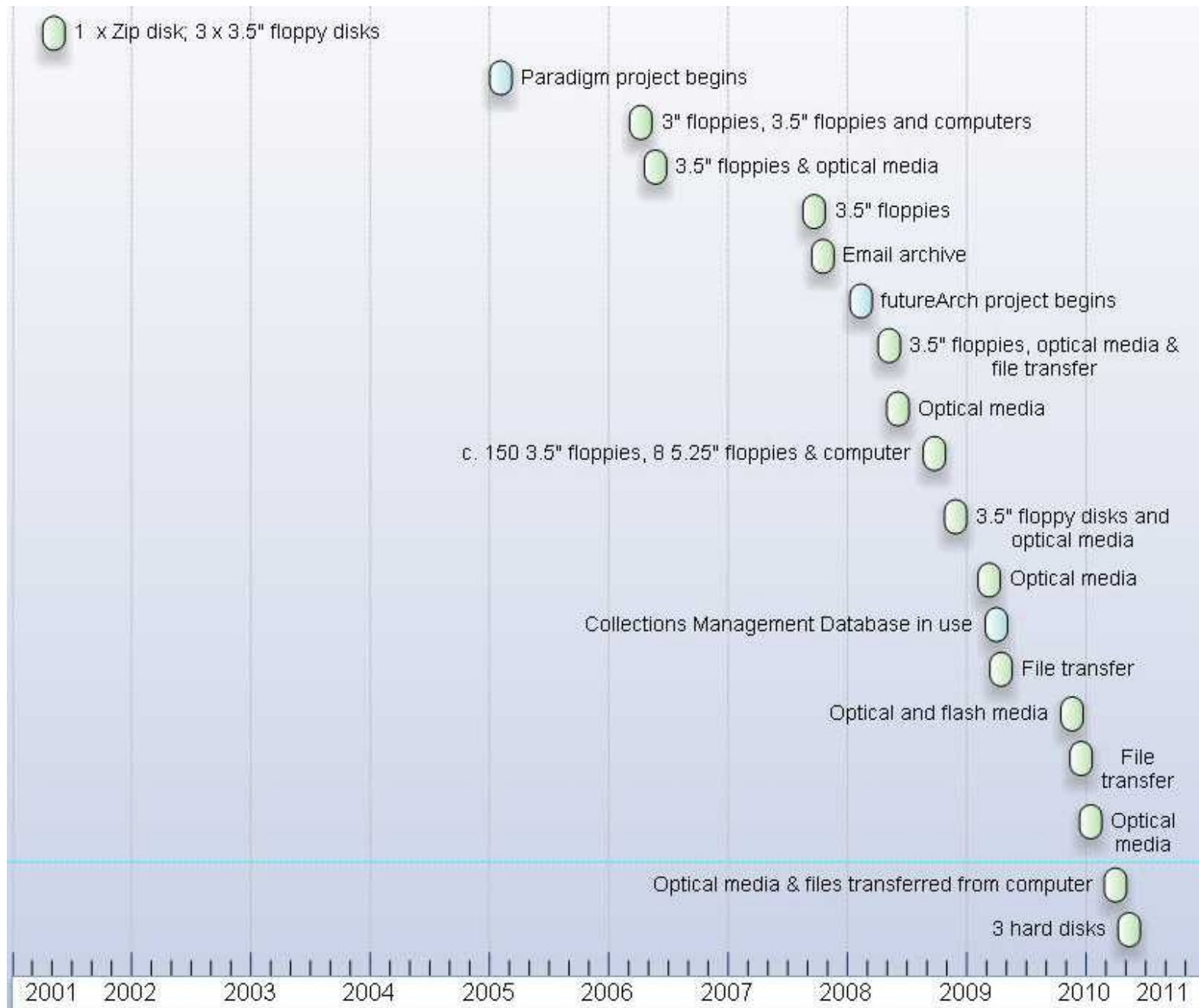


Digital object metadata



Metadata export From FTK used by cataloguer

Growth of born-digital



Many b-d colls in process

- Various phases:
 - Separation
 - Capture
 - Appraisal
 - Cataloguing
 - Preparation for publishing
- Supporting a number of project cataloguers/curatorial reviewers
- Scalability issues with current set-up

Questions?

Email me:

susan.thomas@bodley.ox.ac.uk

More online...

Bodleian Electronic Archives & Manuscripts

<http://www.ouls.ox.ac.uk/beam>

futureArch project blog

<http://futurearchives.blogspot.com/>