Universitat de Barcelona

Universitat Pompeu Fabra Barcelona

Facultat de Biblioteconomia i Documentació

Màster de Gestió de Continguts Digitals

Curs 2012-2013 – 2n semestre

Disseny d'una unitat d'anàlisi forense digital en una biblioteca

Treball Final d'Estudis

Autor: Teodoro López del Castillo Wilderbeek

Tutor: Dr. Miquel Térmens Graells

Barcelona, 10 de juny de 2013

Agraïments

A la meva família pel seu suport constant al llarg de tot el temps d'elaboració del treball, i en especial al meu germà Francisco, ja que la seva fe en la meva capacitat em va ajudar en tot moment,

als companys i companyes del màster, que m'han estimulat a millorar contínuament i m'han motivat per no defallir mai,

i molt especialment, vull donar les gràcies al Dr. Miquel Térmens, que amb els seus consells, coneixements i constant disposició m'han permès finalitzar el treball amb èxit i contemplar nous horitzons en l'àmbit acadèmic i professional.

SUMARI

RESUM EXECUTIU **1**

1. INTRODUCCIÓ **3**

2. PRESENTACIÓ **5**

2.1. DESCRIPCIÓ I IDENTIFICACIÓ DEL PROBLEMA 5

2.2. ABAST 6

2.2.1. Maquinari i programari acceptat 6

2.2.2. Maquinari i programari no acceptat 10

2.2.3. Tècniques de preservació digital emprades 15

2.3. OBJECTIU GENERAL 15

2.4. OBJECTIUS ESPECÍFICS 16

2.5. ANÀLISI INTERN I EXTERN - DAFO 16

2.6. METODOLOGIA 18

3. ESTUDI D'UNITATS JA EXISTENTS **20**

3.1. STANFORD UNIVERSITY LIBRARIES 20

3.2. BODLEIAN LIBRARY 23

3.3. NATIONAL LIBRARY OF AUSTRALIA 29

3.4. YALE UNIVERSITY 32

3.5. EMORY UNIVERSITY 34

4. REQUERIMENTS ESPECÍFICS DE LA UNITAT **38**

4.1. NIVELL BÀSIC 38

4.2. NIVELL AVANÇAT 41

5. RECURSOS NECESSARIS **44**

5.1. PROGRAMARI 44

5.1.1. Creació d'imatges de disc 44

5.1.1.1. FTK Imager (AccessData) 44

5.1.1.2. KryoFlux DiskTool Console (DTC) 49
5.1.2. Generació de *checksums* 53
5.1.2.1. MD5summer 53
5.1.2.2. HashX (BoilingBit) 55
5.1.3. *Suites* d'anàlisi forense digital 57
5.1.3.1. Autopsy (The Sleuth Kit) 57
5.1.3.2. AccessData Forensic Toolkit (FTK) 60
5.1.3.3. EnCase Forensic v7 (GuidanceSoftware) 64
5.1.4. Identificació i validació de fitxers 68
5.1.4.1. DROID (The National Archives) 68
5.1.4.2. JHOVE (JSTOR i Hardvard University Library) 69
5.2. MAQUINARI 72
5.2.1. Ordinadors 'Rosetta' 72
5.2.2. Estacions de treball forense 74
5.2.2. Targetes controladores 77
5.2.2.1. KryoFlux (The Software Preservation Society) 77
5.2.2.2. FC5025 (Device Side Data) 79
5.2.3. Unitats de disquet 81
5.2.3.1. Unitat de disquet de 5 ¼ polzades 81
5.2.3.2. Unitat de disquet de 3 ½ polzades 82
5.2.3.3. Unitat de disquet Zip 83
5.2.4. Docking stations 84
5.2.5. Write-blockers 84
5.3. PERSONAL 85
5.3.1. Competències 86
5.3.2. Formació 87
5.3.2.1. Digital Stewardship Certificate 87
5.3.2.2. Digital Forensics for Curation of Digital Collections 88
5.3.2.3. Cursos de formació de programari comercial 89
5.4. ENTORN DE TREBALL 90
5.5. PROCEDIMENTS DE TREBALL 91
5.5.1. Disquets 92

5.5.2. Discs durs 95
5.5.3. Altres 99

6. PROPOSTA DE PLA D'ACCIÓ 100
6.1. PRESSUPOST 100
6.1.1. Maquinari (nivell bàsic) 102
6.1.2. Maquinari i programari d'anàlisi forense 103
6.1.3. Formació 103
6.1.4. Cost de personal 104
6.1.5. Pressupostos totals (nivells bàsic, avançat amb unitat FRED i amb unitat FRED SR) 105
6.2. IMPLEMENTACIÓ 108
6.3. CALENDARI 111
6.4. SEGUIMENT I AVALUACIÓ 111
6.5. POLÍTICA DE DIFUSIÓ 113

7. CONCLUSIONS 114

GLOSSARI 116

BIBLIOGRAFIA 119

ÍNDEX DE FIGURES I TAULES 123

ANNEX. FORMULARI DE DONACIÓ 127

RESUM EXECUTIU

El present treball és una proposta de disseny per a la instal·lació d'una unitat d'anàlisi forense digital a la Biblioteca de Catalunya (model exportable a altres biblioteques) que permeti resoldre els problemes que plantegen els diferents suports d'emmagatzematge de dades tot fent ús (encara que no exclusivament) de maquinari i programari molt especialitzat, el qual habitualment està destinat a la investigació criminal, però en aquest cas adaptat a la preservació del patrimoni cultural dins les biblioteques.

L'objectiu, per tant, és la concreció dels passos necessaris i mínims per a la instal·lació d'aquesta unitat, sempre tenint en compte les necessitats i particularitats de la BC, un centre que té com a missió (entre d'altres) *vetllar per la conservació i la preservació del patrimoni bibliogràfic de Catalunya*[1], així que la integració de la unitat ha de ser coherent amb la missió esmenada. Un cop completada la instal·lació, la unitat haurà de ser capaç de preservar amb èxit i garantir la integritat dels continguts *born digital* dins els plans de preservació digital de la institució, encara que aquests es presentin en suports obsolets com disquets o altres de vigents com les memòries USB.

El pla de disseny contempla dins el seu abast quins suports físics s'inclouen al pla de treball i quins no, en funció d'un anàlisi dels més utilitzats dins l'àmbit territorial català i també els tipus de continguts acceptats, que són els que corresponen a arxius personals.

Per tal de facilitar un estat de la qüestió i donar exemples de donacions de diferents solucions emprades per superar els múltiples problemes que plantegen els suports amb contingut *born digital,* s'han estudiat unitats ja existents al Regne Unit i als EUA. Aquestes dades han servit com a referent per després escollir les millors opcions per als requeriments de la unitat.

[1] *Missió i funcions de la BC*. Disponible a: <http://www.bnc.cat/Coneix-nos/Qualitat-i-estrategia/Missio-i-funcions-de-la-BC> [data de consulta: 5 maig 2013]

A continuació s'han proposat dos possibles dissenys d'unitat d'anàlisi forense; un de nivell bàsic, amb els elements mínims i necessaris per a la recuperació de dades, i un de nivell avançat, amb les solucions actuals d'última generació.

Un cop finalitzat l'anàlisi de referents i els possibles dissenys, s'han exposat tots els recursos necessaris amb les solucions de programari i maquinari existents, quin tipus de personal s'hauria de contractar per operar amb la unitat, les especials característiques de l'espai necessari i una proposta de procediments de treball que inclouen tres diagrames per a casos concrets.

Per últim, s'ha elaborat una proposta de planificació d'implementació de la unitat, amb diferents pressupostos per tal de contemplar diferents necessitats, detall de fases i de tasques, calendari, dos tipus de propostes per avaluar els resultats de la unitat i quina seria la política de difusió per al projecte.

Paraules clau: anàlisi forense digital, unitat d'anàlisi forense digital, preservació digital, biblioteques, arxius, arqueologia digital

1. INTRODUCCIÓ

El concepte d''anàlisi forense digital' (*digital forensics* en anglès) és un camp de recerca utilitzat principalment en investigacions policials especialitzades per descobrir delictes on l'ordinador ha jugat un paper rellevant. L'any 2001 un grup d'experts proposà la següent definició:

Ús de mètodes científicament provats i derivats cap a la preservació, col·lecció, validació, identificació, anàlisi, interpretació, documentació i presentació d'evidències digitals derivades de fonts digitals amb l'objectiu de facilitar o fomentar la reconstrucció d'esdeveniments que hagin resultat ésser criminals, o ajudar a preveure accions no autoritzades que s'hagin demostrat com a perjudicials per a les operacions previstes[2].

En el cas d'Espanya, la Guardia Civil té un grup especialitzat en aquest tema, el *Grupo de Delitos Telemáticos*[3] *i* el cos de la Policia Nacional, que té la *Brigada de Investigación Tecnológica*[4], que actuen contra casos de fraus informàtics, pornografia infantil o pirateria informàtica, entre d'altres. A l'àmbit geogràfic de Catalunya els Mossos d'Esquadra compten amb la *Unitat Central de Delictes Informàtics* (UCDI).

Les eines i mètodes utilitzades a l'anàlisi forense digital representen una gran oportunitat per al sector del patrimoni cultural, ja que les biblioteques i arxius cada cop més reben noves donacions per part de personalitats que no es limiten a les tradicionals col·leccions en suport paper, sinó que també inclouen suports informàtics com disquets i discs durs (i de vegades, ordinadors). Les eines i processos d'anàlisi forense digital que les autoritats policials utilitzen per recuperar proves contra un possible delinqüent també poden ser útils als bibliotecaris per garantir la integritat dels

[2] *A Road Map for Digital Forensic Research*. Disponible a: <http://www.dfrws.org/2001/dfrws-rm-final.pdf> [data de consulta: 5 maig 2013]

[3] Pàgina web disponible a: <https://www.gdt.guardiacivil.es/> [data de consulta: 10 oct. 2012]

[4] Pàgina web disponible a: <http://www.policia.es/org_central/judicial/udef/bit_alertas.html> [data de consulta: 10 oct. 2012]

continguts digitals; a més, una eina de recuperació de dades aparentment esborrades del disc dur pot servir per recuperar (de forma parcial o total) documents importants que s'hagin perdut de forma accidental. Aquest treball es centrarà, per tant, en els beneficis que poden aportar aquestes eines i processos a les institucions culturals que treballin amb continguts nascuts digitals, els quals plantegen tres problemes greus:

- Maquinari. Un suport antic, com un disquet, no es pot llegir en les unitats actuals perquè no disposen de cap dispositiu adient.
- Programari. Les dades originals podrien haver estat creades amb eines inexistents actualment i totalment incompatibles amb les actuals.
- Sistema operatiu. Si les dades s'han creat amb sistemes operatius totalment incompatibles amb els actuals, s'hauria d'optar per solucions com l'emulació o la utilització dels ordinadors originals amb els seus sistemes operatius.

La tecnologia avança molt ràpidament, i això planteja un repte als centres que preserven la memòria cultural, atès que bona part de les dades digitals són úniques i no es poden reemplaçar si es destrueixen o es perden. Actualment, pràcticament tothom té els seus arxius personals en format digital (text, fotografies, vídeos, etc.) i la tendència es la substitució del paper pels suports informàtics. Es fàcil preveure que d'aquí a uns anys el suport digital serà majoritari als arxius personals i per tant, les biblioteques i els arxius han d'estar preparats per a tots els casos que es puguin presentar.

2. PRESENTACIÓ

2.1. DESCRIPCIÓ I IDENTIFICACIÓ DEL PROBLEMA

Els continguts d'origen digital (*born digital* en anglès) representen un greu problema per a les biblioteques i arxius, ja que a les seves col·leccions es guarden milers de dades amb formats obsolets, a diversos nivells (format de fitxers, sistema de fitxers, sistema operatiu, programari, maquinari i suport), que fan molt complex per al públic l'accés a aquestes dades i també hi ha el risc de pèrdua. Un cas hipotètic pot ser:

- Un document de text creat amb el programari *WordStar*, que funcionava sota el sistema operatiu MS-DOS, és molt possible que no s'hi pugui accedir amb les *suites* ofimàtiques actuals sota el sistema operatiu Windows 7. I encara que s'hi pogués accedir, molt possiblement la visualització del document no seria la correcta perquè no s'ha utilitzat el mateix programari amb què es va crear el fitxer.
- El document de text està guardat dins un disquet de 5 ¼, un suport obsolet del qual ja no es fabriquen unitats lectores i a més les plaques base dels ordinadors actuals no permeten la seva instal·lació.
- El disquet té una antiguitat que fa prioritari fer-ne una còpia de seguretat per tal de no perdre el seu contingut i s'ha de fer de manera que es pugui garantir que les dades originals no s'han alterat o modificat.

En totes aquestes qüestions les tècniques emprades en l'anàlisi forense digital constitueixen un ajut valuós per als agents implicats en la preservació del patrimoni cultural. Una d'elles és la creació d'imatges de disc, que consisteix en capturar els continguts bit per bit, la qual cosa assegura que totes les dades que es trobin al suport es preservaran íntegrament per a la seva anàlisi i recuperació posterior. Aquesta tècnica presenta grans avantatges, ja que cada suport d'emmagatzematge acostuma a contenir diversos fitxers i és més senzill i ràpid el tractament tècnic d'una sola imatge de disc que fer-ho amb cadascú dels fitxers (que podrien ser centenars).

Molts dels problemes que generen els continguts digitals es poden solucionar si les institucions culturals poden comptar amb una **unitat forense digital**, amb la qual s'aconseguiria un tractament documental adequat de les dades.

2.2. ABAST

Per definir l'abast s'han previst quins serien els suports que rebria la BC per part dels donants, els tipus de fitxers emprats habitualment per a arxius personals i els sistemes operatius i de fitxers utilitzats originalment. També s'ha fet un estudi dels fons i les col·leccions existents al centre. Així doncs, s'ha fet una tria de maquinari i programari que doni el menor nombre de problemes tècnics i legals possibles.

2.2.1. Maquinari i programari acceptat

La unitat de anàlisi forense digital que es planteja s'ocuparia de les següents funcions:

- Anàlisi i recuperació de dades que es trobin als suports següents:
    - **Disquets de 5 ¼ polzades**. El seu ús massiu als ordinadors personals per guardar dades durant els anys 70 i 80 fan imprescindible la seva integració a la unitat. Presenten dificultats per integrar-los a la unitat, com la impossibilitat de connectar unitats de lectura i escriptura a plaques base actuals, però no són insalvables.

Figura 1. Disquet de 5 ¼ polzades (Apple)

- **Disquets de 3 ½ polzades**. Encara utilitzats avui en dia per tenir còpies de seguretat de sistemes antics[5], foren àmpliament per guardar dades als ordinadors personals durant els anys 90 i per tant han de formar part de la unitat. Tècnicament, la seva integració no presenta complicacions excessives.

Figura 2. Disquet de 3 ½ polzades

- **Discs durs que segueixin els estàndards d'interfície IDE i SATA**. Gran part dels continguts digitals personals es guarden avui en dia dins discs durs que segueixen aquests estàndards. Per altra banda, és senzilla la seva anàlisi i recuperació amb el maquinari adequat.

Figura 3. Discs durs SATA i IDE

- **Discs Zip Iomega**. Durant la segona meitat dels anys 90, foren utilitzats en substitució dels disquets i per tant seria necessària la seva integració a la unitat. Per altra banda, no plantegen dificultats tècniques excessives.

[5] *The Floppy Disk*. Disponible a: <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/floppy/transform/> [data de consulta: 11 nov. 2012]

Figura 4. Disc Zip Iomega

- **Discs òptics de dades (CD-ROMs i/o DVD-ROMs).** Encara que s'utilitzen principalment per a la distribució de programari, també s'han utilitzat per guardar dades d'ús personal que requereixen un important espai en bytes. La seva integració en la unitat, per altra banda, seria senzilla.

Figura 5. CD-ROM

- **Memòries USB**. Avui en dia és el dispositiu principal per guardar i conservar dades fora del disc dur i a més, la universalitat de la connexió USB no presenta cap problema per a la unitat.

Figura 6. Memòria USB de 4 GB (Sandisk)

- **Discs durs externs amb connexió USB**. Encara que no s'utilitzen a gran escala, es pot preveure que d'aquí a un anys serà un gran problema la seva anàlisi i recuperació.

Figura 7. Disc dur USB extern (Western Digital)

- Els continguts digitals amb què treballaria la unitat serien de les següents tipologies:
  - Documents de text.
  - Fotografies.
  - Vídeos.
  - Gravacions d'àudio.
  - Presentacions.
  - Fulls de càlcul.
  - Correus electrònics.

- Els diferents estàndards informàtics que han existit al llarg del temps ha generat un nombre important de sistemes de fitxers. La unitat treballaria amb els més utilitzats a Espanya, que serien:
  - FAT. Sistema desenvolupat per a MS-DOS i utilitzat posteriorment per a sistemes Windows.
  - NTFS. Sistema desenvolupat per a Windows NT i també inclòs a versions de Windows més avançades, com Windows XP i Windows 7.
  - HFS+. Sistema d'arxius per a OS X.

- ISO 9660. Norma d'arxius per a CD-ROM. Aquest estàndard inclou també el sistema UDF, que s'utilitza per als DVD-ROM.
- ext. Sistema d'arxius per a Linux.
- Solaris UFS. Sistema d'arxius per a Linux.

2.2.2. Maquinari i programari no acceptat

La unitat no s'ocuparia de les següents funcions:

- Anàlisi i recuperació de dades que es trobin als suports següents:
  - **Disquets de 8 polzades**. La seva antiguitat (es començaren a comercialitzar l'any 1971[6]), l'escassa difusió que van tenir i les dificultats tècniques que implicarien la seva integració en la unitat aconsellen descartar aquest suport, ja que una unitat de lectura i escriptura d'aquest tipus és una peça molt rara i extremadament costosa.

Figura 8. Disquet de 8 polzades

  - **Disquets de 2, 2 ½ i 3 polzades**. Degut a les limitacions tècniques del disquet de 5 ¼, diverses companyies van desenvolupar nous suports que

[6] *20th century disk storage chronology*. Disponible a: <http://www-03.ibm.com/ibm/history/exhibits/storage/storage_chrono20.html> [data de consulta: 3 des. 2012]

no van aconseguir imposar-se fins l'arribada dels disquets de 3 ½. La seva integració a la unitat presentaria excessives complicacions tècniques i per altra banda, la possibilitat que arribin aquests materials *born digital* a la unitat és molt baixa.

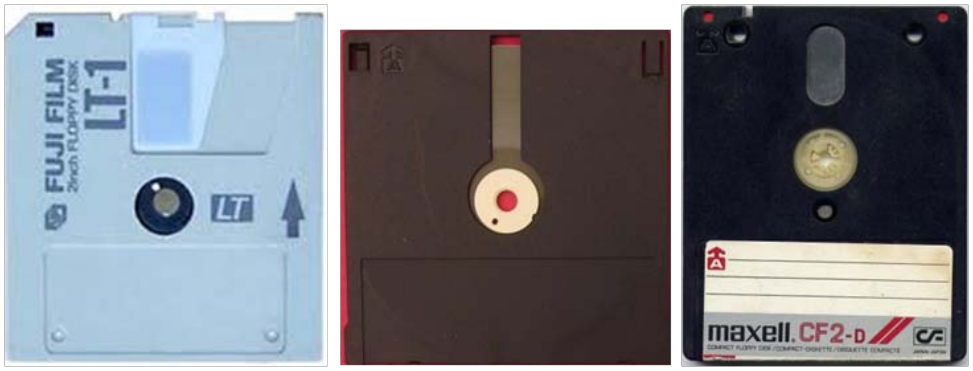


Figura 9. Disquets de 2, de 2,8 i de 3 polzades (Fuji Film, Smith Corona i Maxell)

- **SuperDisk**. També conegut com LS-120 (per la seva capacitat de 120 MB) i posteriorment LS-240 (capacitat de 240 MB), era un disc totalment compatible amb els disquets de 3 ½ polzades i estava destinat a ser el seu substitut. Però l'èxit superior del disc Zip per una banda i el fort abaratiment dels CD-ROM per una altra, van deixar aquest estàndard com a molt minoritari i per tant, no se l'hauria de considerar per ser integrat a la unitat.

Figura 10. Disc LS-120 (Imation)

- **Discs magnetoòptics**. La gran varietat de formats i estàndards i l'alt cost d'adquisició del maquinari necessari els fan inviables per a la unitat.

Figura 11. Discs magnetoòptics (Sony, Fujitsu i Olympus)

- **Cintes de casset**. Aquest suport fou usat intensament durant la dècada dels 80 per emmagatzemar dades a ordinadors personals de 8 bits ZX Spectrum o l'Amstrad CPC 464, però el seu ús per a arxius personals fou molt minoritari, ja que aquest suport s'utilitzà especialment per a jocs.

Figura 12. Ordinador ZX Spectrum amb lector/gravador de casset incorporat

- **Targetes de memòria**. L'enorme varietat de formats (PC Card, SmartMedia, Memory Stick , SD, miniSD, microSD, MMCmicro, XQD, etc.) fa que sigui molt complicat integrar aquest tipus de suport a la unitat. No obstant, es podria considerar utilitzar-ne un nombre limitat.

Figura 13. Targetes de memòria (Sandisk, Olympus i Sony)

- **Targetes de mòbil SIM**. Poden emmagatzemar adreces i contactes, i missatges SMS, que no són continguts amb als que operaria la unitat.

Figura 14. Targetes de mòbil SIM

- **Memòries RAM**. L'anàlisi de les memòries volàtils és un camp més propi de la investigació criminal. Per altra banda, no és un suport convencional on es guardin documents.

Figura 15. Memòries RAM

- **Targetes perforades**. L'adaptació de maquinari per aquest suport resultaria excessivament costosa i per altra banda, no és un suport que s'hagi utilitzat per guardar continguts digitals d'interès per a les

institucions culturals, sinó més aviat per qüestions tècniques, com l'ús de llenguatges de programació[7].

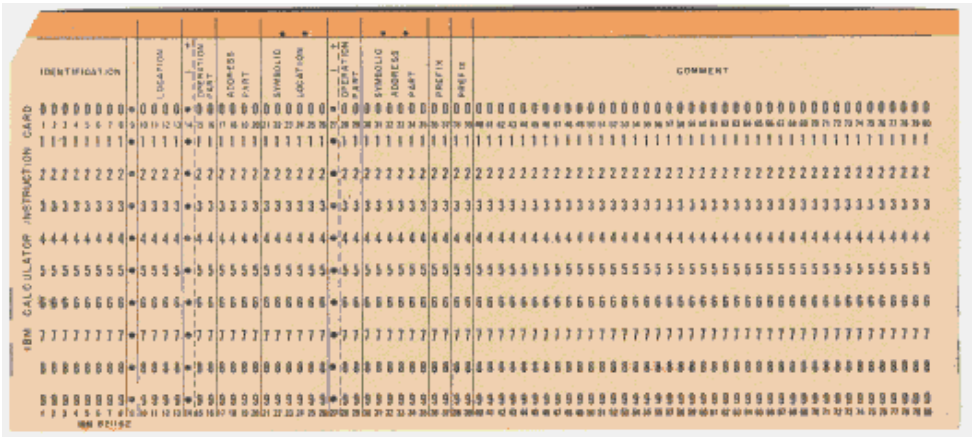


Figura 16. Targeta perforada (IBM)

- **Cintes magnètiques d'emmagatzematge de dades**. Aquest tipus de suport requereix d'un maquinari molt especialitzat i costós, que fa inviable la seva integració a la unitat. Per altra banda, aquest suport s'ha utilitzat majoritàriament en empreses per a dades de caràcter industrial, i no per arxius de caire personal.

Figura 17. Cinta magnètica (3M)

[7] *Punched cards. A brief illustrated technical history*. <http://homepage.cs.uiowa.edu/~jones/cards/history.html> [data de consulta: 15 des. 2012]

- Qualsevol mitjà que no contingui dades informàtiques digitals, com vídeo en cintes VHS, àudio en cintes de casset, discs òptics de DVD-Video, vídeo en LaserDisc, en Blu-Ray, etc.
- Suports amb protecció anticòpia.

- No es faria anàlisi i recuperació de dades dels següents continguts digitals:
  - **Videojocs en qualsevol suport i format**. Encara que els videojocs són objectes culturals d'interès per a les biblioteques, és un programari que implica excessives complicacions tècniques pels problemes de compatibilitat degut a la multiplicitat de plataformes de programació i els conflictes per drets d'autor que presenten: moltes companyies de creació de videojocs ja no existeixen i caldria un esforç excessiu contactar amb els programadors, que són els titulars dels drets.
  - Qualsevol altre programari comercial que no es trobi en el domini públic.

2.2.3. Tècniques de preservació digital emprades

Dins les tècniques existents, la unitat només contempla l'opció de *refreshing,* ja que la migració i l'emulació impliquen la creació de programari fet a mida per a cada cas. Per tant, la BC hauria de decidir si és necessari aplicar aquestes tècniques a les seves col·leccions. No obstant, el present treball incorpora exemples de migració i d'emulació (per al primer cas, vegeu *3.2. BODLEIAN LIBRARY* i per al segon cas, vegeu *3.5. EMORY UNIVERSITY*).

2.3. OBJECTIU GENERAL

El present treball pretén ser una guia de referència per tal de crear una unitat d'anàlisi forense digital a la **Biblioteca de Catalunya**, les característiques de la qual són equivalents i exportables a altres biblioteques nacionals o universitàries grans.

2.4. OBJECTIUS ESPECÍFICS

- Identificar les situacions més habituals d'anàlisi forense en una biblioteca o un arxiu històric:
    - Analitzant les fonts de referència exhaustivament.
    - Establint conclusions.

- Establir el funcionament d'unitats ja existents d'anàlisi forense:
    - Consultant i analitzant la bibliografia disponible sobre les unitats.
    - Estudiant el maquinari i programari utilitzat a les unitats ja existents.
    - Extraient les dades més rellevants.

- Establir procediments de treball en aquesta unitat:
    - Fent anàlisi i recuperació de dades amb programari i maquinari.
    - Analitzant els fluxos de treball dins altres unitats.
    - Elaborant fluxos de treball precisos.

- Definir un pla de treball per posar en marxa la unitat:
    - Consultant preus del maquinari i del programari als fabricants i/o empreses distribuïdores.
    - Calculant costos d'enviament i impostos.
    - Establint costos de personal i altres despeses necessàries.

2.5. ANÀLISI INTERN I EXTERN - DAFO

El següent anàlisi DAFO s'ha realitzat, en primer lloc, per resumir el conjunt de fortaleses (factors propis desfavorables) i debilitats (factors interns desfavorables) que presenta la Biblioteca de Catalunya en relació a la instal·lació d'una unitat forense digital i, en segon lloc, per recollir el conjunt d'oportunitats (factors que es poden

controlar) i amenaces (factors que no es poden controlar) que ha de tenir en compte la BC.

ANÀLISI INTERN. Fortaleses:

- **Ajuda en els objectius de les institucions culturals.** Les biblioteques i arxius històrics (i molt especialment la Biblioteca de Catalunya) tenen com a funcions la preservació del patrimoni i això inclou el material *born digital*. Una unitat forense digital seria un gran ajut per l'acompliment d'aquesta missió.
- **Possibilitat de descobrir col·leccions noves i desconegudes.** Els impediments tècnics dels suports obsolets són un obstacle per als usuaris, que es poden solucionar amb la instal·lació de la unitat.

ANÀLISI INTERN. Debilitats:

- **Manca d'experiències prèvies a Catalunya.** L'anàlisi forense digital és un camp que s'ha treballat molt als països anglosaxons, però a Catalunya encara no s'ha fet res de semblant i per tant aquest és un camp nou i molt desconegut.
- **Manca de personal qualificat.** El personal de la Biblioteca de Catalunya no té una formació acadèmica en anàlisi forense i per tant es requeriria invertir part del pressupost en formació especialitzada.
- **Pressupost limitat.** Les institucions culturals a Catalunya han patit fortes retallades durant l'any 2012 i això pot representar un problema per a l'adquisició dels recursos necessaris.
- **Necessitat d'adequar un espai per a la unitat.** Degut a les característiques físiques de les unitats d'anàlisi forense, s'hauria d'adaptar un lloc per tal que pugui executar les seves funcions.

ANÀLISI EXTERN. Oportunitats:

- **Unitats ja existents**. A diferents biblioteques del Regne Unit i dels Estats Units, ja s'han desenvolupat unitats d'anàlisi forense amb funcions de recuperació de dades digitals amb bons resultats, que serviran de referents per al projecte.

- **Cooperació amb biblioteques amb experiència en la matèria**. La possibilitat de compartir coneixements amb centres que ja han treballat amb èxit en casos de recuperació de dades és una gran oportunitat per ampliar i millorar les competències dins la preservació digital.
- **Programari lliure**. Existeixen solucions de programari lliure molt efectives i totalment gratuïtes que ajudarien a abaratir els costos.

ANÀLISI EXTERN. Amenaces:

- **Maquinari de difícil adquisició.** Una part de les unitats lectores de suports obsolets ja no es fabriquen ni es comercialitzen; per altra banda, el maquinari dissenyat específicament per a l'anàlisi forense representa un alt cost.
- **Compra de llicències informàtiques.** Part del programari d'anàlisi forense és comercial, i no totes les institucions tenen pressupost suficient per costejar-ho.
- **Gran varietat de suports.** Cada suport informàtic està desenvolupat d'una manera concreta i això implica que els processos de recuperació de dades en cada cas diferirà necessàriament.
- **Coneixements informàtics.** Es necessita un mínim de coneixements per a l'ús del programari i del maquinari dedicat a l'anàlisi forense digital. La gran varietat de sistemes de fitxers i d'estàndards informàtics també representen un obstacle.

2.6. METODOLOGIA

La metodologia que s'ha emprat per a l'estudi de casos ja existents, s'ha basat en l'anàlisi exhaustiu d'informes elaborats pel personal dels centres i l'estudi de la bibliografia disponible.

Per a la selecció dels requeriments obligatoris i optatius que haurà de complir la unitat, s'ha analitzat quins són els processos tècnics de la Biblioteca de Catalunya relacionats amb la preservació digital per tal de conèixer les seves prioritats primàries i

secundàries. Les fonts d'informació inclouen webs institucionals i les seves publicacions oficials.

A continuació s'ha elaborat l'apartat de recursos necessaris, a partir de proves realitzades amb el programari i el maquinari seleccionats en funció dels requeriments, la documentació posterior per avaluar els resultats i amb l'anàlisi dels entorns de treball a les unitats ja existents.

Finalment, la proposta de pla d'acció s'ha creat a partir d'una selecció dels recursos i el seu cost, en funció de les diferents necessitats de la unitat.

3. ESTUDI D'UNITATS JA EXISTENTS

3.1. STANFORD UNIVERSITY LIBRARIES

L'any 2008, les Stanford University Libraries van fer un estudi per tal de quantificar el volum, distribució i antiguitat d'ítems *born digital* dins les seves col·leccions. El resultat va permetre identificar més de 18.000 ítems, amb un creixement que augmenta any rere any i que presenten un gran risc de pèrdua de dades. Gràcies als consells i recomanacions vers maquinari i programari d'anàlisi forense que van rebre per part de Jeremy Leighton John (British Library) i Susan Thomas (Bodleian Libraries), l'any 2009 les Stanford University Libraries van iniciar la seva pròpia unitat d'anàlisi forense digital, amb l'adquisició de dues unitats FRED[8] de l'empresa Digital Intelligence (per a més detalls, vegeu *5.2.2. Estacions de treball forense*), de llicències de programari forense comercial (EnCase i FTK) i d'una càmera rèflex digital per fotografiar els ítems. El mateix any, les Stanford University Libraries es van convertir en membres del projecte associatiu AIMS, el qual també el formen la University of Virginia Library, la University of Hull, i la Yale University, que té com objectiu definir directrius de bones pràctiques per a diferents casos institucionals en quant a la gestió del material *born digital*.

Actualment, el personal assignat a la unitat consisteix en:

- Michael Olson, cap de projecte.
- Glynn Edwards, cap de la secció de manuscrits.
- Peter Chan, arxiver digital.
- Henry Lowood, curador d'història de ciència i tecnologia.

Les col·leccions amb què s'està treballant són:

- Arxiu de Stephen Jay Gould. Paleontòleg, biòleg evolutiu i historiador de la ciència, la seva col·lecció conté obres diverses, correspondència, recerca i

[8] *FRED*. Disponible a: <http://www.digitalintelligence.com/products/fred/> [data de consulta: 5 gen. 2013]

suports informàtics antics. Els suports consisteixen en 60 disquets de 5 ¼ i 3 ½ polzades, targetes perforades i tres cintes magnètiques. Els disquets contenen bases de dades bibliogràfiques i esborranys de les obres de Gould, mentre que les targetes perforades contenen conjunts de dades utilitzats en el seu treball de recerca.

- Arxiu de Robert Creeley. Poeta, novel·lista, escriptor, editor i assagista, a la seva col·lecció trobem poema i prosa, així com còpies de seguretat de correus electrònics dins 53 disquets de 3 ½ polzades, 5 disquets Zip i 3 CD-ROMs.
- Arxiu de Peter Koch. Impressor i director de la fundació CODEX, la qual està dedicada a la preservació i la promoció del llibre entès com objecte artístic, la seva col·lecció conté un disc dur amb correspondència i fitxers d'arts gràfiques.
- Col·lecció del projecte Xanadu. Ted Nelson inicià l'any 1960 el projecte Xanadu, que fou el primer projecte d'hipertext i el precursor de l'actual World Wide Web. La col·lecció consisteix en 6 discs durs amb documentació relacionada amb el projecte.

La unitat encara no té un *workflow* concret i definit d'aplicació general, però sí que es disposa d'informació vers els passos que s'han donat fins ara.

Els primers esforços en capturar i processar els continguts es van concentrar en l'arxiu de Stephen Jay Gould. En un principi, es van crear imatges de disc mitjançant el programari ImageTool i una interfície per llegir discs de 5 ¼ polzades a una unitat FRED. No obstant, ImageTool no donava cap informació que confirmés la creació d'una imatge de disc amb el contingut exacte de l'ítem original. Per tant, es va optar per utilitzar un ordinador antic amb una unitat lectora de disquets 5 ¼ ja incorporada; per crear les imatges de disc s'utilitzà el programari FTK Imager[9], que a més genera informes que confirmen que s'ha generat una imatge de disc correcta i també llistats de fitxer dels continguts del disquet.

[9] Programa disponible a: <http://www.accessdata.com/support/product-downloads> [data de consulta: 5 maig 2013]

Per processar les dades, s'usà el programari Forensic Toolkit (FTK) per extreure les metadades tècniques (mida del fitxer, dates de creació, de darrera modificació, format de fitxer, etc.) dels fitxers trobats a les imatges de disc. Dins aquest procés, s'identifiquen també dades sensibles, com les corresponents a targetes de crèdit, a la seguretat social o les qualificacions acadèmiques mitjançant les funcions de cerca de patrons i de text complet; aquests continguts són marcats com "Privilegiats" i no són d'accés públic.

La col·lecció es va classificar per sèries i subsèries, en funció de paraules clau. Per identificar els continguts dels fitxers amb formats obsolets s'utilitzà el visualitzador intern del FTK que permet la lectura de més de 200 formats de fitxer. Es van generar informes en format XML/HTML dels fitxers per ser exportats al repositori Hypatia[10], que actualment permet l'accés a continguts *born digital* de Stephen Jay Gould, Robert Creeley i Peter Rutledge Koch, entre d'altres col·leccions d'institucions membres del projecte AIMS. Dins aquest repositori, hi ha accés directe a part dels fitxers preservats, juntament amb una fotografia del suport original i la imatge de disc corresponent o bé els fitxers originals.

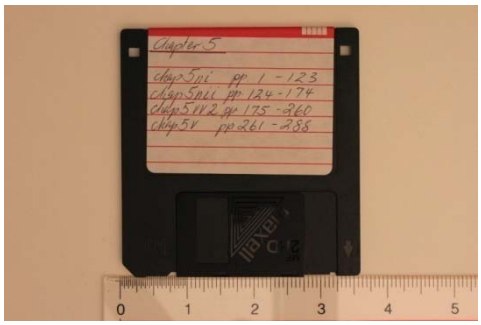


Figura 18. Disquet de la col·lecció Stephen Jay Gould

La col·lecció de Robert Creeley va presentar nombrosos problemes. Com que Robert Creeley va esborrar fitxers abans de donar els disquets i aquest contingut es podria revelar si es fessin imatges de disc, es va optar per treballar amb imatges lògiques. El maquinari emprat fou una unitat de disquet especialment dissenyada i construïda per Peter Chan i el programari escollit fou FTK Imager de l'empresa AccessData. Una

[10] Pàgina web disponible a: <http://hypatia-demo.stanford.edu> [data de consulta: 2 març 2013]

qüestió tècnica que resultà insalvable fou un format de còpies de seguretat utilitzat als disquets, que el programari no va poder reconèixer (possiblement creats amb programari propietat de l'empresa Iomega, creadora dels discs Zip).

Per altra banda, gran part dels continguts corresponien a correus electrònics i part d'ells estaven comprimits en un sol fitxer; en un primer moment, s'estimava una quantitat de 50.000 correus, però la quantitat real era superior a les 80.000. Per tal d'extreure informació útil i recuperable posteriorment, la unitat va decidir testejar l'ús de diagrames de xarxa, una eina molt utilitzada dins l'anàlisi de xarxes socials. El procés consistí en guardar les dades de la capçalera dels correus electrònics (que contenen la informació del remitent, el destinatari, l'assumpte i la data) en un fitxer csv, i obrir-lo amb Gephi[11], programari gratuït de visualització de xarxes. El resultat fou el següent diagrama, que mostra els noms dels emissors i els receptors, així com el volum de correspondència.

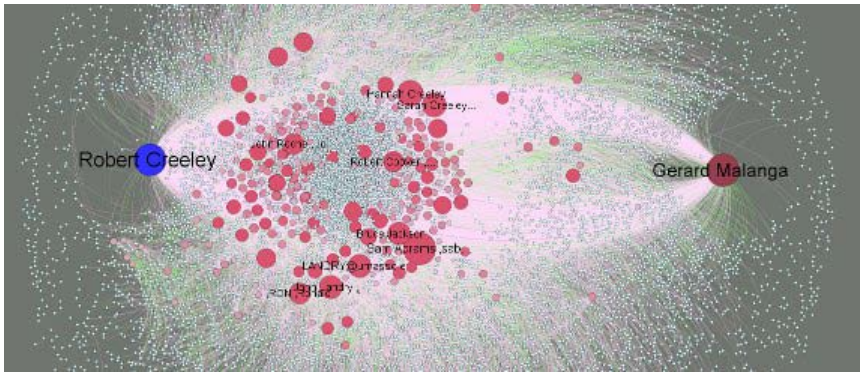


Figura 19. Diagrama de la xarxa de correu electrònic de Robert Creeley

3.2. BODLEIAN LIBRARY

Les Bodleian Libraries integren el servei de biblioteques de la Universitat d'Oxford (Regne Unit), les col·leccions de les quals formen un total de més de onze milions de documents. Dins d'aquesta xarxa es troba la Bodleian Library, que és la biblioteca de recerca principal i té la missió de rebre el dipòsit legal del Regne Unit, juntament amb

[11] Programa disponible a: <http://gephi.org/> [data de consulta: 4 abr. 2013]

la British Library, la Cambridge University Library, la National Library of Scotland, la Library of Trinity College de Dublin i la National Library of Wales[12]. La seva missió, per tant, és la mateixa que la Biblioteca de Catalunya en quant a la preservació del patrimoni documental, que inclou el material *born digital*, el qual arriba en diversos suports (disquets, ordinadors sencers, etc.) i formats (àudio, text, correu electrònic, fulls de càlcul, etc.). Per tal d'assegurar la seva preservació i el seu accés, es creà una secció dins la biblioteca per tal de proveir un repositori digital, el qual rep el nom de Bodleian Electronic Archives and Manuscripts (BEAM), que ha de proveir els mitjans per reunir, descriure, gestionar i preservar els components digitals dins les col·leccions d'arxius i manuscrits.

Per tal de desenvolupar el BEAM, la primera fase fou crear el projecte futureArch l'any 2009, el qual va establir els següents objectius:

- Polítics. Establir el *workflow* del tractament documental de les col·leccions híbrides (formades per diversos suports, com paper, vídeo, àudio, etc.), establir rols i responsabilitats i implementar canvis pressupostaris.
- Culturals. Crear eines, sistemes i formació per al personal arxiver per tal que estiguin preparats per treballs amb col·leccions híbrides (formades per diversos suports, com paper, vídeo, àudio, etc.).
- Suport a la recerca. Desenvolupar serveis segurs d'accés per a usuari, amb informació de les metadades dels arxius híbrids.
- Col·leccions digitals. Desenvolupar noves relacions amb els creadors i donants dels arxius, així com habilitar transferències electròniques de documentació a la biblioteca. Per altra banda, també es vol processar el registre dels materials *born digital* i desenvolupar l'arxiu web de les Bodleian Libraries.
- Infraestructura. Actualment, el BEAM ja actua com el repositori digital per a materials d'arxiu en format digital i comparteix els seus serveis amb la resta

[12] *Agency for the Legal Diposit Libraries*. Disponible a: <http://www.legaldeposit.org.uk/index.html> [data de consulta: 7 abr. 2013]

d'institucions a les Bodleian Libraries, com l'Oxford University Research Archive.

Algunes col·leccions amb les que s'està treballant són:

- Arxiu del Partit Conservador. En creixement continu, el material *born digital* és considerable: dossiers de premsa, transcripcions de discursos, fullets, pòsters, imatges, vídeos o fitxers del Departament de Recerca del Partit.
- Arxiu literari de la impremta Clutag. Fundada l'any 2000 per Andrew McNeillie, aquesta impremta ha publicat obres poètiques d'autors britànics importants com Seamus Heany o Tom Paulin. La Bodleian ha estat adquirint aquest fons de forma híbrida des del seu origen.
- Arxiu de la baronessa Nicholson. L'extens arxiu d'Emma Nicholson, membre de la Cambra dels Lords, inclou material *born digital*.
- Arxiu de Sir Isaiah Berlin. Considerat un dels principals pensadors liberals del segle XX, el seu arxiu inclou entrevistes en fitxers d'àudio.
- Arxiu d'Edmund Dell. Polític i fundador del canal de televisió Channel Four, va donar cassets d'àudio que contenien els seus diaris.
- Arxiu d'Alan Bennett. Dramaturg, actor, novel·lista i guionista, els seus documents ocupen uns quinze metres lineals, produïts entre 1960 i 2008, que inclouen onze disquets de 3 ½, produïts entre 1992 i 2002.
- Arxiu de Barbara Castle. Una de les dones polítiques del Partit Laborista més importants, el seu arxiu inclou material digital com correspondència o esborranys de les seves memòries escrits als anys 80 i 90. El detall de la donació consistí en:
    - 800 capses, produïdes entre 1903 i 2002.
    - 1 capsa de vídeo i 1 capsa d'àudio.
    - 31 disquets Amstrad PCW de 3 polzades, produïts entre 1989 i 1992.
    - 2 ordinadors; un que fou utilitzat entre 1992 i 1997 i un altre que fou utilitzat entre 1997 i 2002.
    - 4 disquets de 3 ½ polzades, utilitzats com a còpies de seguretat.

En el cas de l'arxiu de Barbara Castle, la Bodleian Library va aconseguir amb èxit una migració de les dades de text escrites originalment en el programari LocoScript de Locomative Software dissenyat per l'ordinador personal Amstrad PCW. Aquest ordinador fou creat originalment l'any 1985 i com a suport físic utilitzava disquets de 3 polzades. La solució final consistí en programari i maquinari dissenyats específicament per a la migració: es van connectar un ordinador Amstrad PCW en funcionament i un ordinador portàtil amb sistema operatiu Linux que executava una 'màquina virtual' amb un sistema operatiu Windows 95 mitjançant un cable de comunicació de dades, mentre que un programari de migració va transformar les dades en LocoScript a ASCII o bé RTF, que va permetre la lectura dels fitxers originals.

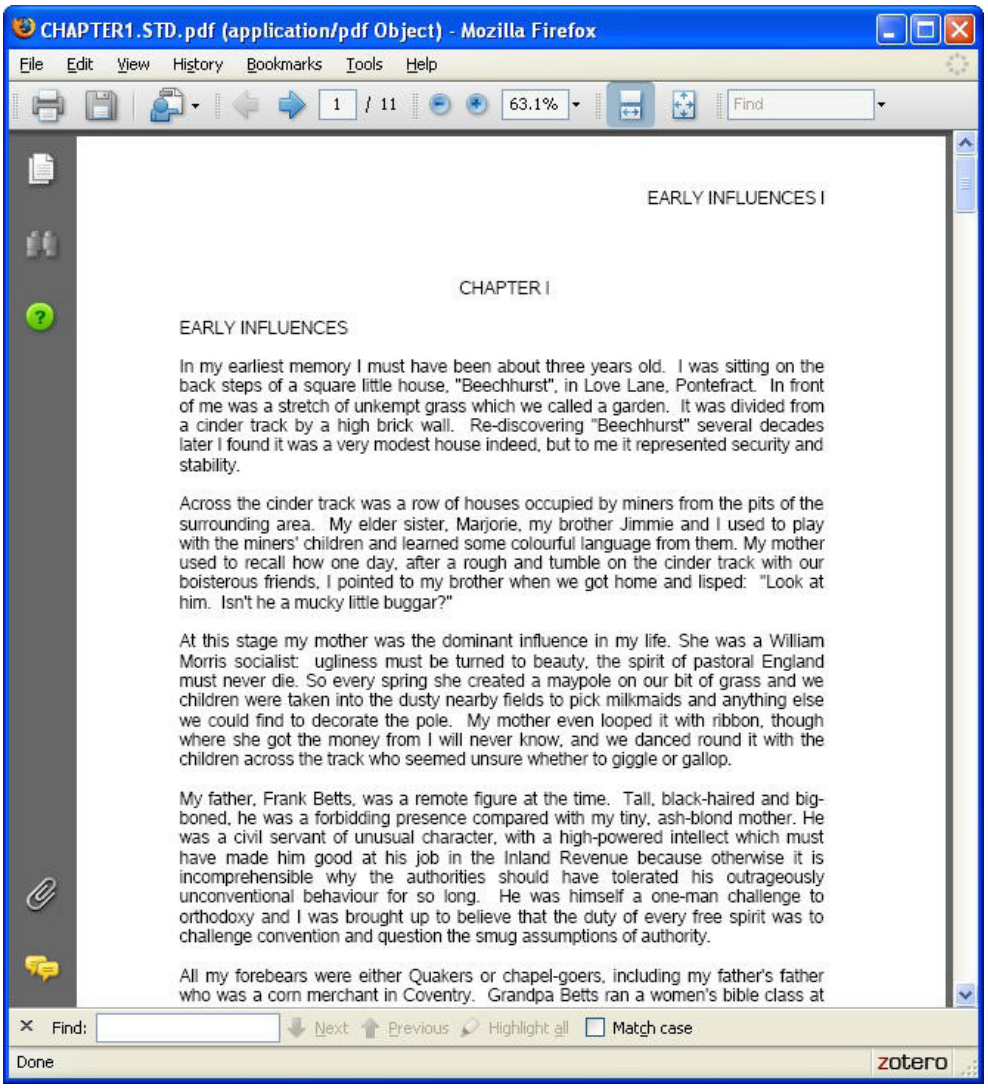


Figura 20. Text migrat al format PDF (arxiu de Barbara Castle)

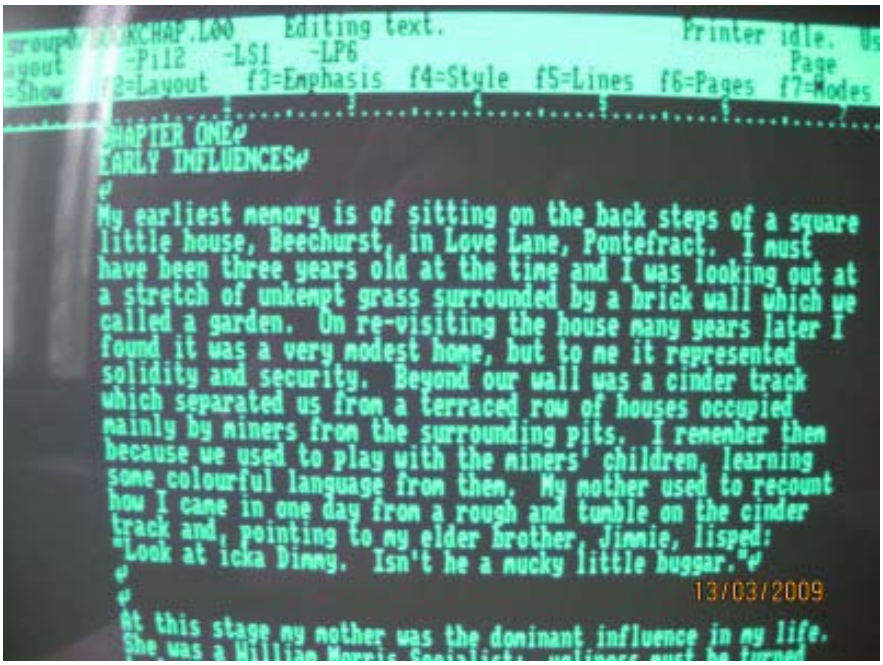


Figura 21. Text original en format LocoScript de l'ordinador Amstrad PCW (arxiu de Barbara Castle)

El *workflow* general que s'aplica als arxius híbrids consisteix en:

- Separació. Els ítems d'emmagatzematge digital es separen del material en suport paper i s'incorporen al repositori BEAM, el qual garanteix la integritat del contingut original. Tant els materials tradicionals com els digitals es gestionen dins una base de dades de col·leccions per tal d'evitar la dispersió de l'arxiu.

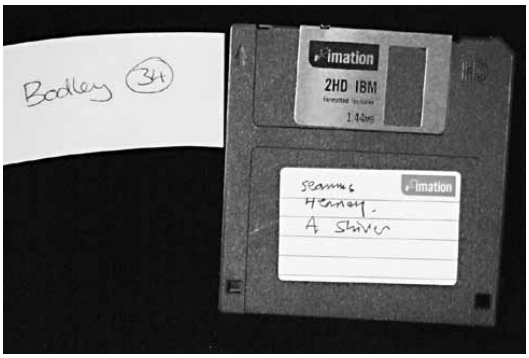


Figura 22. Disquet de 3 ½ polzades corresponent a l'arxiu de la impremta Clutag

- Captura. A cada ítem de la col·lecció se li assigna un número de inventari i se'n fa una fotografia per tal de crear un registre de referència a futurs catalogadors

i investigadors. Posteriorment, es fan servir eines de maquinari i programari forenses per crear una imatge de disc de cada ítem. Per tal de evitar qualsevol afectació del contingut, s'utilitza un dispositiu *write-blocker*. Per últim, es verifica la imatge creada per tal d'assegurar que les dades d'aquesta imatge són una representació exacta de les dades del suport original.

Figura 23. Creació de la imatge d'un disc dur mitjançant una unitat FRED

A l'acabar el procés, s'adquireixen i es guarden al repositori de preservació (juntament amb metadades de col·lecció i d'accés) les següents dades:

- Una còpia fidedigna dels continguts de cada ítem, agrupats en un objecte digital (la imatge de disc) i un valor *hash* que identifica de forma unívoca cada objecte.
- Un llistat dels continguts del disc, que inclou els tipus de metadades i els valors *hash* de cada objecte.
- Metadades relacionades amb el procés de creació d'imatges.
- Fotografies dels ítems.

Hi ha casos en què l'arxiver digital fa una captura selectiva de materials, com per exemple un fitxer de text concret. En aquests casos, s'utilitza un dispositiu USB amb *write-blocking* incorporat i programari per crear imatges de disc, amb un disc dur extern que serveix per emmagatzemar les dades en trànsit. Un cop les dades s'han incorporat al repositori, el seu procés tècnic és similar al descrit més amunt.

Per altra banda, BEAM utilitza eines d'anàlisi forense que permeten identificar problemes amb els objectes digitals, quins formats hi estan presents i preparar l'arxiu per al seu tractament tècnic. L'usuari pot fer cerques per format, per atributs de l'arxiu, per etiquetes o també fer cerques de text complet. Els catalogadors tenen un paper important dins l'edició de metadades, ja que l'arxiver només incorpora camps bàsics, però s'han d'incloure altres de descripció com les restriccions d'accés, ítems duplicats o identificar els ítems que s'han de eliminar.

Els reptes que afronta la unitat per al futur són:

- Millora de les interfícies i de les eines per a arxivers, catalogadors i investigadors.
- Establiment de polítiques de preservació digital.
- Obrir vies d'adquisició de materials del web i del 'núvol' digital.
- Millorar les condicions d'accés als investigadors.

3.3. NATIONAL LIBRARY OF AUSTRALIA

Des de fa uns trenta anys, la National Library of Australia (NLA) ha experimentat un lent però important creixement en la seva col·lecció de material *born digital* que inclou disquets, CD-ROMs i DVD-ROMs. Fins fa poc, la biblioteca s'ocupava d'aquesta col·lecció manualment i cas per cas, però aquest sistema no era viable degut al constant increment de documents i pel gran volum de suports i de formats. Per tant, l'any 2008 l'entitat va iniciar el Digital Preservation Workflow Project amb els següents objectius:

- Diferenciar les col·leccions digitals de les col·leccions tradicionals.
- Migrar els continguts digitals dels suports físics abans de que aquests es deteriorin.
- Facilitar maquinari i programari per accedir als continguts de suports físics obsolets.

Com a resultat d'aquests esforços, es va produir l'aplicació Prometheus, un "procés escalable i semi-automatitzat per transferir dades des de suports físics a un sistema d'emmagatzematge de preservació digital[13]".

Figura 24. Unitat personalitzada per a la duplicació de CD-ROMs a la NLA

L'aplicació permet crear una imatge del contingut digital, verificar la seva integritat, fer l'edició de les metadades necessàries i transferir-la al dipòsit digital de preservació. El maquinari que s'utilitza consisteix en una unitat que permet la instal·lació de múltiples unitats de CD-ROM, DVD-ROM, dispositius USB, targetes de memòria o disquets de 3 ½ polzades. Aquesta unitat es pot connectar a qualsevol estació de treball del personal de la biblioteca que disposi d'una targeta de connexió SATA.

Tot el programari que s'utilitza en el procés de recuperació de contingut *born digital* és de lliure i gratuït i és el següent:

- Escanejat de virus: ClamAV[14]
- Creació d'imatges de disc: dd[15] i cdrdao[16]
- Verificació i càlcul del *checksum*: Jaxsum[17]
- Identificació de fitxers: DROID[18]

[13] *Prometheus. Digital preservation workbench.* Disponible a: <http://prometheus-digi.sourceforge.net/> [data de consulta: 8 abr. 2013]

[14] Programa disponible a: < http://www.clamav.net/lang/en> [data de consulta: 5 maig 2013]

[15] Programa disponible a: <http://www.chrysocome.net/dd> [data de consulta: 5 maig 2013]

[16] Programa disponible a: <http://cdrdao.sourceforge.net> [data de consulta: 5 maig 2013]

[17] Programa disponible a: <http://sourceforge.net/projects/jacksum> [data de consulta: 5 maig 2013]

- Validació de formats de fitxer: JHOVE[19]
- Extracció de metadades: NLNZ Metadata Extractor[20].

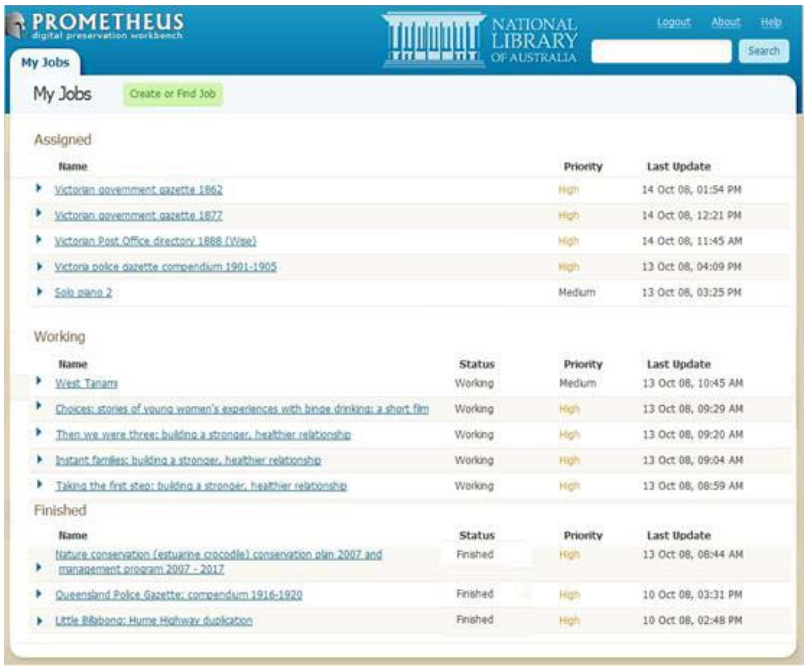


Figura 25. Llistat de tasques a Prometheus

Prometheus, que utilitza una interfície web, ja s'utilitza per part dels catalogadors de la NLA. El seu *workflow* és, a grans trets, el següent:

- Crear una nova tasca.
- Afegir el material *born digital* a la tasca, amb suport per a diversos formats.
- Assignar identificadors persistents, generar la imatge de disc amb el *checksum* corresponent.
- La imatge de disc es copia a l'estació de treball i es verifica la transferència amb el *checksum*.
- La imatge es munta i es desempaqueta el sistema de fitxers i els fitxers.
- Es genera un fitxer METS que documenta el mapa del sistema i l'estructura de directoris.

[18] Programa disponible a: <http://droid.sourceforge.net> [data de consulta: 5 maig 2013]

[19] Programa disponible a: <http://sourceforge.net/projects/jhove> [data de consulta: 5 maig 2013]

[20] Programa disponible a: <http://meta-extractor.sourceforge.net> [data de consulta: 5 maig 2013]

- L'antivirus analitza els fitxers per tal d'evitar la presència de programari maliciós.
- DROID i JHOVE s'ocupen de donar informació dels formats de fitxer i NLNZ Metadata Extractor genera les metadades amb format METS i PREMIS.
- Es guarda la informació de l'anàlisi de fitxers a la base de dades.
- Si no s'han de crear més imatges de disc, es passa al pas següent.
- Es comprova que s'hagin capturat les metadades correctament.
- Es confirma que s'hagi completat la tasca.
- Es fa la ingesta del contingut al Digital Object Storage System (DOSS) de la NLA, que tindrà dues còpies del contingut: una serà la imatge de disc i l'altra, els fitxers juntament amb el sistema de fitxers que s'han extret.

Amb data de juliol de 2011, la NLA té 589.030 fitxers emmagatzemats al DOSS provinents de material *born digital* de suports físics.

3.4. YALE UNIVERSITY

La Yale University compta amb diverses seccions per gestionar les seves col·leccions, però la que té una quantitat més important de contingut *born digital* és la de Manuscripts and Archives, fundada l'any 1969[21]. L'abast temàtic del seu fons és ampli, ja que adquireix documentació sobre història, arquitectura, ciència, medicina i cultura.

Com que la Yale University forma part del projecte AIMS, algunes dades sobre les seves col·leccions són accessibles al repositori Hypatia. Es tracta de les següents:

- Arxiu de James Tobin. Guanyador d'un Premi Nobel i professor d'economia, la seva col·lecció inclou 25 disquets de 3 ½ polzades. A Hypatia es pot accedir a un fitxer EAD (Encoded Archival Description) que inclou referències als disquets. No es facilita accés públic al material, però els investigadors en poden demanar còpies.

[21] *About Manuscripts and Archives: introduction*. Disponible a: <http://www.library.yale.edu/mssa/about_intro.html> [data de consulta: 7 abr. 2013]

- Arxiu de Henry Ashby Turner Jr. Historiador i acadèmic, dins la seva col·lecció hi ha nombroses dades de recerca en format de bases de dades Microsoft Access i Filemaker Pro. Es pot accedir el fitxer EAD a Hypatia.
- Arxiu de James Welch. Poeta, novel·lista i professor especialitzat en els nadius americans, el seu material *born digital* inclou esborranys de les seves obres.
- Fundació 'Love makes a family'. Creada per promoure la igualtat d'oportunitats entre els col·lectius gais, bisexuals i transsexuals per crear famílies, el seu arxiu *born digital* compta amb 36 GB que inclouen correus electrònics, fitxers, materials audiovisuals, fotografies, pàgines web i contingut de xarxes socials. Es pot accedir al fitxer EAD a Hypatia.

El *workflow* que s'aplica és el següent:

- Etiquetatge. Els suports físics, un cop adquirits, són etiquetats individualment i es generen metadades sobre la seva descripció física i el contingut dels seus fitxers. Aquestes metadades es guarden dins una base de dades que rep el nom de Media Log.
- Transferència dels continguts. El suport es connecta a una estació de treball local, utilitzant un *write-blocker*, i es genera una imatge de disc. Si es tracta d'un disquet, s'ha d'utilitzar un dispositiu especial com KryoFlux o FC5025 *(*vegeu *5.2.2. Targetes controladores).*
- Validació de les dades. Es genera un valor *hash* i es confirma que les dades de la imatge de disc són correctes.
- Extracció de metadades. El programari d'anàlisi forense fiwalk[22] reconeix el format dels fitxers, genera valors *hash* per a cadascun dels fitxers trobats al suport físic, comprova que no hi hagin virus i genera un fitxer xml amb els resultats finals.
- Ingesta. El fitxer xml i la imatge de disc són transferits al repositori Rescue, que no permet l'accés públic als materials, ja que només s'hi pot accedir internament i amb els permisos corresponents. Aquest procés d'ingesta consta de quatre etapes:

[22] Programa disponible a: <http://afflib.org/software/fiwalk> [data de consulta: 10 des. 2012]

- Verificació i validació del fitxer. Per verificar el fitxer, s'utilitza el programa JHOVE, que genera el *checksum* identificador únic per al fitxer. Un cop s'ha verificat el fitxer ara s'ha de validar; en aquest cas, JHOVE comprova el tipus de fitxer. Per exemple, un fitxer .jpg podria tenir un contingut que no es correspongui amb una imatge.
- Copiar el fitxer al repositori Rescue. Un cop s'ha completat amb èxit la verificació i la validació, es fa una còpia del fitxer dins Rescue.
- Post-verificació del fitxer. Un cop la còpia s'ha completat, JHOVE fa una segona verificació per comprovar que el *checksum* que s'ha generat prèviament correspon amb el fitxer.
- Generació de fitxer de registre. Un cop s'ha fet la post-verificació, es genera un fitxer de registre que detalla les tasques que s'han fet durant el procés d'ingesta del fitxer.

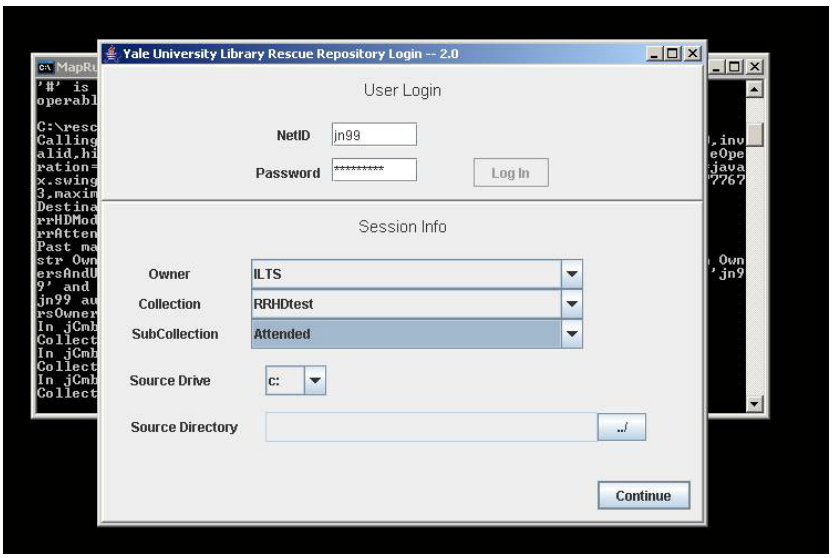


Figura 26. Pantalla d'inici de sessió del repositori Rescue

3.5. EMORY UNIVERSITY

Aquesta universitat situada a l'estat de Geòrgia, als EUA, mereix un especial esment, ja que la seva secció Manuscript, Archives, and Rare Book Library (MARBL) va rebre l'any 2010 la donació de l'arxiu personal del famós escriptor britànic Salman Rushdie, que

no només contenia cartes, manuscrits o llibretes, sinó també un disc dur i quatre ordinadors, que presentaven les següents característiques:

- Un ordinador Macintosh Performa 5400/180
- Un ordinador portàtil Macintosh PowerBook 5300c
- Dos ordinadors portàtils Macintosh PowerBook G3
- Un disc dur portàtil de 60 GB

Figura 27. Ordinadors donats per Salman Rushdie

L'adquisició d'aquests aparells va motivar la creació d'un grup de treball específic per tractar amb aquests materials *born digital,* el Born-Digital Archives Working Group, que va rebre la missió de trobar la millor manera de presentar i preservar la informació dins d'aquests dispositius, a més de respectar la privacitat de l'escriptor; la documentació inclosa als dispositius contenia dades sensibles, com informació financera i personal de Rushdie i de la seva família.

El personal assignat, de tres arxivers i de tres enginyers informàtics, van acordar tractar el material com a col·lecció híbrida, formada per materials físics i digitals i van decidir utilitzar la següent política de preservació i accés:

- Emular la interfície d'accés original.
- Permetre consultar l'estructura i la descripció (sèries, subsèries, tipus de contingut, etc.) de la col·lecció a la base de dades, però no el seu contingut.
- Accés restringit a la interfície emulada dins les instal·lacions de la universitat.
- El contingut només apareixerà a l'emulació.

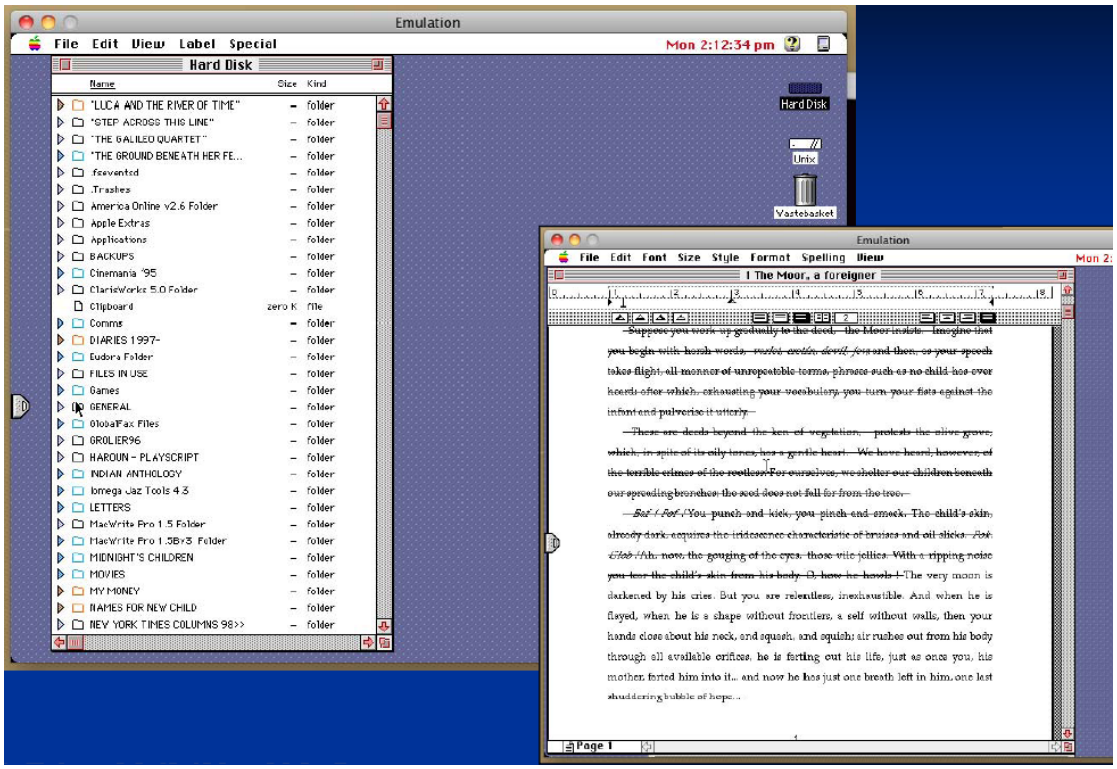


Figura 28. Emulació del sistema operatiu del Macintosh Performa 5400

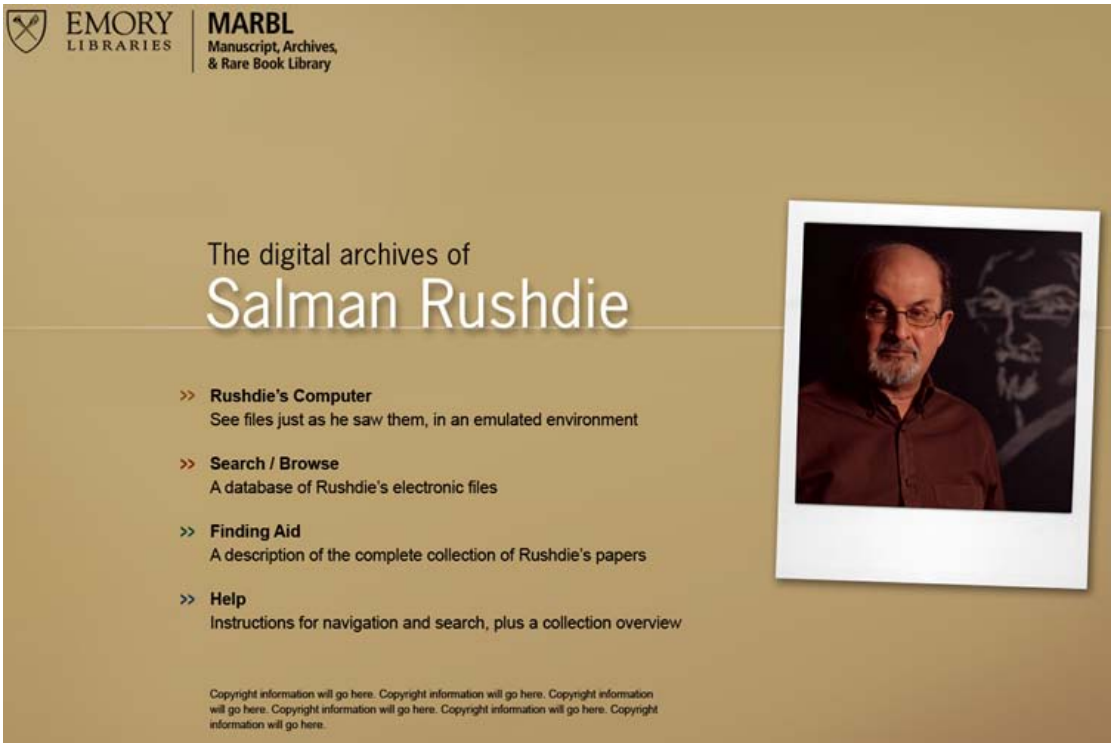


Figura 29. Base de dades de l'arxiu digital de Salman Rushdie

Només els investigadors en tenen accés a l'arxiu. Poden fer les seves recerques de dues maneres:

- Entorn emulat. L'investigador pot obrir els fitxers que es trobaven als ordinadors originals en un entorn que simula el que feia servir Salman Rushdie i pot accedir als continguts originals.
- Base de dades. L'investigador navega entre l'estructura de directoris dels ordinadors. Es pot consultar els tipus de fitxers, la seva extensió i es poden fer cerques, però no a text complet.

Al concloure el projecte, el grup de treball va avaluar els processos per tal d'evitar problemes en el futur i trobar millores. Les seves conclusions foren:

- És necessària la col·laboració entre professionals de diferents àmbits.
- És important consultar amb els donants quina és la informació sensible que s'ha de restringir a l'accés públic.
- S'ha fer una avaluació i tria dels materials curosa.
- S'han de definir els processos i *workflows*.
- S'han de desenvolupar les eines de treball amb assessorament de professionals de diferents àmbits.

Encara que l'aproximació de l'Emory University vers el material *born digital* va ser molt diferent dels altres centres, ja que no es van utilitzar eines d'anàlisi forense estrictament per fer l'emulació, la seva solució s'ha de tenir en compte per la seva intenció de preservar el 'look-and-feel' de l'entorn de treball de Salman Rushdie. De fet, aquesta experiència va servir per què la institució s'adonés de la importància del tractament d'aquest material i ja s'està treballant per a la creació d'un laboratori d'anàlisi forense al centre.

4. REQUERIMENTS ESPECÍFICS DE LA UNITAT

Per tal de dissenyar la unitat i fixar quins elements de maquinari i programari es requereixen, s'han plantejat dos escenaris: un de nivell mínim, on s'exposen quins són els elements mínims que necessitaria la unitat per recuperar i analitzar dades i un altre de nivell avançat, on s'exposen les prestacions màximes existents actualment per l'anàlisi forense digital.

4.1. NIVELL BÀSIC

En funció de l'abast de sistemes descrits a l'apartat 2.2., els requeriments mínims d'aquesta unitat serien:

- Suport per a interfícies SATA.
- Suport per unitats de disquet de 3 ½ i de 5 ¼ polzades.
- Connectivitat a la xarxa d'Internet.
- Connectivitat per a diversos dispositius USB.
- Duplicació i anàlisi de diversos sistemes de fitxers.
- Validació de la integritat d'imatges de disc i de fitxers.

Una estació de treball amb aquests requeriments pot presentar la següent configuració de maquinari:

- Torre amb capacitat per a connectar un mínim de quatre unitats SATA.
- Placa base amb suport per a SATA i sis dispositius USB.
- Processador Intel.
- Compartiment per a connectar tres unitats extraïbles.
- Disquetera interna de 3 ½ polzades.
- Disquetera interna de 5 ¼ polzades.
- Targeta controladora per a disqueteres.
- Disquetera Zip externa.

- Disc dur intern.
- Targeta Ethernet 10/100.
- Lector de CD-ROM/DVD-ROM.
- *Docking station* amb interfície IDE i SATA per connexió USB.
- *Write-blocker* extern.

Figura 30. Torre d'ordinador amb compartiments externs

El programari per a anàlisi forense pot ser gratuït, ja que existeixen eines que donen resultats adients als objectius de la unitat. Per tant, la configuració que es requereix seria d'aquest tipus (per a més detalls sobre el programari, vegeu *5.1. PROGRAMARI*).

- Arrencament dual amb Windows XP i Windows 7. En alguns casos es requereix utilitzar programari compatible amb MS-DOS (la qual cosa es pot fer amb Windows XP) i en altres, es requereix el Windows 7 degut a requeriments del programari forense. És perfectament possible fer una instal·lació dels dos sistemes operatius en un o més discs durs.
- Creador d'imatges FTK Imager (AccessData).

- Eina d'anàlisi forense digital Autopsy[23] (The Sleuth Kit).
- Identificador de fitxers DROID (The National Archives).
- Generador de *cheksums* MD5summer[24].
- Antivirus avast![25] (Avast Software).
- Validador de formats de fitxer JHOVE (JSTOR i Harvard University Library).

Opcionalment, el maquinari pot presentar les següents característiques:

- Disquetera Zip interna (obligatori si no es disposa de disquetera externa).
- *Write-blocker* intern (obligatori si no es disposa de *Write-blocker* extern).

Una segona opció és utilitzar les estacions de treball ja instal·lades a la institució i només adquirir el maquinari necessari per a l'adquisició d'imatges de disquets i de discs durs. En aquest cas, caldrà aconseguir els següents accessoris:

- Targeta controladora Kryoflux.
- Disquetera externa de 3 ½ polzades.
- Disquetera externa Zip de 750 GB.
- Disqueteres internes de 3 ½ i de 5 ¼ polzades.
- Docking-station amb interfície IDE i SATA per connexió USB.
- *Write-blocker* extern.

Juntament amb la instal·lació del programari gratuït, això seria suficient per superar l'obstacle principal que representa el suport físic: l'accés a les dades.

[23] Programa disponible a: <http://www.sleuthkit.org/autopsy> [data de consulta: 6 abr. 2013]

[24] Programa disponible a: <http://www.md5summer.org> [data de consulta: 6 abr. 2013]

[25] Programa disponible a: <http://www.avast.com/es-es/index> [data de consulta: 6 abr. 2013]

4.2. NIVELL AVANÇAT

Existeixen diverses solucions integrades al mercat per fer anàlisi forense, sense necessitat de complexes instal·lacions, i que integren el maquinari i el programari adient per a investigacions avançades. L'opció més adient (que ja estan utilitzant altres institucions del Regne Unit i els EUA) seria una unitat **FRED** (Forensic Recovery of Evidence Device), un sistema dissenyat i optimitzat per a l'anàlisi forense. És semblant a una torre convencional, però disposa de tots els recursos necessaris ja integrats de sèrie. Per exemple, per analitzar un disc dur només cal connectar-lo directament a una de les safates habilitades a tal efecte, sense haver d'utilitzar dispositius externs de connexió (per a detalls tècnics, vegeu *5.2.2. Estacions de treball forense*). Les opcions de configuració són molt àmplies i contemplen la possibilitat de connectar diversos discs durs a la vegada i amb diferents tipus de processador, en funció de les necessitats del client. Aquesta opció seria molt adient per a una unitat estàtica, però també es pot fer esment d'una altra, que té encara millors prestacions:

Figura 31. Unitat FRED SR

FRED SR (amb processador Dual Xeon) és l'estació de treball més potent entre totes les que ofereix Digital Intelligence. Permet connectar directament discs durs IDE, EIDE, ATA, SATA, ATAPI, SAS, Firewire i USB i memòries USB; gravació d'imatges forenses a Blu-Ray, DVD, CD o discs durs i adquisició de dades de Blu-Ray, CD-ROM, DVD-ROM i targetes de memòria.

A més de les prestacions d'una unitat FRED convencional, FRED SR també inclou una capsa on es troben:

- Un CD-ROM de recuperació del sistema.
- Claus de sistema per a les portes i els discs durs extraïbles.
- Adaptadors i cables.
- Càmera digital per documentar el maquinari.
- Equip de tornavisos per si és necessari obrir torres informàtiques per extreure discs durs.

En quant al programari, s'ha de contemplar la compra de llicències comercials, les quals contenen funcions més avançades que els programes gratuïts (per a detalls tècnics, vegeu *5.1. PROGRAMARI*). Actualment les solucions d'anàlisi digital forense amb més implantació són les següents:

EnCase Forensic v7 (Guidance Software)

Solució integral per a l'adquisició, anàlisi i informe de dades digitals. Les seves funcionalitats inclouen:

- Creació d'imatges de disc forenses.
- Verificació d'imatges amb generació de valors *hash*.
- Automatització de tasques amb l'extensió EnScript.
- Sistema de cerca avançat que permet identificar dades irrecuperables amb altres aplicacions.
- Cerca per correus electrònics.
- Adquisició integrada de dades a mòbils i tablets.
- Personalització d'informes.
- Format propi d'imatge EnCase, acceptat com a prova criminal davant els tribunals.

Forensic Toolkit – FTK (AcessData)

Plataforma d'anàlisi forense digital amb les següents funcionalitats:

- Anàlisi de correus electrònics.
- Visualització de dades personalitzable.
- Desencriptació de contrasenyes.
- Mòdul Cerberus d'anàlisi de programari maliciós o *malware*.

5. RECURSOS NECESSARIS

Per a la creació de la unitat s'han d'adquirir múltiples recursos, alguns més complicats d'aconseguir que altres. S'han seleccionat solucions de maquinari i programari apropiades en funció de l'anàlisi a les unitats ja existents, amb el disseny dels perfils necessaris del personal, una aproximació a l'entorn de treball més adequat i els fluxos de treball més habituals.

5.1. PROGRAMARI

Dins la selecció de programari, no s'ha triat cap eina de gestió/extracció de metadades tècniques degut a que es tracta d'una tasca massa complexa per les diferents tipologies de fitxers i l'enorme varietat de metadades que es poden extreure. No obstant, la institució hauria de considerar com gestionar les metadades en funció de les característiques de la seva col·lecció.

5.1.1. Creació d'imatges de disc

Dins d'aquesta secció s'analitzaran les eines existents més adients per a la creació d'imatges de disc (o imatges forenses) dels diferents suports físics acceptats a la unitat.

5.1.1.1. FTK Imager (AccessData)

Programari gratuït per a Windows XP i posteriors.

Funcionalitats

- Crea imatges de discs durs, disquets, discs Zip, CD-ROMs, DVD-ROMs, carpetes o fitxers individuals.
- Vista prèvia dels fitxers i carpetes a discs durs, disquets, discs Zip, CD-ROMs i DVD-ROMs.

- Munta la imatge per tal de visualitzar el contingut de la imatge exactament com l'usuari amb la unitat original.
- Exportar fitxers i carpetes d'imatges de disc.
- Veure i recuperar fitxers que s'han esborrat des de la paperera de reciclatge, però que encara no s'han sobreescrit a la unitat.
- Crear *hashes* de fitxers mitjançant dues funcions de *hash*: MD5 i SHA-1.
- Generar informes de *hashes* per a fitxer i per a imatges de disc per comprovar la integritat dels continguts. El *hash* és la prova que els fitxers no s'han alterat ni modificat en cap cas.

Característiques tècniques

Sistemes de fitxer acceptats:

- FAT 12, FAT 16, FAT 32
- NTFS
- HFS
- HFS+
- Ext2FS
- Ext3FS
- Ext4FS
- CDFS
- exFAT
- ReiserFS3
- VXFS

Formats d'imatge de CD i DVD acceptats:

- Alcohol (*.mds)
- CloneCD (*.ccd)
- Nero (*.nrg)
- ISO
- Virtual CD (*.vc4)
- VCD
- DVD+MRW
- DVD-RW
- DVD+RW Dual Layer
- BD-R SRM-POW
- BD-R SRM
- HD DVD-R
- IsoBuster CUE
- CD-ROM XA
- Roxio (*.cif)
- Pinnacle (*.pdi)
- CD-RW
- CD-ROM
- DVCD
- DVD-VFR
- BD-R DL
- HD DVD-RW DL
- HD DVD-RW
- SVCD
- PlexTools (*.pxi)
- DVD+VR
- DVD+R DL
- DVD-VRW
- HD DVD-R DL
- HD DVD
- DVD-RAM
- CD-MRW
- DVD+R
- BD-RE
- BD-ROM
- BD-R RRM

- VD-R
- DVD+RW

Formats d'imatge de disc dur acceptats:

- EnCase, inclosa la versió 6.12
- Safeback 2.0
- Linux DD
- Ghost (nómes imatges forenses)
- Imatge lògica AccessData (AD1)
- SnapBack
- Expert Witness
- ICS
- SMART
- Advanced Forensics Format (AFF)

Proves

Per provar el programari s'ha creat una imatge de disc amb un pendrive Toshiba d'1 GB, mitjançant les opcions *Create Disk Image – Physical Drive* (en el cas de suports inserits a unitats internes com CD-ROMs o DVD-ROMs, es faria servir l'opció *Logical Drive*). Indicarem la carpeta de destinació i ens crearà una imatge del tipus a escollir entre les opcions raw, SMART, E01 i AFF.

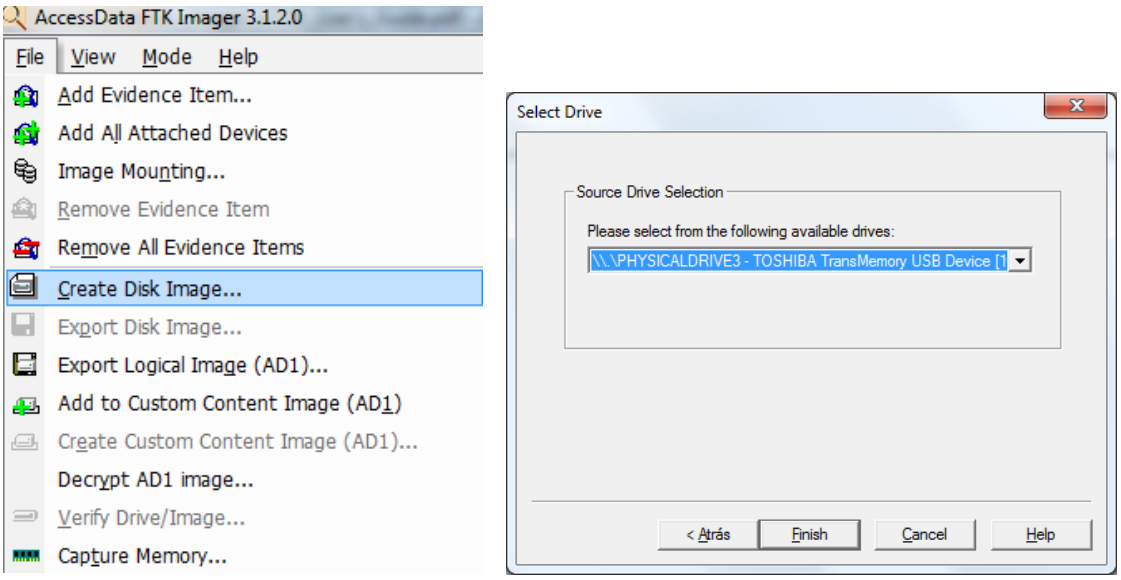


Figura 32. Creació d'imatge de disc amb FTK Imager

Un cop creada la imatge, proporcionarà el següent informe de verificació de *hash* de la imatge, juntament amb altra informació com el temps que s'ha trigat en adquirir la imatge, quin era el tipus de suport original, etc.:

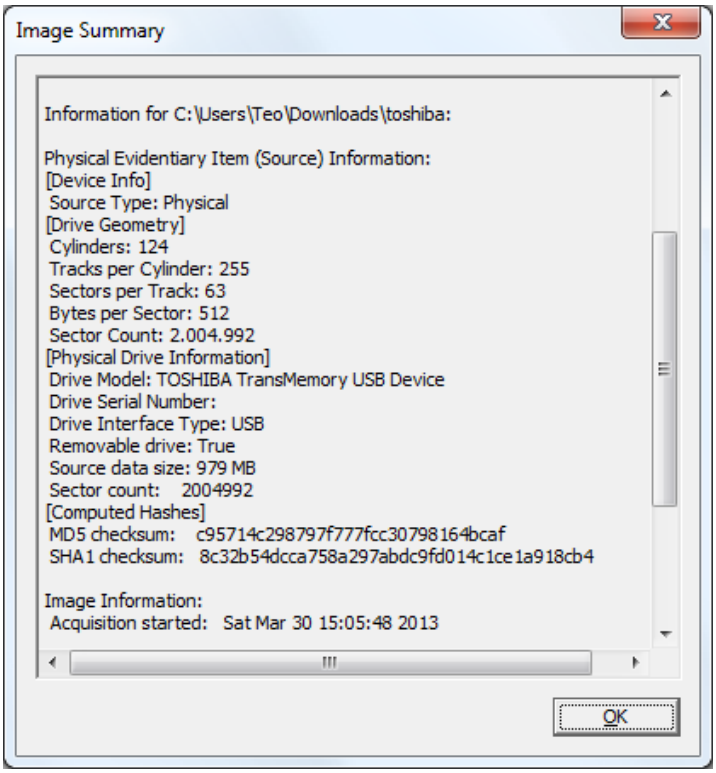


Figura 33. Informe de la imatge amb *hash* MD5 i SHA1 en FTK Imager

Es pot obrir posteriorment aquesta imatge i visualitzar l'estructura del sistema de fitxers, amb les seves particions originals i els fitxers i directoris esborrats, que es troben marcats amb una creu.

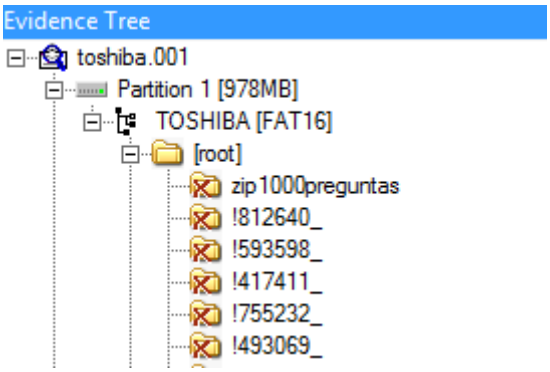


Figura 34. *Evidence Tree* en FTK Imager

Dins *Evidence Tree*, és possible navegar dins l'estructura de fitxers per seleccionar fitxers i directoris concrets, o bé seleccionar tota l'estructura i extreure-la a una destinació concreta. Això mateix es pot fer també per extreure *hash* de fitxers i directoris.

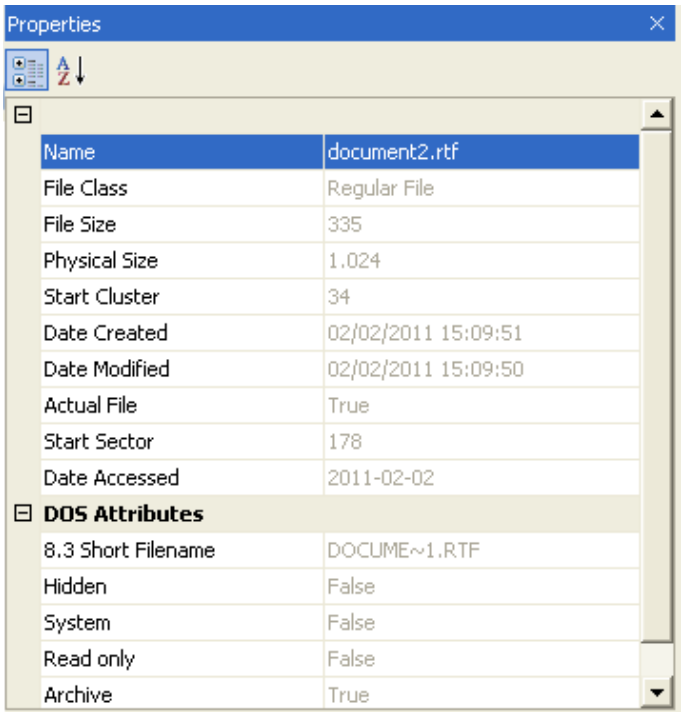


Figura 35. Opció de propietats de fitxer en FTK Imager

Dins *Properties*, tenim accés a diferents metadades com la data que es va crear el fitxer, la data de modificació, la mida del fitxer, etc. Mancarien dades més concretes com el format de codificació del fitxer, que en aquest cas és el RTF, un format propietari de Microsoft.

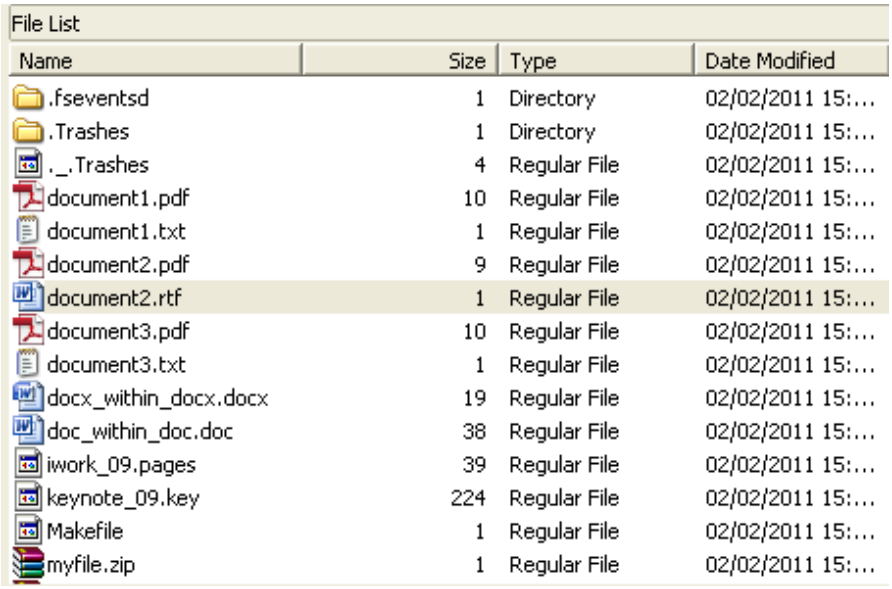


Figura 36. Visualització de la llista de fitxers en FTK Imager

Per últim, *File List* ens permet navegar fitxer per fitxer i la finestra inferior ens permetrà veure el contingut, bé directament (si es tracta dels formats PDF o text simple en ASCII) o bé mitjançant el visor hexadecimal, on també podem fer cerques de text o explorar el contingut. Com que no sempre és possible que la institució tingui el programari comercial instal·lat en certs casos, aquesta és una bona opció per recuperar dades.

Un exemple seria el fitxer doc_within_doc.doc, que conté dades de dos correus electrònics. Al visor hexadecimal podem fer una cerca per .com (un domini habitual), que ens dóna el següent resultat:

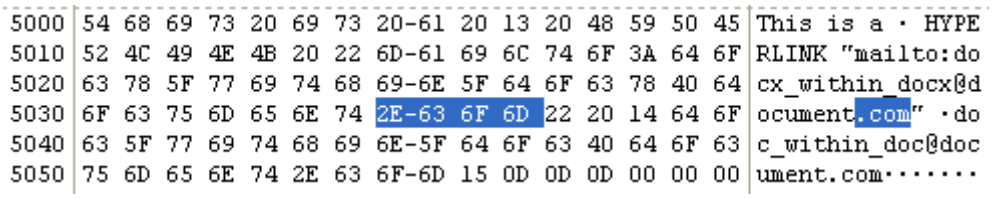


Figura 37. Visor hexadecimal a FTK Imager

Conclusions

Aquesta eina és fonamental per a la creació d'imatges de disc, ja que, encara que existeixen altres eines semblants, aquesta permet crear imatges per a anàlisi forense, la qual cosa permet tenir informació com fitxers i directoris esborrats, sistema de fitxers original, nombre de particions o informació del suport original (en aquest cas, una unitat USB extraïble).

5.1.1.2. KryoFlux DiskTool Console (DTC)

Programari gratuït per a ús exclusiu amb la targeta controladora Kryoflux (vegeu *5.2.2. Targetes controladores*). Compatible amb Windows (XP o superior), Mac OS X (10.5 o superior) i Linux. A Windows, és necessari utilitzar l'aplicació 'Símbol del sistema', semblant a l'antic sistema operatiu MS-DOS.

Funcionalitats

- Creació d'imatges de disquets de 3 ½ i de 5 ¼ polzades.
- Suport per a múltiples formats d'imatge.
- Permet la lectura i l'escriptura d'imatges de disc en disquets.
- Permet la lectura i escriptura amb disquets de simple, doble i alta densitat.
- Creació d'imatges de disc per tal de ser utilitzats en emuladors.

Característiques tècniques

Lectura dels següents tipus d'imatge de disc:

- Fitxers que segueixen el protocol propi de Kryoflux.
- CT Raw, 84 pistes, simple i doble densitats, 300 revolucions, codificació en MFM.
- Imatge en sectors FM, 40/80 pistes, simple i doble densitats, 300 revolucions, codificació en FM.
- FM XFD (Atari 8 bits).
- Imatge en sectors MFM, 40/80+ pistes, simple, doble i alta densitats, 300 revolucions, codificació en MFM.
- MFM XFD (Atari 8 bits).
- Imatge en sectors AmigaDos, 80+ pistes, doble i alta densitats, 300 revolucions, codificació en MFM.
- Imatge en sectors CBM DOS, 35+ pistes, doble densitat, 300 revolucions, codificació en GCR.
- Imatge en sectors Apple DOS 3.2, 35+ pistes, doble densitat, codificació en GCR.
- Imatge en sectors Apple DOS 3.3+, 35+ pistes, doble densitat, codificació en GCR.
- DSK, DOS 3.3.
- Imatge en sectors Apple DOS 400K/800K, 80+ pistes, doble i alta densitats, codificació en CLV i GCR.
- Imatge en sectors Emu, 35+ pistes, simple i doble densitats, 300 revolucions, codificació en FM.

- Imatge en sectors Emu II, 80+ pistes, simple i doble densitats, 300 revolucions, codificació en FM.
- Imatge en sectors Amiga DiskSpare, 80+ pistes, simple, doble i alta densitats, 300 revolucions, codificació en MFM.
- Imatge en sectors DEC RX01, 77+ pistes, baixa densitat, 360 revolucions, codificació en FM.
- Imatge en sectors DEC RX02, 77+ pistes, baixa i doble densitat, 360 revolucions, codificació en FM i MFM.

La lectura de disquets en codificació FM i MFM permet el suport per a un gran nombre de sistemes informàtics. Per exemple, les unitats de disquet de 3 ½ polzades utilitzen el codificat MFM.

Escriptura dels següents tipus d'imatge de disc:

- IPF (Interchangeable Preservation Format), format obert per a diverses plataformes.
- ADF, format per als ordinadors Amiga.
- CBM G64, format per als ordinadors Commodore CBM.

Proves

Per provar el programari es va crear una imatge del contingut d'un disquet de 5 ¼ de baixa densitat (360 KB), el qual contenia fitxers DRW, que s'utilitzaven originalment amb el programa de presentacions Lotus Freelance Graphics (IBM). Es va crear la imatge amb el següent comandament:

Figura 38. Creació d'imatge de disc amb DTC

En aquest cas es van utilitzar les següents opcions:

- -d<id>: número de la unitat (en aquest cas 0, ja que només havia una unitat connectada).

- -g<side>: ús de disquets amb una sola cara (en aquest cas 0, ja que només hi havia una cara gravada).
- -f<name>: nom de fitxer. S'ha indicar tant nom de fitxer com d'extensió.
- -i<type>: tipus d'imatge. Aquí s'ha indicat el tipus d'imatge en sectors MFM, 40/80+ pistes, simple, doble i alta densitats, 300 revolucions, codificació en MFM.

Un cop s'introdueix el comandament, el programa fa la lectura del disquet pista per pista. Si troba errors de lectura, farà diversos intents per recuperar la informació i minimitzar possibles pèrdues de dades.

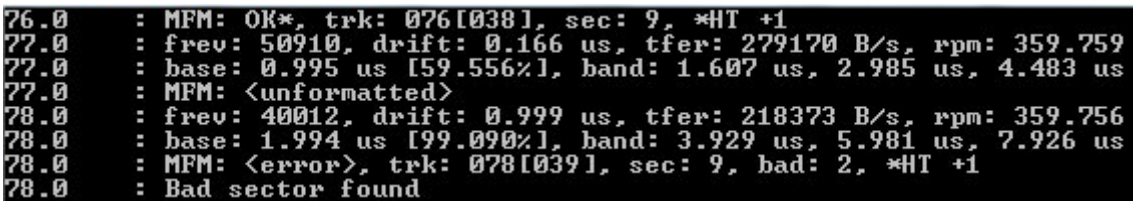


Figura 39. Lectura de pistes de disquet amb DTC

Algunes pistes contenien errors i per tant algunes dades no es van poder recuperar. Posteriorment el fitxer .img es pot obrir i analitzar amb FTK Imager.

Conclusions

Si la unitat no té una unitat de disquet incorporada, aquest programari és realment imprescindible per aconseguir imatges de disc fiables i que preservin el contingut original, sempre i quan es faci servir la targeta Kryoflux. Les seves nombroses opcions fan aquest programari molt interessant si es treballa amb una quantitat important de disquets, però el seu ús és molt poc intuïtiu i requereix tenir coneixements tècnics avançats per al seu aprofitament òptim, com la capacitat d'emmagatzematge dels discs, el sistema de codificat, etc.

5.1.2. Generació de *checksums*

Hi ha casos en què és necessari l'ús d'un programari específic per verificar la integritat de les dades i assegurar que no s'hagi fet cap modificació dins el material *born digital* que hagi adquirit la institució. Per tant, ha de ser capaç de crear i verificar els *hashes* necessaris MD5 i SHA1 que identifiquen una imatge forense.

5.1.2.1. MD5summer

Generador de *checksums* per a Windows de codi obert i gratuït.

Funcionalitats

- Creació de *hashes* MD5 i SHA1 amb múltiples fitxers i carpetes.
- Verificació de *hashes* MD5 i SHA1 amb múltiples fitxers i carpetes.

Proves

Per provar el programari s'han aprofitat els *hashes* ja generats amb FTK Imager per verificar que les dades coincideixen si es verifica novament la imatge de disc del pendrive Toshiba amb MD5summer, el fitxer toshiba.001. Partint que la imatge té els següents *checksums*:

- MD5 checksum: c95714c298797f777fcc30798164bcaf
- SHA1 checksum: 8c32b54dcca758a297abdc9fd014c1ce1a918cb4

Amb MD5summer, primer s'ha de seleccionar el tipus de *checksum* (MD5 o SHA1) i després es selecciona el fitxer toshiba.001, que és la imatge de disc creada amb FTK Imager:

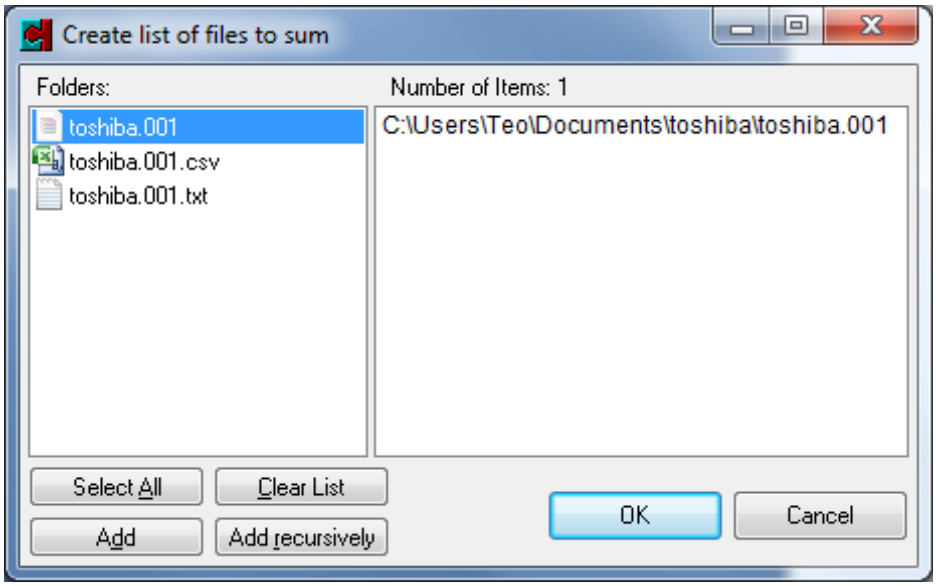


Figura 40. Selecció de fitxers i carpetes a MD5summer

A continuació, es generarà el *checksum* MD5 i es veu que coincideix amb el que va generar FTK Imager:

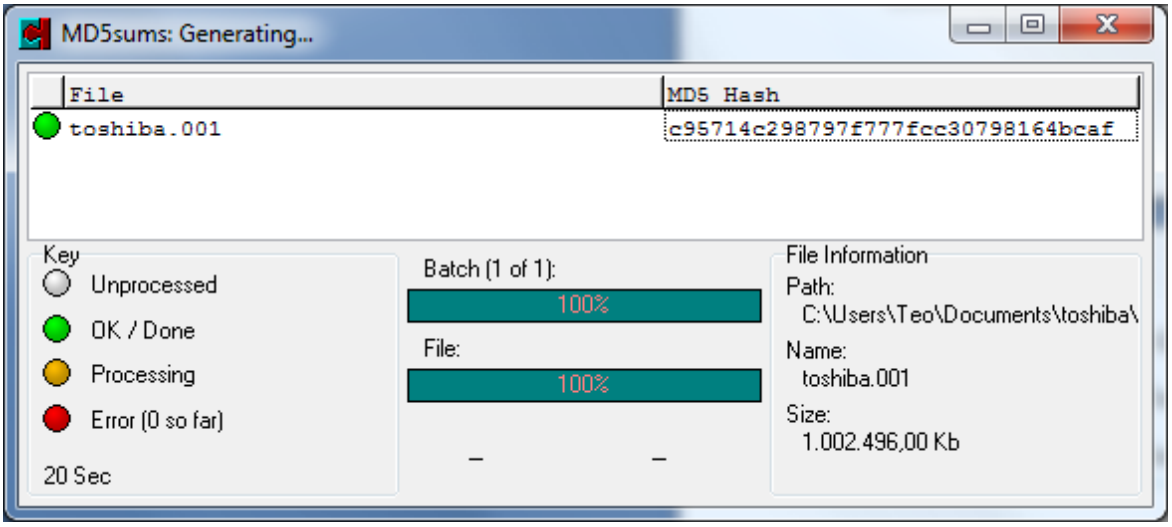


Figura 41. Generació de *checksums* a MD5summer

Aquest codi es pot guardar després en un fitxer amb extensió MD5; el nom de fitxer serà el mateix que el de la imatge forense original. En aquest cas, toshiba.md5. Si es volgués crear i/o verificar el *hash* SHA1, s'hauria de repetir el procés, seleccionant el tipus de *checksum* corresponent.

Conclusions

Aquest programa és òptim per a generar *checksums* de múltiples fitxers, però té l'inconvenient que no permet crear codis MD5 i SHA1 al mateix temps; primer s'ha de fer una operació per aconseguir els MD5 i després una segona per aconseguir els SHA1. No obstant això, és molt ràpid i fiable i la seva capacitat per treballar amb múltiples fitxers i carpetes és ideal si, a més de la imatge de disc, s'han de generar *hashes* de fitxers amb contingut de text, so, etc.

5.1.2.2. HashX (BoilingBit)[26]

Generador de *checksums* per a Windows gratuït.

Funcionalitats

- Creació de *hashes* dels tipus CRC32, GOSTHash, MD2/MD4/MD5 i SHA1,SHA2-256/384/512 en formats amb lletres minúscules (amb espais i sense espais) i amb lletres majúscules (amb espais i sense espais).
- Comparació de *hashes* dels tipus CRC32, GOSTHash, MD2/MD4/MD5 i SHA1,SHA2-256/384/512 en formats amb lletres minúscules (amb espais i sense espais) i amb lletres majúscules (amb espais i sense espais).

Proves

S'ha tornat a utilitzar la imatge de disc toshiba.001 per fer les proves. Val a dir que HashX no té la funció de generar *hashes* amb múltiples fitxers i carpetes, sinó només amb fitxers individuals, però sí permet treballar amb més tipus de *checksums*. En aquest cas, es va provar la generació de *hash* SHA1, el qual està format pel text 8c32b54dcca758a297abdc9fd014c1ce1a918cb4. Un cop es generà, es va comprovar que coincidia copiant i pegant el text a la casella 'Compare with' i efectivament, coincidia.

[26] Programa disponible a: <http://www.boilingbit.com/products/hashx> [data de consulta: 28 abr. 2013]

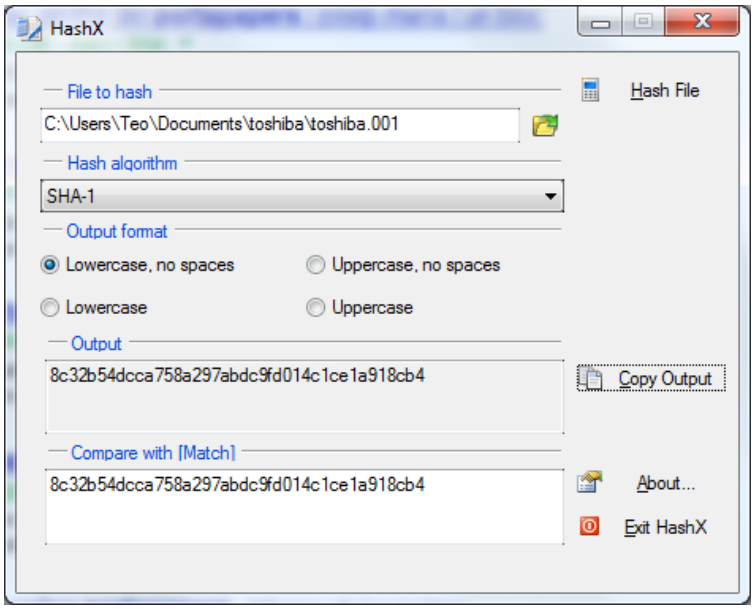


Figura 42. Generació de *checksum* SHA1 amb HashX

Les diferències entre els diferents tipus de checksums es poden veure a la següent taula:

Taula 1. Comparativa de *checksums*

Tipus de *checksum*	**Valor *hash* del fitxer toshiba.001 (979 MB)**
CRC32	3d723207
GOSTHash	09bdc5f1d1a27e82b5ff874cfafa5f994d881f22744ed1de3c337901b4ca34e0
MD2	85990829f53e017fb3a439efc9b68a08
MD4	5703cd989eae3cbf1e7fccad2a90fa35
MD5	c95714c298797f777fcc30798164bcaf
SHA1	8c32b54dcca758a297abdc9fd014c1ce1a918cb4
SHA-2 256	d2a9f9bb3c09c34153cd4badecbd89480bd631e417148037a07e518ae87c9b27
SHA-2 384	498ad484053fc116d14b924f7b0ff1fa07dbd43bb039d955bb57b68a92cd241e27 2a28c5d59133e3da817c18a6e2ba56
SHA-2 512	648cc071c06377c25bcf9759396af9f382f8088fdfe966a6a9bef3ad7f9f82ba4356 1328ecae16c9d840e3c39f1b01a5b49e93d1709ef199709ff113928bb780

El més segur de tots seria el SHA-2 512, que genera paraules de 64 bits i és quasi impossible de replicar. Una prova de la seva fiabilitat és la que les diferents variants de

SHA-2 són actualment utilitzats a la majoria d'aplicacions informàtiques de les agències federals dels EUA[27].

5.1.3. *Suites* d'anàlisi forense digital

5.1.3.1. Autopsy (The Sleuth Kit)

Interfície gràfica d'anàlisi forense digital que forma part de la biblioteca d'eines forenses The Sleuth Kit per a Windows, Unix i Linux. Programari de codi obert i gratuït.

Funcionalitats

- Anàlisi d'imatges de disc en els formats:
  - RAW
  - E01
- Sistemes de fitxers acceptats:
  - NTFS
  - FAT12, FAT16, FAT32
  - HFS+
  - ISO9660
  - Ext2, Ext3
  - UFS
- Sistema de cerca per paraula clau.
- Anàlisi del registre.
- Extracció de metadades EXIF de fitxers JPEG.
- Visualització de l'activitat recent.
- Anàlisi de la línea de temps.
- Cerca automàtica d'adreces de correus electrònics.
- Reproducció de vídeo i imatges.
- Visualització de miniatures.

[27] *NIST's policy on hash functions.* Disponible a: <http://csrc.nist.gov/groups/ST/hash/policy.html> [data de consulta: 30 abr. 2013]

- Extracció de text de HTML, Office, PDF i RTF.
- Extracció de cadenes de text en Unicode en diversos idiomes (àrab, xinès, japonès, etc.)
- Marcadors.
- Generació d'informes.

Proves

S'ha fet un anàlisi amb la imatge de disc de la memòria USB Toshiba creada amb FTK Imager. El programa té una opció que permet la recuperació de dades, la localització de "fitxers orfes", que són fitxers que es van esborrar del dispositiu però el seu contingut encara es pot recuperar, ja sigui parcial o totalment. Aquests fitxers s'identifiquen amb una creu.

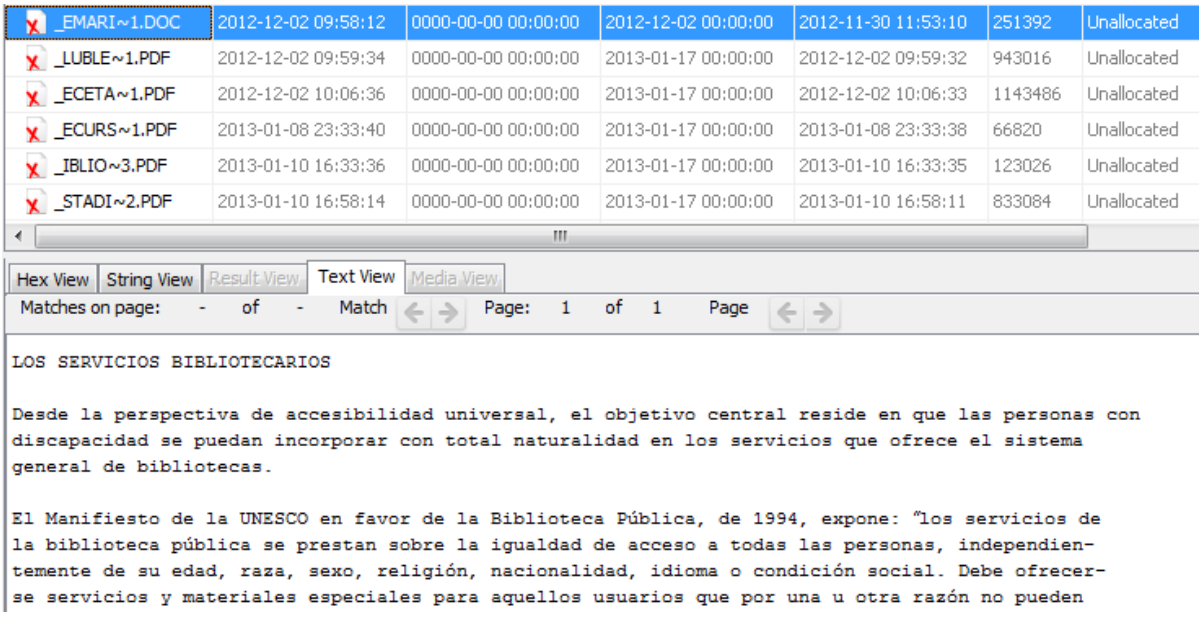


Figura 43. Visualització de fitxers orfes amb el text recuperat a Autopsy

Si és possible, Autopsy recuperarà el contingut esborrat, amb previsualització del text, la imatge o el vídeo. També permet l'extracció dels fitxers.

Els fitxers orfes s'han de tenir molt en compte quan arriben dades personals i sensibles, ja que poden revelar informació que el donant pot creure esborrada, però

que encara es troba present i es pot recuperar. A tall d'exemple, una cerca amb Autopsy amb la paraula clau 'DNI' va recuperar 54 resultats, tots de fitxers orfes.

El programa també permet la visualització directa dels fitxers d'especial interès per a la unitat: els documents de text, les imatges, els vídeos i els àudios. En aquesta visualització també s'inclouen els fitxers orfes.

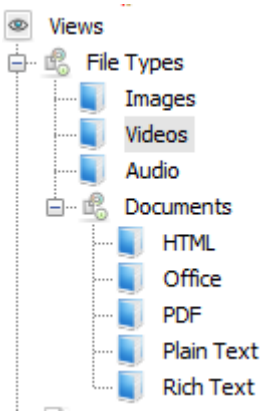


Figura 44. Visualització de tipus de fitxers a Autopsy

Per altra banda, es poden generar informes en HTML, Excel i text pla que permetrà tenir un registre de:

- Nom de cada fitxer present a la imatge de disc.
- Mida en bytes de cada fitxer present a la imatge de disc.
- Marcadors.
- Galetes (o *cookies*).
- Descàrregues d'Internet.
- Resultats de les cerques per paraula clau.
- Adreces de correu electrònic trobades.
- Metadades en format EXIF.
- Documents recents.
- Historial de pàgines web.
- Resultats de cerques a motors de cerca a Internet.

Conclusions

Tot i que es tracta d'un programari gratuït, les prestacions d'Autopsy són nombroses i molt útils per localitzar informació privada i sensible, la qual s'ha de tractar amb cura per tal d'evitar possibles perjudicis al donant. Les dades que pot amagar un disc dur o una memòria USB són nombroses, sobre tot amb les capacitats actuals (entre 1 i 2 TB els discs durs i els 16 i 64 GB les memòries USB) que faciliten molt que quedin fitxers orfes. Per tant, aquesta és una molt bona solució per a unitats amb un pressupost limitat.

5.1.3.2. AccessData Forensic Toolkit (FTK)

Programari comercial creat com a solució integral d'anàlisi forense digital per a Windows XP, Windows 7 i Windows Server 2003 i 2008. Una prova de la seva fiabilitat és que és acceptat el seu ús per a la creació de proves digital acceptades als tribunals de justícia.

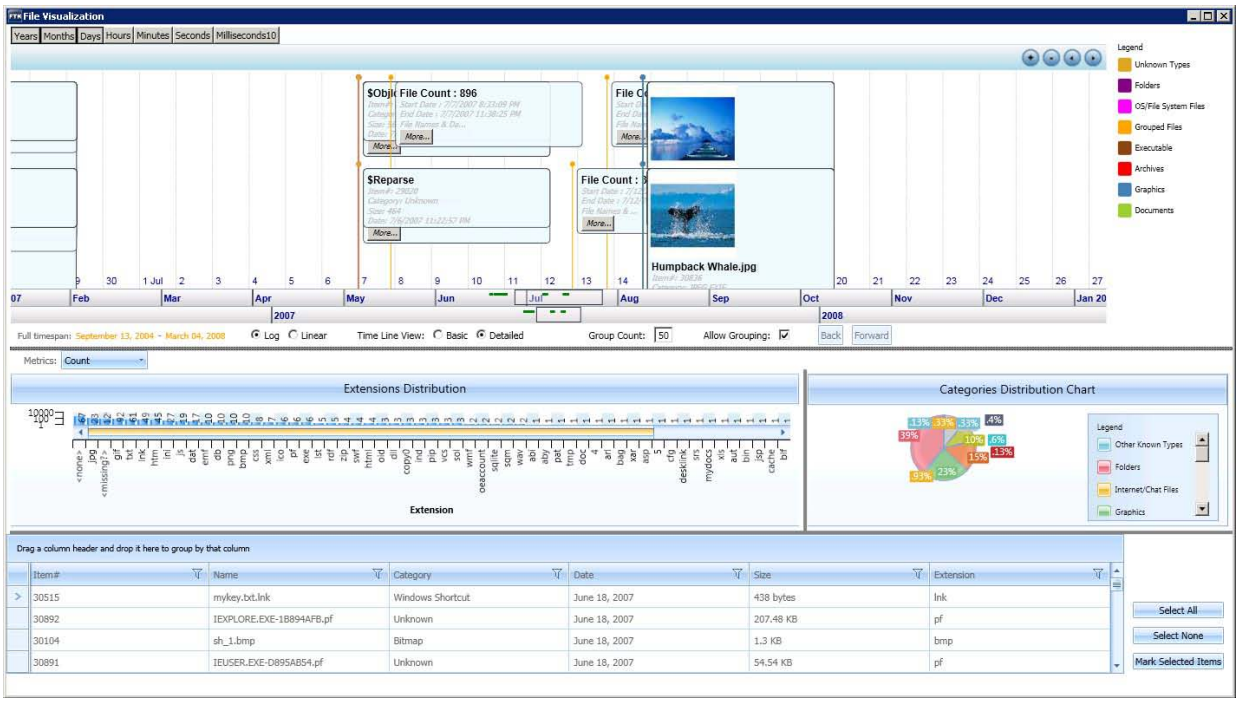


Figura 45. Opcions de visualització de dades a FTK

Funcionalitats

- Visualització de dades en múltiples formats (línies de temps, gràfics, etc.).
- Creació d'imatges forenses.
- Recuperació de contrasenyes.
- Biblioteca amb 45 milions de *hashes*.
- Anàlisi de memòria volàtil.
- Suport per a més de 700 tipus de fitxers.
- Suport per a múltiples sistemes de fitxers.
- Creació d'imatges Advanced Forensic Format (AFF).
- Anàlisi de correus electrònics.
- Identificació de PDFs encriptats.
- Generació d'informes en els formats HTML, PDF, XML i RTF.
- Creació de fitxers csv que es poden importar a Excel o a una base de dades.
- Diferents opcions de configuració en funció del nombre d'ordinadors.
- Mòdul específic (Cerberus) per identificar programari maliciós.

Proves

Com que es tracta d'un programari comercial que no té disponible una versió de prova, no ha estat possible fer proves directes amb ell. No obstant, per tal d'oferir les opcions de la interfície, s'ha consultat un informe del Master of Science in Information Security (MSIS) de la Lewis University (Romeoville, Illinois, EUA), on s'exposaven els diferents menús de visualització, un cop s'ha obert una imatge forense corresponent a un disc dur personal.

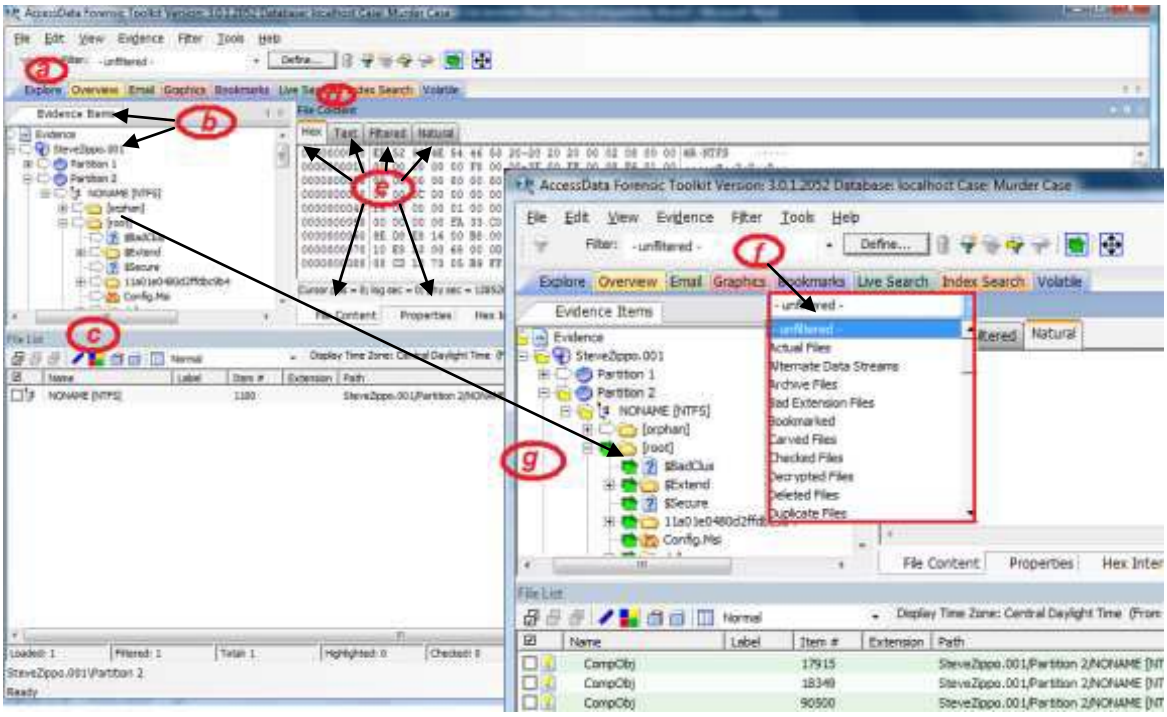


Figura 46. Visualitzador de l'explorador a FTK

a) Exploració. És la primera opció per navegar dins els continguts de la imatge, que té el nom de SteveZippo.001.

b) Evidències. En aquest cas s'aprecia que la imatge correspon a un disc dur que té dos particions i les seves llistes de carpetes.

c) Llista d'arxius. Amb qualsevol arxiu que s'hagi marcat, es podrà veure el seu contingut i propietats a l'apartat d).

d) Contingut de l'arxiu

e) Opcions de contingut de l'arxiu:
   - Hexadecimal
   - Text
   - Filtres

f) Filtres. Aquesta opció permet filtrar els tipus de contingut que es vol consultar.

g) És possible marcar ítems individualment per tal de ser exportats o marcats per ser analitzats posteriorment. Es pot apreciar quins arxius són per la pestanya a l'esquerra de cada fitxer.

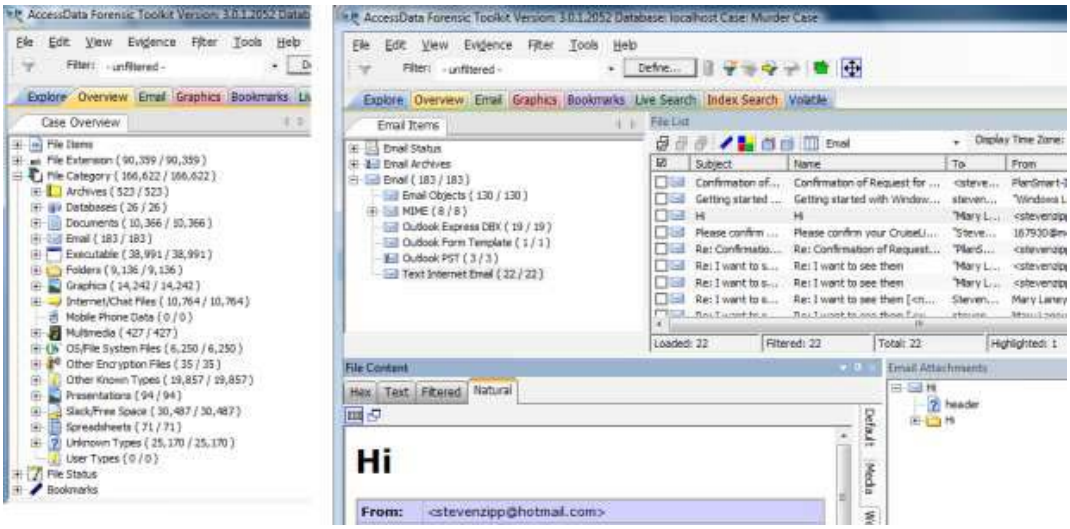


Figura 47. Visualització general i del correu electrònic a FTK

La pestanya de visualització general ('Overview') permet desglossar per tipus de fitxer (en funció de l'extensió), categoria (arxius, bases de dades, carpetes, gràfics, etc.) i estat. Les pestanyes de visualització del correu electrònic i dels gràfics permeten filtrar els resultats en funció del tipus de gestor de correu electrònic en el primer cas i per tipus de gràfics/imatges en el segon.

Hi ha dues opcions de cerca: 'Live Search' i 'Index Search' localitza patrons de cerca com números de telèfon, de targetes de crèdit o de la seguretat social. Per últim, 'Volatile' importa i exporta fitxers de dades volàtils procedents de memòries RAM.

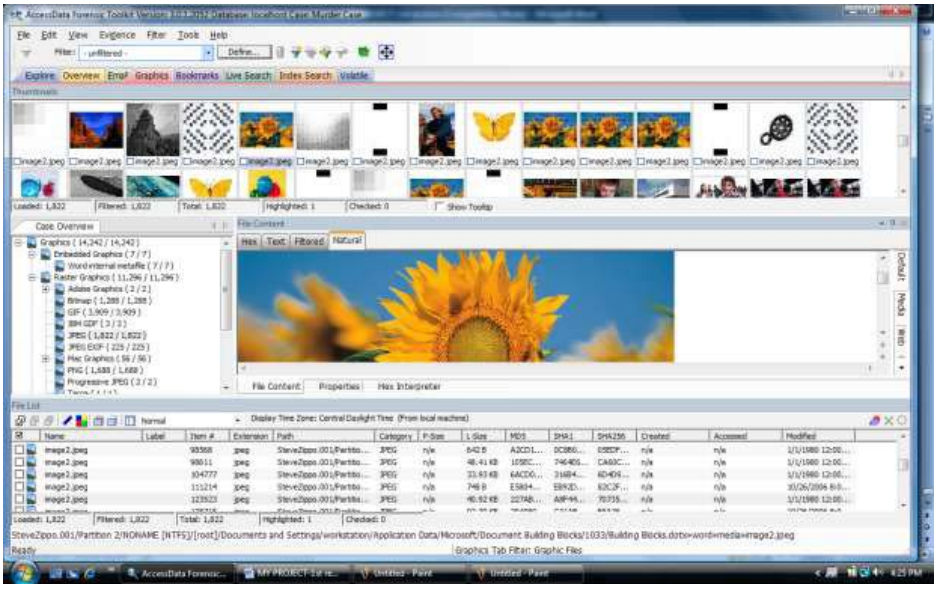


Figura 48. Visualització dels gràfics a FTK

Conclusions

Tot i la impossibilitat de poder establir conclusions definitives, perquè no ha estat possible fer proves directes amb el programa, les diferents opcions que s'han exposat són molt semblants a les d'Autopsy, però més avançades. La localització d'informació privada és en principi més fàcil, hi ha un millor suport per a diferents tipus de fitxers i el processament de dades seria més ràpid, ja que AccessData ha afirmat que el seu programari pot processar més d'un Terabyte d'informació en només dotze hores[28].

5.1.3.3. EnCase Forensic v7 (GuidanceSoftware)

L'empresa Digital Intelligence comercialitza la tecnologia EnCase, la qual inclou EnCase Forensic, EnCase Cybersecurity, EnCase eDiscovery i EnCase Portable. El producte més interessant per a la unitat seria la *suite* dedicada a l'anàlisi forense digital, EnCase Forensic.

Funcionalitats

- Adquisició de dades de telèfons mòbils i de *tablets*.
- Suport per a encriptació de dades en format AES-256.
- Format d'imatge forense E01 i L01.
- Automatització de tasques comuns, com la recuperació de carpetes, anàlisi de *hashes*, cerca de paraules clau, etc.
- Motor de cerca optimitzat per a investigadors forenses.
- Suport per a múltiples tipus de fitxers i de sistemes de fitxers.
- Interfície senzilla d'investigació de correus electrònics.
- Creació d'etiquetes.
- Creació d'informes en els formats RTF, HTML, XML, PDF i text pla.

[28] *Processing over a terabyte of complex data in 12 hours.* Disponible a: <http://marketing.accessdata.com/acton/attachment/4390/f-01b9/1/-/-/-/-/file.pdf> [data de consulta: 30 abr. 2013]

Proves

De la mateixa manera que amb FTK, no ha estat possible accedir a una versió de prova d'aquest programa, però s'ha consultat un article al portal de notícies LTN – Law Technology News[29] on es feia un anàlisis del producte provant el seu rendiment amb un ordinador amb els següents components:

- Lenovo ThinkPad T520 (processador de doble nucli Intel i7-2860QM CPU a 2.5 GHz)
- 8 GB RAM.
- Sistema operatiu Windows 7 (64-bit).
- Disc dur de 165 GB a 7200 rpm.

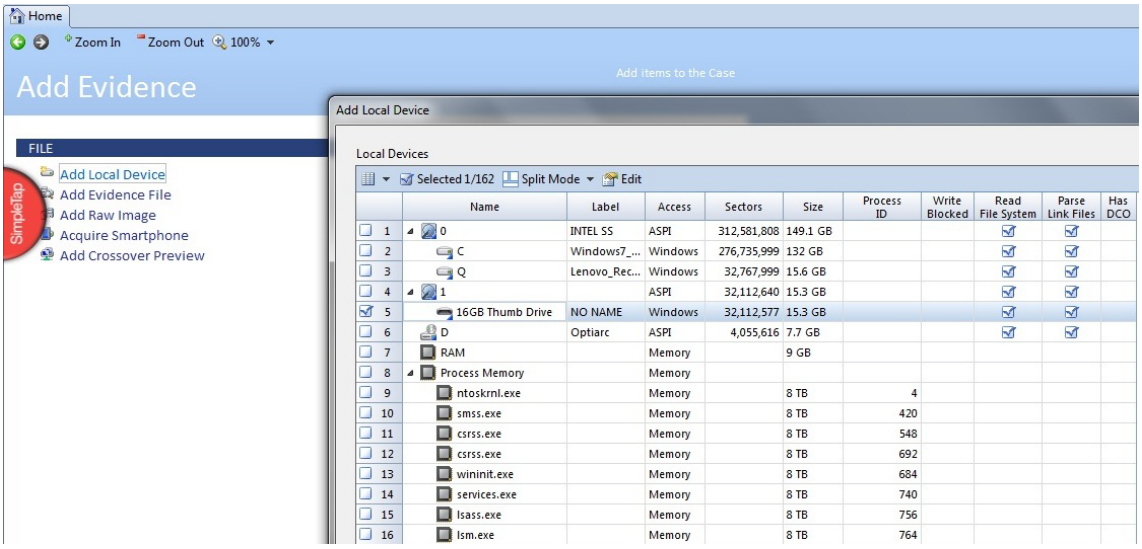


Figura 49. Opcions per obrir dispositius locals a EnCase Forensic

És possible començar un nou cas, crear una imatge forense dels dispositius connectats a l'ordinador o bé seleccionar fitxers lògics. Si es comença un nou cas, una capsa de diàleg demanarà la introducció del nom del cas, la carpeta on es guardarà, i la carpeta on es guardarà la memòria cau. Un cop s'ha seleccionat l'opció desitjada (en aquest cas, es va seleccionar adquirir el contingut d'una memòria USB), una nova capsa de diàleg s'obrirà per tal d'especificar les metadades per a cada adquisició, com el

[29] Disponible a: <http://www.law.com/jsp/lawtechnologynews/index.jsp> [data de consulta: 5 maig 2013]

número de cas i el nom de l'investigador. A continuació, començarà el procés de creació d'imatge amb el format .ex01 i el *hash* corresponent.

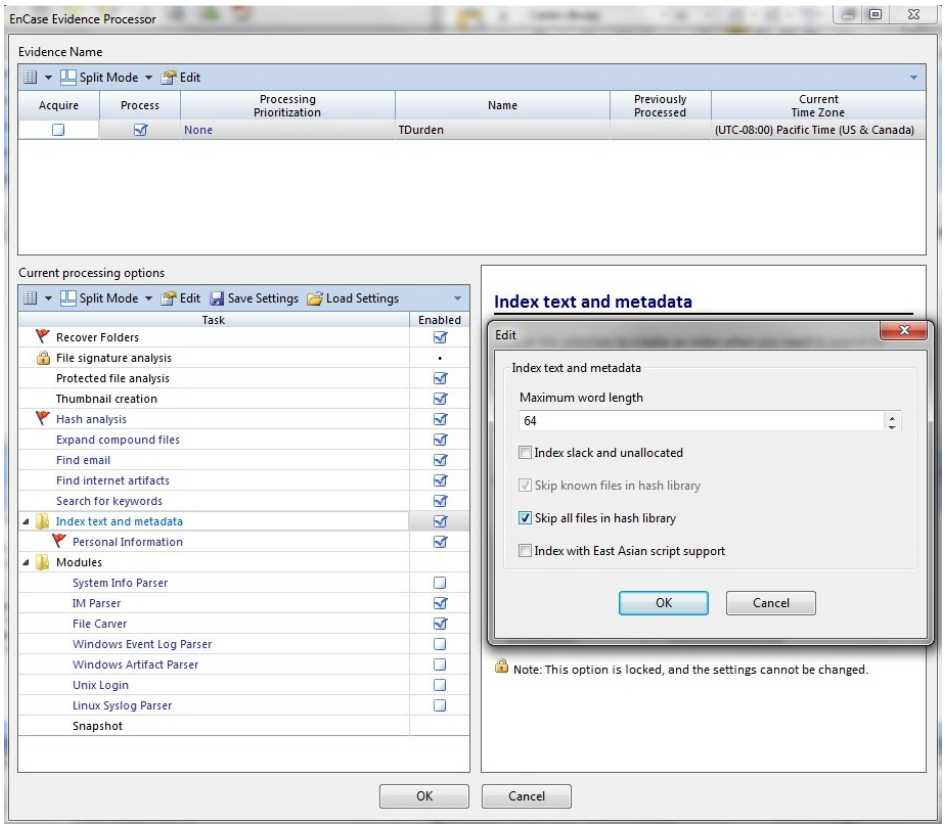


Figura 50. Opcions d'EnCase Forensic per processar fitxers

Una opció que permet recuperar fitxers esborrats o corruptes a sistemes de fitxers FAT i NTFS és la de 'Recover Folders', que també habilita la verificació de *hashes* MD5 i SHA1 i comprovar si són idèntics amb altres fitxers si escau. També és possible crear índexs i metadades i determinar el màxim de caràcters per processar a cada paraula (per defecte, el màxim de caràcters és 64).

Es va fer una prova amb el sistema de cerca mitjançant el cognom del propietari de la memòria USB, 'Tyler', juntament amb el nom propi amb més resultats, 'John'. La cadena de cerca 'John AND Tyler' va resultar en 53 documents personals (documents i correus electrònics especialment), que es podien marcar i afegir etiquetes (ja configurades per defecte) com 'Review', 'Add to Report', 'Follow Up with Submitter',

'Ignore' i 'Important'. És possible crear etiquetes pròpies i ordenar els fitxers en funció dels criteris de l'investigador.

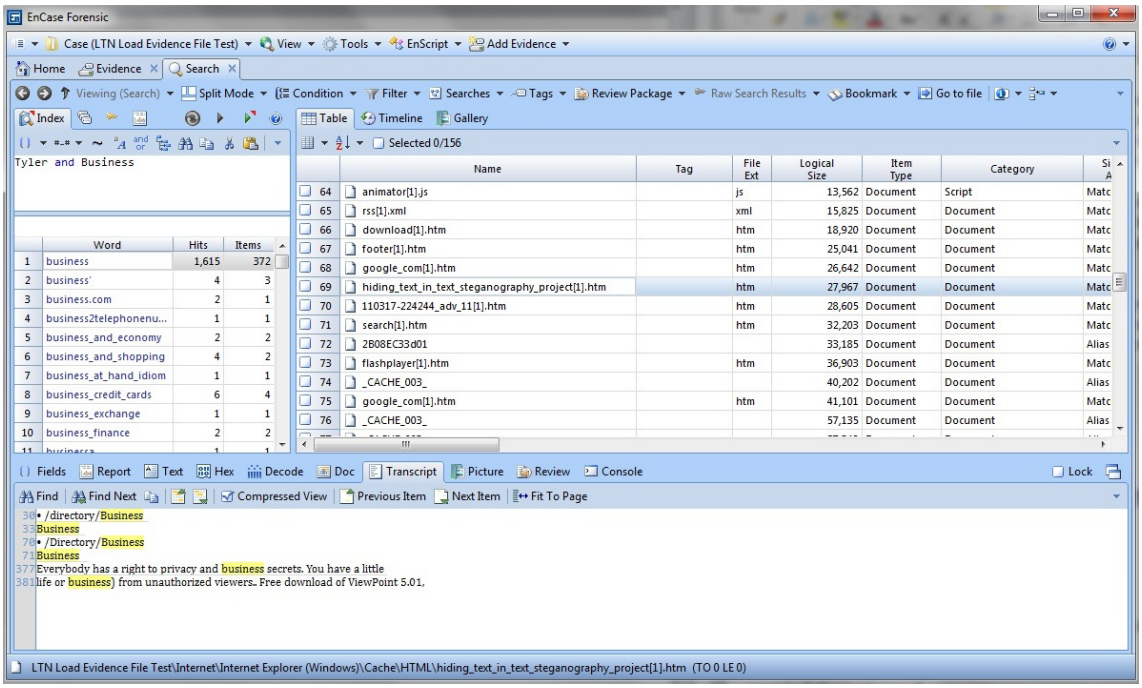


Figura 51. Sistema de cerca a EnCase Forensic

Per últim, la generació d'informes permet detallar els fitxers per tipus, com correus electrònics, imatges o documents en format HTML.

Conclusions

Les funcionalitats són prou semblants a les exposades en FTK, amb alguns canvis com en les funcions de cerca. Però, la interfície sembla una mica més complicada i menys intuïtiva que la de FTK .

Si una institució hagués de triar una de les dues solucions per a la seva unitat d'anàlisi forense digital, el més adient seria adquirir totes dues, ja que alguns experts en aquest camp consideren que ambdues eines són eficients per a tasques concretes i no necessàriament una és millor que l'altra[30].

[30] Obayi, Karl. 2011. Comentari del 13 de juny a "FTK vs. EnCase". *Investigators and consultants* <http://investigatorsandconsultants.com/ftk-vs-encase> [data de consulta: 5 maig 2013]

5.1.4. Identificació i validació de fitxers

5.1.4.1. DROID (The National Archives)

DROID (Digital Record and Object Identification) és una eina desenvolupada per The National Archives, arxiu nacional del Regne Unit, que realitza identificacions automatitzades de formats de fitxers. Es tracta d'un programari lliure de codi obert disponible sota llicència BSD i es pot descarregar al repositori Github. Es requereix la instal·lació de la versió de Java 1.6 Standard Edition (no funciona amb Java 7).

Funcionalitats

- Identificació de formats de fitxers.
- Reconeixement de fitxers segons l'esquema PUID (PRONOM Persistent Unique Identifier)[31]. Actualment es reconeixen un total de 962 formats.
- Creació de perfils d'adquisició.
- Interfície de línies de comandaments.
- Interfície gràfica d'usuari.
- Elaboració d'informes dels següents tipus:
    - Desglossament complet.
    - Tipus i mides de fitxers.
    - Tipus i mides de fitxers per extensió de fitxer.
    - Tipus i mides de fitxer per data.
    - Total de fitxers no llegibles.
    - Total de carpetes no llegibles.

Proves

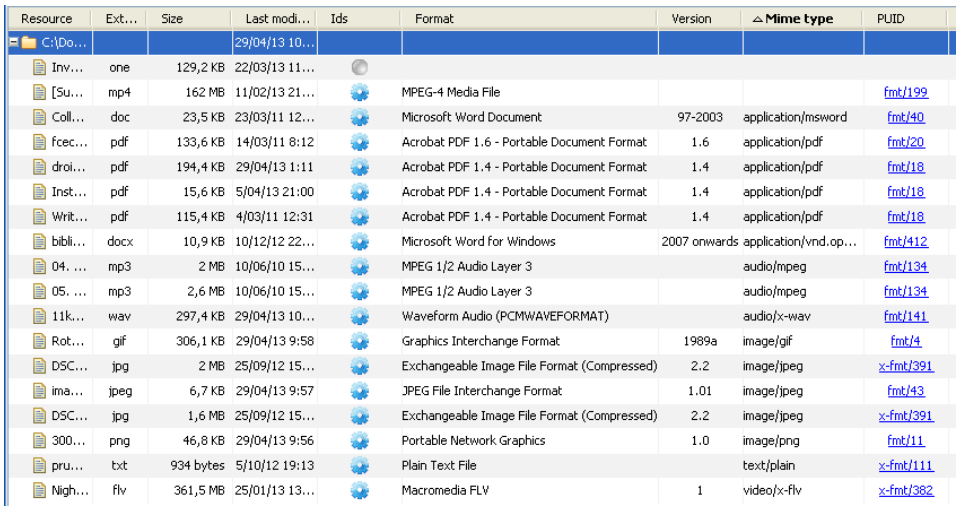
S'ha utilitzat el programa amb una carpeta amb fitxers de text, àudio i imatge, per provar com reconeixia els diferents formats. Va reconèixer els següents fitxers:

- Text. Adobe Acrobat (en diferents versions: 1.4 i 1.6), Microsoft Word versió 97-2003 i 2007 i text pla.

[31] Es pot consultar la totalitat de fitxers reconeguts, amb informació detallada de cadascun d'ells a <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> [data de consulta: 30 abr. 2013]

- Àudio. MPEG (el propi del mp3) i PCM (el corresponent al wav).
- Imatge. Formats png (versió 1.0), jpeg (versions 1.01 i 2.2) i gif (versió 1989a).
- Multimèdia. Formats Macromedia FLV i MPEG-4.

També es van fer proves amb fitxers d'extensions menys comuns, com .one (el propi de Microsoft OneNote), que no van ser reconeguts.



Resource	Ext...	Size	Last modi...	Ids	Format	Version	Mime type	PUID
C:\Do...			29/04/13 10...					
Inv...	one	129,2 KB	22/03/13 11...					
[Su...	mp4	162 MB	11/02/13 21...		MPEG-4 Media File			fmt/199
Coll...	doc	23,5 KB	23/03/11 12...		Microsoft Word Document	97-2003	application/msword	fmt/40
fcec...	pdf	133,6 KB	14/03/11 8:12		Acrobat PDF 1.6 - Portable Document Format	1.6	application/pdf	fmt/20
droi...	pdf	194,4 KB	29/04/13 1:11		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18
Inst...	pdf	15,6 KB	5/04/13 21:00		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18
Writ...	pdf	115,4 KB	4/03/11 12:31		Acrobat PDF 1.4 - Portable Document Format	1.4	application/pdf	fmt/18
bibli...	docx	10,9 KB	10/12/12 22...		Microsoft Word for Windows	2007 onwards	application/vnd.op...	fmt/412
04. ...	mp3	2 MB	10/06/10 15...		MPEG 1/2 Audio Layer 3		audio/mpeg	fmt/134
05. ...	mp3	2,6 MB	10/06/10 15...		MPEG 1/2 Audio Layer 3		audio/mpeg	fmt/134
11k...	wav	297,4 KB	29/04/13 10...		Waveform Audio (PCMWAVEFORMAT)		audio/x-wav	fmt/141
Rot...	gif	306,1 KB	29/04/13 9:58		Graphics Interchange Format	1989a	image/gif	fmt/4
DSC...	jpg	2 MB	25/09/12 15...		Exchangeable Image File Format (Compressed)	2.2	image/jpeg	x-fmt/391
ima...	jpeg	6,7 KB	29/04/13 9:57		JPEG File Interchange Format	1.01	image/jpeg	fmt/43
DSC...	jpg	1,6 MB	25/09/12 15...		Exchangeable Image File Format (Compressed)	2.2	image/jpeg	x-fmt/391
300...	png	46,8 KB	29/04/13 9:56		Portable Network Graphics	1.0	image/png	fmt/11
pru...	txt	934 bytes	5/10/12 19:13		Plain Text File		text/plain	x-fmt/111
Nigh...	flv	361,5 MB	25/01/13 13...		Macromedia FLV	1	video/x-flv	x-fmt/382

Figura 52. Reconeixement de formats de fitxer a DROID

Conclusions

El programa dóna un bon rendiment amb els tipus de fitxers amb els que treballaria la unitat i els identifica ràpidament. La funció de creació d'informes és una bona eina que ajuda a documentar tots els tipus de fitxers (per data, tipus, etc.) fàcilment en funció de les necessitats de la institució.

5.1.4.2. JHOVE (JSTOR i Hardvard University Library)

JHOVE és un programari lliure desenvolupat en Java per la biblioteca digital JSTOR i la Hardvard University Library i dissenyat per a validar formats de fitxer (entès com el procés de determinar el grau de compliment de les especificacions de cada format) mitjançant una interfície gràfica. Existeix una nova versió, JHOVE2, però no té aquesta interfície. JHOVE funciona dins plataformes Unix, Windows o OS X que suportin Java 1.4.

Funcionalitats

- Identificació dels formats de fitxers AIFF, ASCII, Bytestream, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAV i XML
- Validació de format i comprovació del nivell de conformitat
- Informació de les característiques tècniques de cada format

Proves

La instal·lació del programari és una tasca prou complexa, però gràcies al tutorial del grup de recerca de la UB "Preserva"[32] s'han pogut fer les proves pertinents. S'han provat tres tipus de fitxers: un d'imatge JPG, un document PDF i un de so WAV.

Amb el fitxer d'imatge, s'obre una finestra (RepInfo) amb la informació que s'ha extret, que es pot desplegar dins un cercle blau.

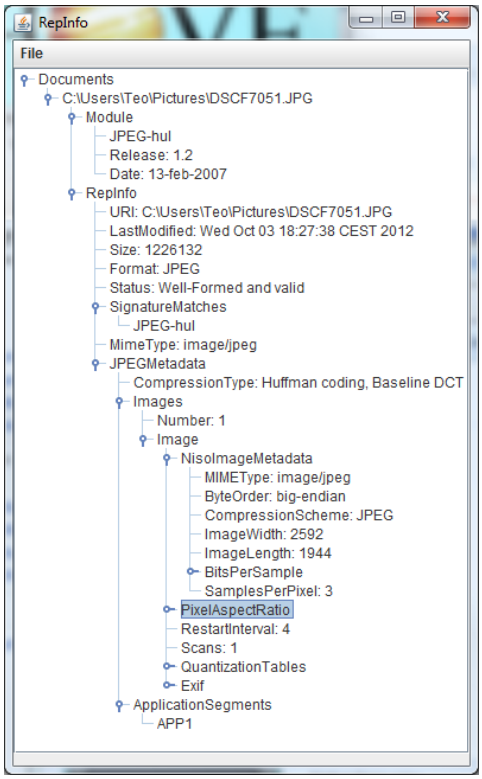


Figura 53. Informació d'un fitxer JPG a JHOVE

[32] Tutorial disponible a <http://bd.ub.edu/preservadigital/sites/default/files/Tutoriales_JHOVE1.6.pdf> [data de consulta: 18 maig 2013]

Aquesta informació es pot guardar en un fitxer de text, XML o Audit. Hi ha diverses metadades que es poden capturar, com la mida d'imatge o el tipus de compressió. En el cas del fitxer PDF, va mostrar la següent informació:

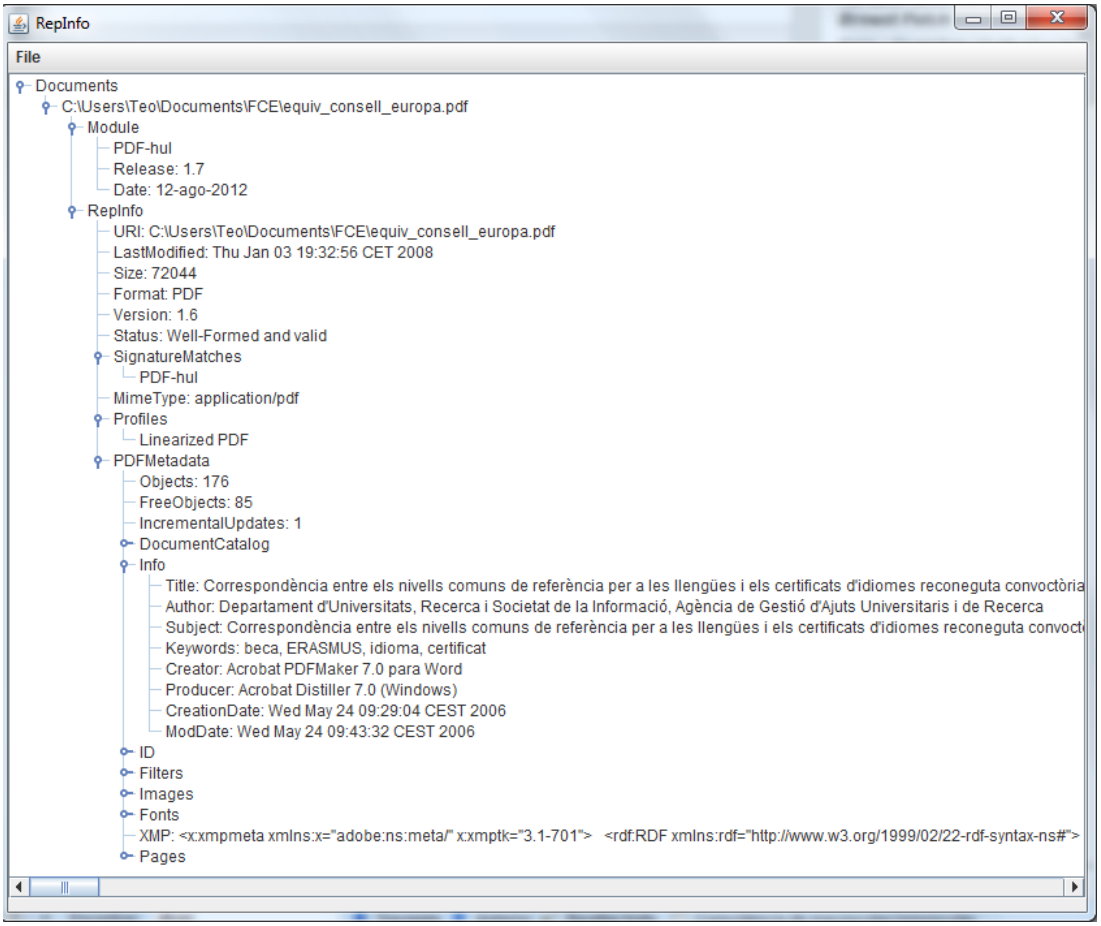


Figura 54. Informació d'un fitxer PDF a JHOVE

En aquest cas, ens dóna informació relativa a la versió amb què es va crear el fitxer originalment, el nombre d'objectes presents al fitxer, metadades de títol, autor, matèria, paraules clau, programa amb què es va crear i la dada de creació. Finalment, el fitxer d'àudio va aportar la següent informació:

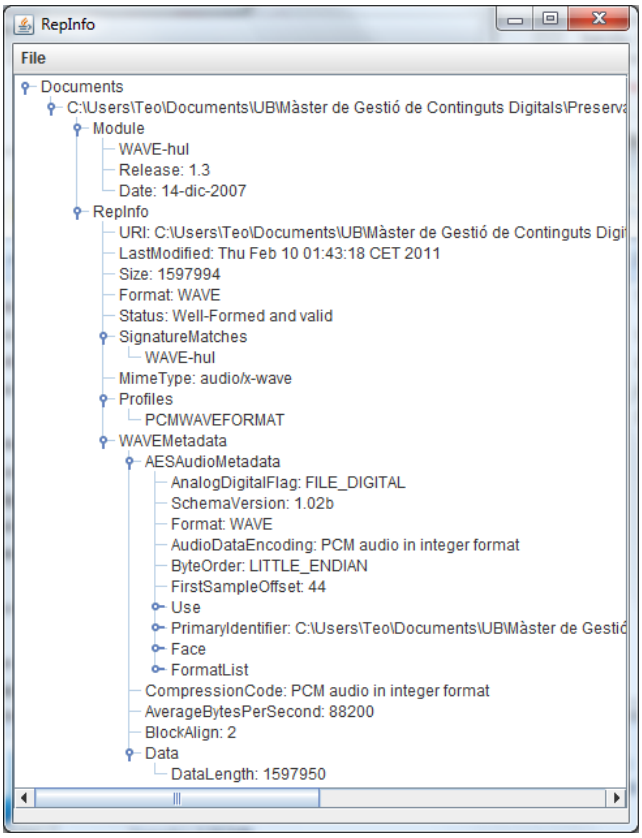


Figura 55. Informació d'un fitxer WAV a JHOVE

Aquí dóna informació vers la codificació d'àudio, durada del fitxer, versió del wav, etc.

Conclusions

Encara que el programa presenta limitacions importants, ja que no reconeix formats de fitxers molt utilitzats com mp3, doc o png, és molt útil per validar diferents tipus de fitxers i així tenir la seguretat que aquests són conformes a les especificacions de cada format.

5.2. MAQUINARI

5.2.1. Ordinadors 'Rosetta'

Hi ha casos en què és aconsellable recuperar contingut *born digital* amb un ordinador compatible amb el que es va crear. És el cas de l'Emory University, que van disposar dels ordinadors Macintosh originals amb què Salman Rushdie va crear els seus

documents (vegeu *3.5. EMORY UNIVERSITY*). L'arxiver digital Doug Reside (University of Maryland) ha utilitzat el terme "màquines Rosetta" per definir els ordinadors que permeten traduir una informació codificada en un sistema obsolet a un altre d'actual, igual que la Pedra de Rosetta va facilitar la traducció dels texts jeroglífics. Un exemple d'això seria la incompatibilitat de Windows XP amb els disquets de 5 ¼ amb format de 360 KB i baixa densitat, que van ser dissenyats per ser utilitzats amb MS-DOS. Un altre exemple seria l'ordinador Macintosh Wallstreet Powerbook G3, que té integrada una unitat SuperDrive, que permet la lectura de disquets de 3 ½ polzades amb el sistema de fitxers d'Apple i format de 800 KB i de 1,44 MB.

Fer un inventari exhaustiu de tots els ordinadors necessaris que caldrien per a tots els casos seria una tasca excessivament llarga, però una alternativa és l'ús d'un ordinador compatible IBM PC que serviria per a una part important dels casos amb què tractaria la unitat.

Figura 56. Ordinador 'Rosetta' amb unitats de disquet instal·lades

Prestacions mínimes:

- Lectura de disquets de 3 ½ i de 5 ¼ polzades.
- Creació d'imatges de disc.
- Accés al sistema de fitxers original.

Especificacions tècniques mínimes:

- Processador Intel Pentium III a 600 MHz.
- Disc dur de 100 GB.
- Memòria RAM de 256 MB.
- Interfície USB.
- Disquetera de 3 ½ polzades.
- Disquetera de 5 ¼ polzades.
- Targeta gràfica compatible SVGA.
- Targeta de so compatible SoundBlaster.
- Sistema operatiu Windows 98 i MS-DOS.

Aquest ordinador pot servir com a solució alternativa per crear imatges de disquets, que després es poden guardar en memòries USB i ser gestionades posteriorment.

5.2.2. Estacions de treball forense

La Stanford University i la Bodleian Library disposen d'unitats FRED de l'empresa Digital Intelligence, que són estacions de treball especialment dissenyades per a l'anàlisi forense digital. Presenten un cost molt elevat, però la seva adquisició pot ser necessària en funció del volum d'adquisicions de la unitat.

Figura 57. Unitat FRED

Les unitats FRED, com que estan optimitzades per a l'adquisició i anàlisi de dades, faciliten la lectura i duplicació de quasi qualsevol dispositiu sense perill d'alterar les dades originals. Permeten la connexió directa de discs durs mitjançant les safates frontals, arrencament dual de dos sistemes operatius (Windows 98/MS-DOS i Windows 7) amb l'opció d'instal·lar Linux, safates optimitzades per a dispositius IDE i SATA que no requereixen apagar el sistema i connexió a xarxa per interfície Ethernet 10/100/1000 Mb que permet el seu ús com estació de treball estàndard.

Presenten les següents especificacions de base (que es poden ampliar):

- Torre amb altura de 23 3/4 polzades, amplada de 8 3/8 polzades i profunditat de 25 1/4 polzades.
- Pes de 36 Kilograms.
- Intel® Core™ i7-3820 CPU (processador multinucli), 3.6 GHz, 10MB Intel® Smart Cache, 5 GT/s DMI.
- Memòria RAM de 16 GB PC3-12800 DDR3 a 1660 MHz.
- Disc dur per a sistema operatiu de 300 GB a 10.000 rpm SATA III.
- Disc dur per a dades de 2 TB a 7.200 rpm SATA III.
- Monitor de 22 polzades widescreen amb altaveus incorporats.
- Sistema operatiu Windows 7 Ultimate 64 bit amb mode Windows XP. També s'inclouen DOS, Windows 98 i Linux 64 bit.
- Write-blocker per a les interfícies:
    - SAS
    - SATA
    - IDE
    - USB 3.0/2.0/1.1
    - FireWire 400/800
- Lector de targetes de memòria amb funció de només lectura i de lectura/escriptura.
- Torre ATX amb 12 safates de 5 ¼ polzades.
- Alimentació de corrent a 1100 watts.

- Placa base i7 amb xip Intel X79 Express.
- 6 ranures PCI-Express 3.0
- Targeta de vídeo Nvidia GT 630 4 GB 128 bits DDR3 PCI-Express amb 1 mini HDMI i 2 ports DVI amb suport a monitors duals.
- Adaptador de xarxa dual 10/100/1000 Mbs.
- Còdec d'àudio de 8 canals d'alta definició.
- 2 ports Intel SATA a 6.0 GB/s.
- 4 ports Intel SATA a 3.0 GB/s.
- 2 ports Marvell SATA a 6.0 GB/s.
- 2 ports PS/2.
- 6 ports USB 3.0.
- 10 ports USB 2.0.
- 1 port USB 2.0/3.0 amb write-blocking.
- 1 port FireWire IEEE 1394a (400 MB/s).
- 2 ports FireWire IEEE 1394b (800 MB/s).
- 2 safates per a discs durs IDE extraïbles.
- 4 safates per a discs durs compatibles IDE i SATA extraïbles.
- Unitat interna de lectura i escriptura BD-R, DVD-R i CD-R.
- Teclat sense fils.
- Ratolí sense fils.
- Cables i adaptadors necessaris.
- Programari:
    - Antivirus Norton Ghost.
    - Creador d'imatges de disc Nero.
    - Anàlisi forense DriveSpy, Image, PDWipe, PDBlock, PART.

5.2.2. Targetes controladores

5.2.2.1. KryoFlux (The Software Preservation Society)

Targeta controladora d'unitats de disquet dissenyada especialment per a la preservació.

Figura 58. Part anterior de la targeta Kryoflux

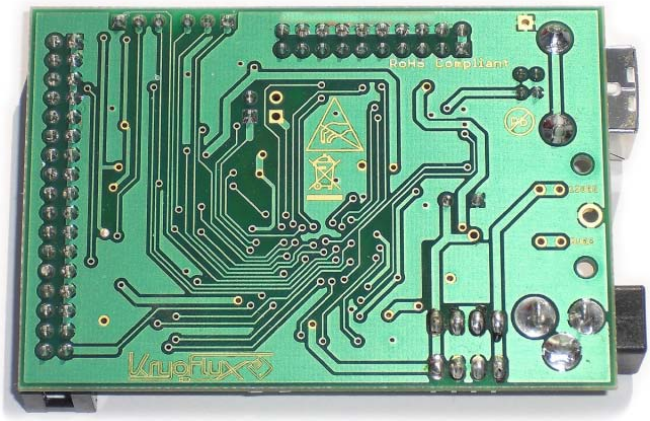


Figura 59. Part posterior de la targeta Kryoflux

Prestacions:

- Permet llegir qualsevol tipus de format de disquet, encara que estigui codificat.
- Permet la lectura de disquets defectuosos.
- Pot crear imatges de disc en format raw i exportar-la a diversos estàndards com Acorn Electron, Apple, Amstrad CPC, Atari ST, Commodore 64, Commodore Amiga, MSX, IBM PC, Sam Coupe, Spectrum, etc.

- Grava imatges de disc i/o fitxers en un altre disquet.
- Amb els disquets que tinguin format doble (com un disquet d'Amiga i Atari ST) es poden crear com a imatge .raw, imatge ADF (de format Amiga) i imatge ST (per a Atari ST).

Especificacions tècniques:

- Compatible amb USB 2.0.
- No requereix de transformador de corrent, ja que la interfície USB de l'ordinador compleix aquesta funció. No obstant, la unitat de disquet sí que requereix d'un alimentador.
- Connexió Shugart de 34 pins (l'estàndard *de facto* per a interfícies de unitat de disquet).
- Compatible amb unitats de 3 ½ i de 5 ¼ polzades. També és possible utilitzar-la amb unitats de 3 i de 8 polzades amb adaptacions especials.
- Suport per a connexió de dues unitats de disquet, però no es pot accedir a les dues simultàniament.

Notes:

- S'aporta programari compatible amb plataformes Windows (XP, Vista, Windows 7), MAC OS X i Linux.
- Format obert (DRAFT) per a l'emmagatzematge d'informació raw.
- Programari gratuït per a ús privat i no comercial.
- Es pot adquirir mitjançant la botiga on-line de la Software Preservation Society[33].
- La seva instal·lació presenta algunes complicacions, així que es recomana que només la faci un tècnic.
- No permet utilitzar les unitats de disquet com a unitats convencionals amb la seva lletra corresponent (A: o B:), així que s'ha d'utilitzar per força el programari administrat per crear les imatges.

[33] Disponible a: <http://webstore.kryoflux.com/catalog/> [data de consulta: 30 oct. 2012]

5.2.2.2. FC5025 (Device Side Data)

Targeta controladora per a unitats de disquet de 5 ¼ polzades.

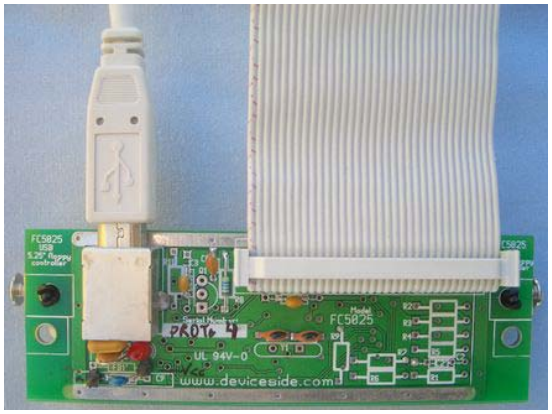


Figura 60. Part anterior de la targeta FC5025

Figura 61. Part posterior de la targeta FC5025

Prestacions:

- Permet llegir diversos formats de disquet com Apple, Atari, Commodore, etc.
- Pot crear imatges de disc en format Apple, Commodore, MS-DOS i altres.

Especificacions tècniques:

- Compatible amb USB 2.0 i USB 1.1.
- No requereix de transformador de corrent, ja que la interfície USB de l'ordinador compleix aquesta funció. La unitat de disquet sí que necessita del seu propi alimentador de corrent.
- Connexió Shugart de 34 pins (l'estàndard *de facto* per a interfícies de unitat de disquet).
- Suport per a una unitat de disquet.

- Permet utilitzar una unitat de disquet de 5 ¼ com unitat extraïble USB.

Presenta les següents limitacions:

- No té suport per a escriptura; per tant, no permet enregistrar imatges de disc ni fitxers als disquets.
- Només és compatible amb disquets de 5 ¼ polzades.
- No pot llegir disquets danyats o amb protecció anticòpia.
- Si el disquet s'ha gravat pel revers a més de l'anvers (la qual cosa és possible amb disquets de 5 ¼), el FC5025 no podrà llegir aquestes dades.

Notes:

- S'aporta programari compatible amb Linux, Mac OS X i Windows que permet la creació de imatges de disc compatibles amb diversos emuladors.
- Els següents tipus de format de disc estan suportats:
    - Apple DOS 3.2 (13 sectors)
    - Apple DOS 3.3 (16 sectors)
    - Apple ProDOS
    - Atari 810
    - Calcomp Vistagraphics 4500
    - Commodore 1541
    - Kaypro 2 CP/M 2.2
    - Kaypro 4 CP/M 2.2
    - MS-DOS
    - North Star MDS-A-D
    - PMC MicroMate
    - Tandy Color Computer Disk BASIC
    - TI-99/4A
- La navegació als fitxers de sistema està suportada pels següents sistemes:
    - ProDos
    - MS-DOS
    - Kaypro

- PMC MicroMate
- Disk BASIC

- L'adquisició de la targeta de moment està limitada als EUA, ja que Device Side Data encara no ha realitzat els tests que la Unió Europea requereix per a la comercialització de dispositius electrònics al seu territori[34].
- L'empresa comercialitza també carcasses per a unitats de disquet, juntament amb alimentadors de corrent, la qual cosa permet connectar la unitat a l'ordinador com qualsevol altre dispositiu USB.

5.2.3. Unitats de disquet

5.2.3.1. Unitat de disquet de 5 ¼ polzades

Són prou difícils d'adquirir actualment, però encara és possible trobar preus raonables per una unitat en bon estat. Per a l'elaboració del present treball s'adquirí una unitat TEAC FD-55GFR, que permet la lectura i escriptura de disquets de les següents capacitats:

- 360 KB i 40 pistes.
- 1.2 MB i 80 pistes.

Figura 62. Part frontal de la unitat de disquet de 5 ¼ TEAC FD-55GFR

[34] Correu electrònic de Device Side data enviat a l'autor en data de 5 juny 2012.

Figura 63. Part posterior de la unitat de disquet de 5 ¼ TEAC FD-55GFR

Val a dir que existeixen unitats que només tenen capacitat per a disquets de 360 KB i per la qual cosa no complirien els requisits mínims de la unitat forense. També s'ha de tenir en compte que només és possible adquirir unitats internes; no hi ha cap model que estigui dissenyat per ser utilitzat externament amb connexió USB.

5.2.3.2. Unitat de disquet de 3 ½ polzades

No presenten cap problema per a la seva adquisició; a diferència de les de 5 ¼, és possible adquirir una unitat externa amb connexió USB. Això té l'avantatge d'una connexió directe a l'ordinador sense haver de fer servir ordinadors amb unitats ja instal·lades o targetes controladores, però també presenta un inconvenient important: només reconeix disquets formatejats a 1.44 MB. No pot treballar amb disquets formatejats amb densitats més baixes, com els de 720 KB o els de 360 KB (molt estesos en sistemes antics, com el MSX).

Figura 64. Unitat de disquet de 3 ½ USB (Freecom)

Per tant, en els casos que correspongui s'ha de fer servir una unitat interna, bé instal·lada a un ordinador clònic (com ja s'ha indicat a 5.2.1.) o bé mitjançant una targeta controladora i connectar-la per USB.

Figura 65. Unitat de disquet de 3 ½ interna (Sony)

5.2.3.3. Unitat de disquet Zip

Els disquets Zip en un principi tenien una capacitat de 100 MB, però posteriorment aquesta capacitat va augmentar a 250 MB i més tard a 750 MB. Existeixen diverses versions d'aquestes unitats: tant internes com externes i compatibles amb connexions IDE, port paral·lel, FireWire o USB. Per a la unitat forense el més adient seria la unitat externa 750 amb connexió USB 2.0, que seria la que permetria la lectura de qualsevol disquet Zip. Es pot adquirir prou fàcilment en botigues d'informàtica especialitzades.

Figura 66. Unitat Zip externa i disquet Zip de 750 MB (Iomega)

5.2.4. Docking stations

Els discs durs i lectors de CD-ROM/DVD-ROM externs no representen cap problema de connexió, ja que han estat dissenyats amb aquesta funcionalitat: poder ser connectats d'un ordinador a un altre sense fer cap tipus d'instal·lació. Els discs durs interns, no obstant, necessiten ser connectats a una placa base. No obstant, hi ha la possibilitat de connectar aquests discs durs interns com si fossin externs mitjançant els *docking stations*, dispositius amb connexió USB i que permeten connectar discs durs i lectors de DVD-ROMs (entre d'altres aparells).

Figura 67. *Docking station* amb connexió SATA superior i connexió IDE frontal (StarTech)

És especialment important que continguin la connexió IDE, ja que aquesta interfície s'ha utilitzat àmpliament les darreres dècades en els discs durs i encara s'utilitza actualment.

5.2.5. Write-blockers

Aquests dispositius són totalment imprescindibles dins l'anàlisi forense digital, ja que permet assegurar que les dades d'un disc dur o dispositiu USB no quedin alterades o modificades un cop es connecten a la unitat; qualsevol ordre del sistema d'escriptura queda bloquejada i això permet assegurar la integritat de les dades, la qual cosa és vital dins l'anàlisi forense: l'alteració d'una prova la invalidaria per presentar-la a un

judici i això s'aplica també a les donacions de particulars. Si el bibliotecari o arxiver manipula inconscientment les dades contingudes al dispositiu, es perdrà informació vital.

Figura 68. Disc dur (Maxtor) connectat a un *write-blocker* (Tableau)

Existeixen diversos models de *write-blocker*, però el més adient per a la unitat seria un de compatible amb les interfícies IDE, USB i SATA, i opcionalment amb les interfícies SCSI i FireWire. També es pot contemplar la instal·lació de *write-blockers* interns.

5.3. PERSONAL

L'equip assignat per fer les tasques d'anàlisi forense hauria de contemplar les següents característiques:

- **Un cap de projecte** amb coneixements avançats tant d'arxivística i documentació com de preservació digital aplicada a les institucions culturals. Titulació de llicenciat en documentació o bé de graduat en informació i documentació i també de màster en gestió de continguts digitals.
- **Un enginyer informàtic** expert en formats de dades, text, imatge, àudio i vídeo, sistemes de fitxers i estàndards informàtics. Titulació d'enginyeria informàtica.

5.3.1. Competències

Segons un estudi del projecte DigCurv[35] (Digital Curator Vocational Education Europe), un projecte del programa Leonardo da Vinci de la Comissió Europea per establir un marc curricular per a la formació en preservació digital, les competències transversals que ha de tenir el personal serien les següents:

- Capacitat de treballar en equip i col·laborar.
- Capacitats de comunicació.
- Interès per la tecnologia.
- Gestionar projectes.
- Saber formar i ensenyar.
- Gestionar pressupostos.
- Capacitats de lideratge.
- Organitzar conferències, tallers i altres actes.

En quant a funcions específiques de preservació digital, l'estudi va concloure que les competències més importants són:

- Planificació per a preservació.
- Garantir l'accés.
- Gestionar les dades.
- Avaluar i seleccionar dades per a preservació a llarg termini.
- Emmagatzematge de dades.
- Ingesta de dades.
- Recerca, desenvolupament i implementació d'entorns de preservació digital.
- Administració de l'arxiu digital.

També s'han de tenir en compte les competències específiques necessàries per treballar amb eines d'anàlisi forense digital. Dins l'entorn de biblioteques i arxius, les més importants serien:

[35] Pàgina web disponible a: <http://www.digcur-education.org> [data de consulta: 2 maig 2013]

- Gestió de *hashes* i *checksums*.
- Creació d'informes acurats.
- Creació d'imatges de disc.
- Coneixement de formats de capçalera de fitxers.
- Coneixement de sistemes de fitxers.
- Anàlisi de múltiples suports informàtics (disquets, CD-ROMs, DVD-ROMs, etc.).
- Tenir cura amb la privacitat de les dades personals.
- Gestió de metadades per a documents d'imatge, text, vídeo i àudio.
- Recuperació de dades.

5.3.2. Formació

El personal requeriria d'una formació complementària, sempre en funció dels requeriments específics de la unitat i del seu pressupost. Existeixen diverses opcions dins l'oferta acadèmica al Regne Unit i als EUA.

5.3.2.1. Digital Stewardship Certificate[36]

Curs en línia de la Graduate School of Library and Information Science del Simmons College (Boston, EUA) amb l'objectiu de formar administradors digitals a biblioteques i arxius. Impartit per Linda Braun, professora de pràctiques; Ross Harvey, professor; Aaron Rubinstein, arxiver digital i Nanette Veilleux, professora associada.

L'alumne ha de superar 15 crèdits dividits en cinc assignatures (tres obligatòries i dues optatives). El pla d'estudis és el següent:

Assignatures obligatòries:

- Administració digital
- Arxiu i preservació de mitjans digitals

[36] Pàgina web disponible a: <http://www.simmons.edu/gslis/academics/programs/post-masters/dsc/index.php> [data de consulta: 2 maig 2013]

- Estudi independent o bé pràctiques a un centre

Assignatures optatives:

- Metadades
- Administració de bases de dades
- Desenvolupament del web i arquitectura de la informació
- XML
- Gestió d'actius digitals a biblioteques, arxius i museus

L'admissió d'alumnes es fa un cop l'any, al setembre, i el curs es pot completar en tres semestres o en cinc semestres.

Els requisits d'admissió són els següents:

- Titulació de màster o equivalent dins el camp de la biblioteconomia i documentació, gestió d'arxius o altre camp relacionat amb l'administració digital.
- Enviament del currículum vital.
- Enviament d'una carta de motivació.
- Enviament del certificat acadèmic del màster.

5.3.2.2. Digital Forensics for Curation of Digital Collections

Curs presencial de la School of Information and Library Science[37] de la University of North Carolina (UNC). Impartit per Christopher A. Lee[38], professor associat de la universitat i expert en la conservació de col·leccions digitals.

El seu temari tracta específicament l'anàlisi digital forense, ja que els alumnes reben formació sobre:

[37] Pàgina web disponible a: <http://sils.unc.edu/> [data de consulta: 2 maig 2013]

[38] *Christopher (Cal) Lee*. Disponible a: <http://www.ils.unc.edu/callee/> [data de consulta: 4 maig 2013]

- Maquinari, programari i mètodes per extraure dades digitals de múltiples suports.
- Extracció de diverses formes de metadades per ser incorporades a *workflows* de conservació digital.
- Dispositius i interfícies més comuns.
- Ús de dispositius *write-blockers* per adquirir dades.
- Estructures de sistemes de fitxers.
- Importància dels valors *hash*.
- Visualització de valors hexadecimals.
- Programari d'adquisició de dades.

Els alumnes tenen a la seva disposició maquinari i programari (tant comercial com de codi obert) digital forense d'última generació per així explorar les possibilitats d'ús en múltiples escenaris per a professionals de la informació.

5.3.2.3. Cursos de formació de programari comercial

Les empreses AccessData i Digital Intelligence ofereixen cursos, presencials i en línia, per a l'ús dels seus respectius programaris. És possible demanar un curs 'a la carta' específicament per a una entitat.

L'oferta de cursos d'anàlisi forense inclouen formació en:

- Fonaments de l'anàlisi forense digital.
    - Sistemes de fitxers.
    - Partició de discs durs.
- Programari FTK (Forensic Toolkit); nivells bàsic i avançat.
- Anàlisi forense digital en els sistemes operatius:
    - Windows XP.
    - Windows Vista.
    - Windows 7.
    - Macintosh.

- Anàlisi de dades.
- Cerca de dades en codi hexadecimal.

Les opcions més adients per a usuaris principiants en FTK i EnCase serien, en el primer cas, l''AccessData 5 Day BootCamp', un curs de cinc dies en línia que forma als alumnes en els rudiments bàsics del programa. En el segon cas, Digital Intelligence ofereix els cursos 'Encase Computer Forensics' I i II; el primer per a usuaris inicials i el segon, per a usuaris intermedis, un cop s'ha dominat el més bàsic.

5.4. ENTORN DE TREBALL

La unitat hauria de tenir un entorn concret dins la institució per tal d'elaborar amb eficàcia les seves tasques. Un entorn d'oficina estàndard seria suficient, amb un espai mínim que permeti el treball amb tres ordinadors. Però, s'han de considerar certs elements:

- El maquinari és car i costós, així que s'ha de restringir el seu accés.
- Es rebrien dades personals en múltiples suports i per tant s'han de tractar amb molt de cura.
- S'han d'evitar possibles robatoris del material.
- Es necessitarien múltiples espais per a endolls; s'han de connectar monitors, ordinadors, targetes controladores, dispositius externs USB, etc.
- També es necessitarien múltiples connexions Ethernet, en funció del nombre d'estacions de treball.
- S'ha de reservar espai per a prestatgeries, i aquestes han d'estar especialment condicionades per a material informàtic: disquets i discs durs especialment.
- Igualment, s'ha de reservar lloc per als diferents cables utilitzats. Per exemple, la targeta Kryoflux necessita de cables IDE, cable de connexió a corrent i cable USB.

Per tant, l'entorn de treball ha de ser tancat amb clau quan el personal no estigui operant a la unitat. També seria recomanable l'ús de videocàmeres i dispositius antifurts.

5.5. PROCEDIMENTS DE TREBALL

Tal com s'ha exposat a l'apartat 3, els procediments de treball presenten semblances, però també diferències importants:

Taula 2. Maquinari i programari forense a les unitats ja existents

	Stanford University Libraries	**Bodleian Library**	**National Library of Australia**	**Yale University**	**Emory University**
Maquinari	Dues unitats FRED amb *write-blocker*	Una unitat FRED amb *write-blocker*	Una unitat portàtil on es poden adaptar unitats de diverses tipologies	Targetes Kryoflux i FC5025 amb *write-blocker*	No es va utilitzar maquinari d'anàlisi forense
Programari	Llicències comercials	Programari forense lliure	Programari forense lliure	Programari forense lliure	No es va utilitzar programari d'anàlisi forense

S'ha donat prioritat a l'ús de programari forense lliure i a maquinari que sigui senzill d'adquirir i utilitzar, atès que la major part dels casos a Espanya de preservació de material *born digital* vindran de disquets i discs durs i no cal fer una inversió excessiva. Aquests procediments són merament orientatius, però poden servir perfectament a la Biblioteca de Catalunya, que fa servir un repositori intern de preservació digital, el COFRE (COnservem per al Futur Recursos Electrònics).

5.5.1. Disquets

- **Descripció física**. Es fa una fotografia del disquet i es generen metadades sobre la seva descripció física i el seu sistema de fitxers, en funció del que digui l'etiqueta o bé en funció de la informació proporcionada pel donant. És important recordar que només s'accepten sistemes de fitxers esmenats a l'apartat 2.2.
- **Bloqueig d'escriptura de disc**. Per tal d'evitar qualsevol alteració en el contingut original, habilitarem la pestanya de protecció contra escriptura en el cas dels disquets de 3 ½ i de 5 ¼ polzades. En el cas de disquets Zip, connectarem el dispositiu *write-blocker* a la unitat externa.
- **Creació d'imatges de disc**.
    - Si es disposa d'un ordinador amb unitat de disquet interna de 3 ½, de 5 ¼ i Zip, es crea la imatge mitjançant FTK Imager.
    - Si no es disposa d'aquest ordinador, es connecta la unitat amb la targeta controladora Kryoflux i es crea la imatge amb el programari subministrat amb la targeta, dtc.exe. S'ha de fer tenint en compte el sistema de fitxers i la capacitat del disquet. En el cas dels discs Zip, es pot utilitzar una unitat externa amb connexió per USB.
- **Verificació de la imatge de disc**.
    - Si s'ha creat la imatge amb dtc, la verifiquem amb el programari gratuït MD5summer, amb el qual generem dos algoritmes *hash*, un del tipus MD5 i un altre del tipus SHA-1. Posteriorment, obrim la imatge de disc amb FTK Imager i verifiquem que els *hash* són correctes.
    - Si s'ha creat la imatge amb FTK Imager, aquest ens crearà un informe amb els *hash* MD5 i SHA-1 corresponents.
- **Extracció dels fitxers**.
    - FTK Imager extreu els fitxers, el llistat dels *hash* de cadascun d'ells (amb extensió csv) i el llistat de directoris (amb extensió csv).
- **Escanejat de virus**.
    - L'antivirus avast! inspeccionarà que tots els fitxers estiguin lliures de programari maliciós.

- **Identificació i validació dels formats de fitxer**.
    - DROID identifica els formats de fitxer, en funció de la seva extensió.
    - JHOVE valida els fitxers, ja que a més aporta informació com el *hash* i el seu format.
- **Cerca de dades privades i/o sensibles**.
    - S'ha d'inspeccionar si els fitxers contenen dades privades i/o sensibles del donant, com informació personal, bancària, filiació política o qualsevol altre que el donant hagi especificat. Aquest contingut ha de quedar bloquejat per a l'accés públic per tal de complir amb la Llei Orgànica de protecció de dades de caràcter personal[39]. Un programa adient per trobar dades sensibles seria Autopsy.
- **Ingesta al repositori**.
    - Un cop s'han validat els fitxers i s'han seleccionat aquells que s'han de preservar, es copien juntament amb la imatge de disc i els fitxers de metadades al repositori institucional.
- **Comprovació de la integritat del fitxer**.
    - Mitjançant MD5summer, comprovem que el *hash* dels fitxers que s'han copiat al repositori corresponen al fitxers originals.
- **Accés al contingut**.
    - Si el contingut és totalment accessible, es podrà accedir-hi en línia lliurement.
    - Si no és totalment accessible, s'hi podrà accedir mitjançant els terminals de la institució.

[39] *Llei orgànica 15/1999, de 13 de desembre, de protecció de dades de caràcter personal* <http://www20.gencat.cat/docs/Adjucat/Documents/ARXIUS/lo15_1999lopdcp.pdf> [data de consulta: 18 maig 2013]

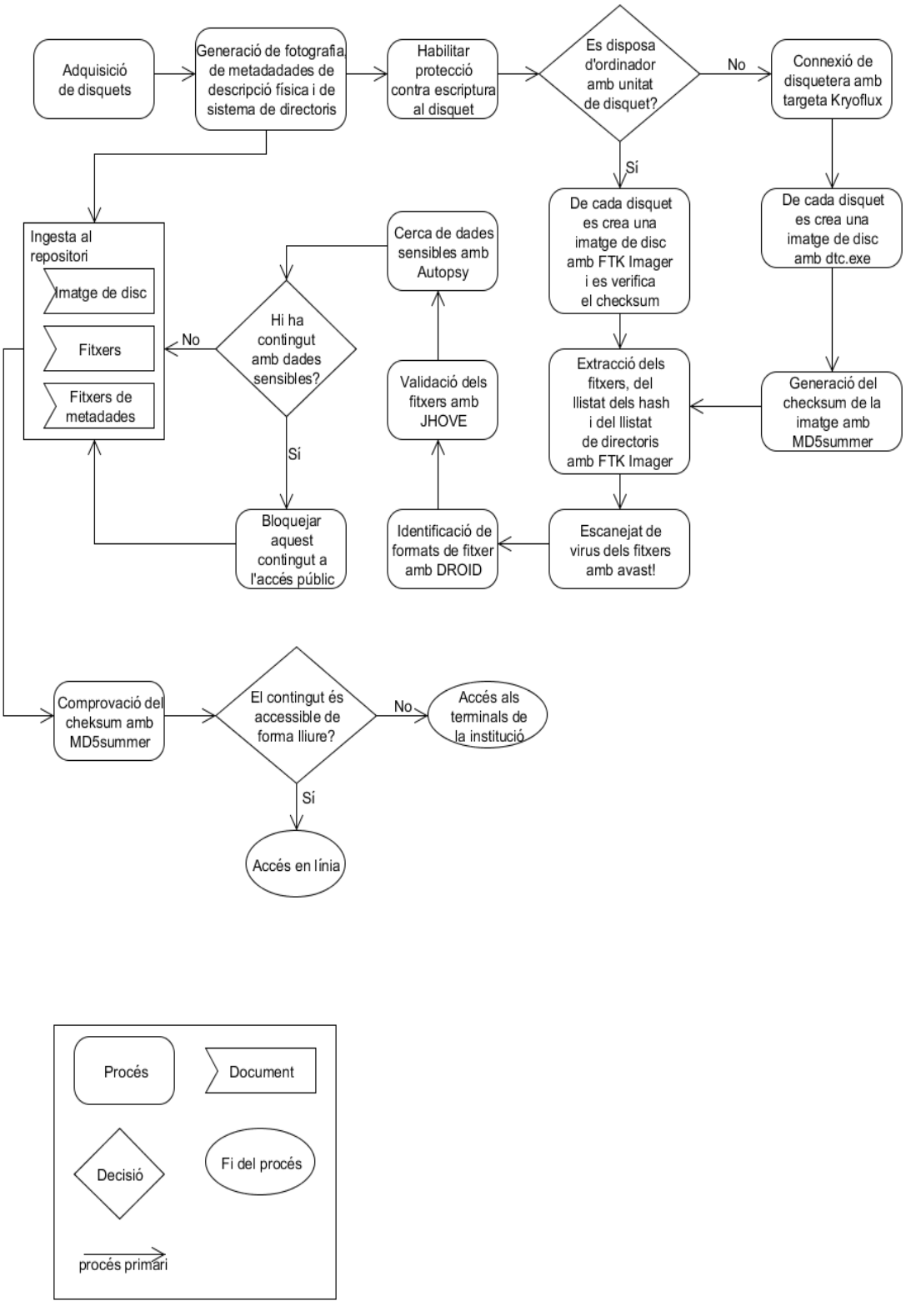


Figura 69. *Workflow* del procediment de treball amb disquets

5.5.2. Discs durs

El procediment és semblant al dels disquets, però amb una diferència important: la quantitat de fitxers en un disc dur és molt elevada en comparació amb un disquet, per tant no seria raonable fer l'extracció de tots els fitxers. L'alternativa seria la creació d'una imatge de disc amb els continguts íntegres i recuperables posteriorment fent una recerca mitjançant eines forenses. El procés en aquest cas seria el següent:

- **Descripció física**. Es fa una fotografia del disc dur i es generen metadades sobre la seva descripció física i seu sistema de fitxers. Aquestes metadades es poden guardar en un fitxer xml amb el format més adient per a la institució.
- **Bloquejar el disc dur contra escriptura**. Per tal d'evitar qualsevol alteració de les dades del disc dur i així assegurar la seva integritat, s'ha de connectar un *write-blocker* (vegeu *5.2.5. Write-blockers*) al disc dur i posar-lo en marxa.
- **Connexió del disc dur a l'estació de treball**.
    - Si el disc dur és del tipus extern i connectable directament per interfície USB, el connectarem d'aquesta manera.
    - Si el disc dur és del tipus intern, utilitzarem un *docking station* (vegeu *5.2.4. Docking stations*) per connectar-lo mitjançant USB.
- **Creació i verificació de la imatge de disc**. S'utilitzarà FTK Imager per crear la imatge, que crearà un informe amb els *hash* MD5 i SHA-1 corresponents. En el cas que el donant hagi especificat un nombre de fitxers concrets per preservar, els seleccionarem i els extraurem.
- **Extracció del llistat de directoris i de fitxers**.
    - FTK Imager extreu el llistat dels *hash* dels fitxers seleccionats (amb extensió csv) i el llistat de directoris (amb extensió csv). En el cas que el donant hagi especificat un nombre de fitxers concrets per preservar, els seleccionarem i els extraurem.
- **Escanejat de virus**. Igual que amb els disquets.
- **Identificació i validació dels formats de fitxer**. Igual que amb els disquets.
- **Cerca de dades sensibles**. Igual que amb els disquets.

- **Ingesta al repositori**. La imatge de disc, els fitxers que haguem seleccionat i els fitxers de metadades es copien al repositori institucional.
- **Comprovació de la integritat del fitxer**. Igual que amb els disquets.
- **Accés al contingut**. Igual que amb els disquets.

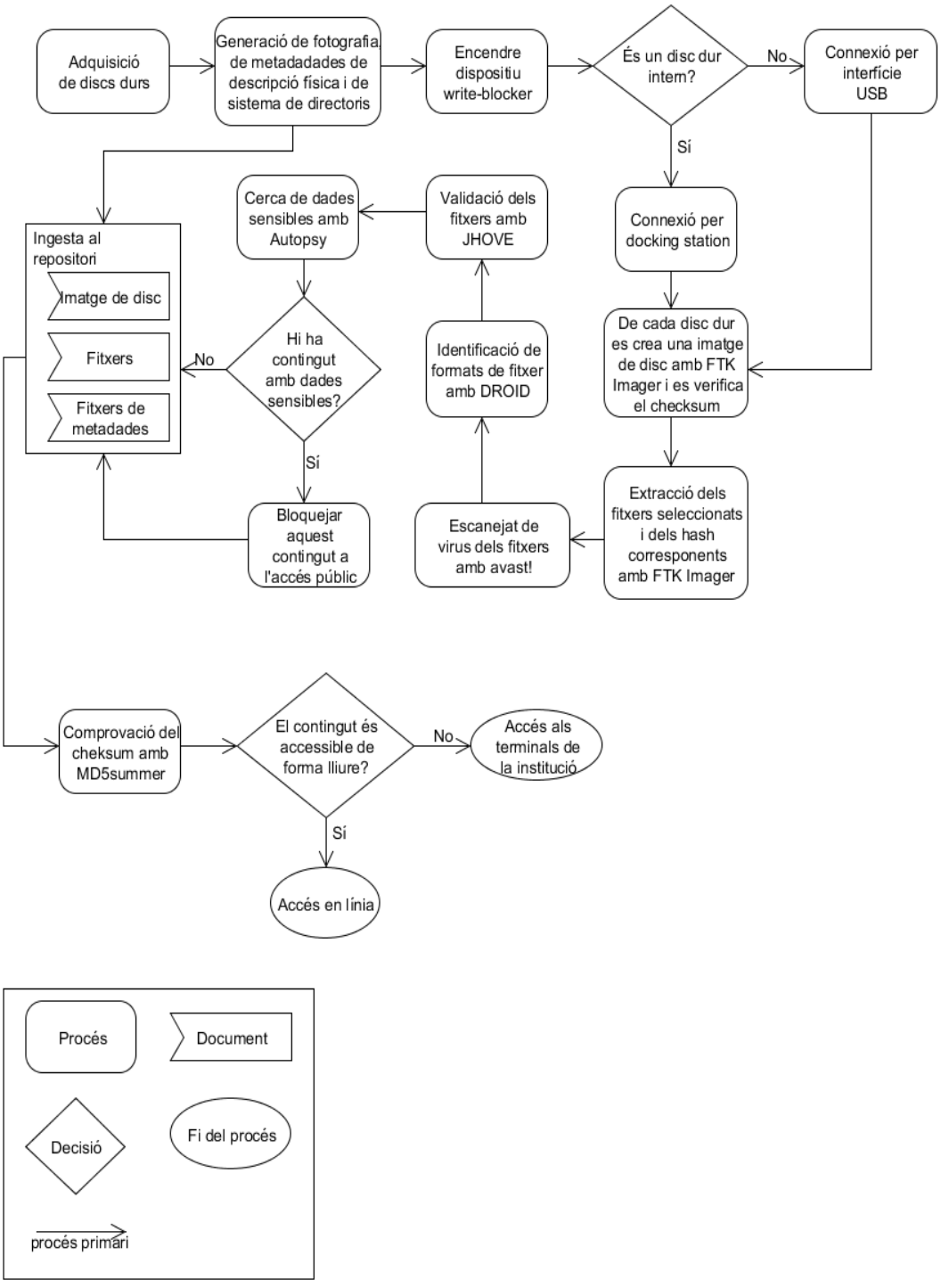


Figura 70. *Workflow* del procediment de treball amb discs durs amb arxius personals

Dins dels procediments de treball amb disc durs s'ha de contemplar l'adquisició de suports que continguin dades de recerca de *data management plans*, els quals són, segons defineix la California Digital Library[40], *document formal que dóna una idea general sobre què es fa amb les dades durant un procés de recerca i després de completar el projecte.* L'objectiu de un *data management plan* és considerar (sempre en un context de recerca) els aspectes de la gestió de dades, la generació de metadades, la preservació de dades i el seu anàlisi abans del començament del projecte i també garantir que les dades són gestionades correctament per a la seva preservació un cop el projecte hagi finalitzat. Ara per ara, a Espanya aquest document no és un requisit indispensable per iniciar un projecte, però a l'àmbit anglosaxó (EUA i el Regne Unit especialment) ja s'exigeix a institucions com el National Science Foundation (NSF)[41], on el *data management plan* ha de tenir un màxim de dues pàgines amb la descripció de la proposta i de quina manera compliria amb la política de la NSF vers la difusió dels resultats de la recerca i com es comparteixen. El document pot incloure:

- Tipus de dades, mostres, col·leccions físiques, programari, materials curriculars i altres materials que s'hagin de produir durant el projecte.
- Estàndards per ser utilitzats en dades i formats de metadades i altres continguts.
- Política d'accés a les dades.
- Política per a nous usos i noves distribucions de les dades.
- Plans per a l'arxiu de dades.

Si aquest pla es comencés a aplicar a universitats espanyoles, les biblioteques universitàries rebrien, si la universitat aplica aquesta política, a més d'una còpia dels projectes de recerca en paper i en altres suports com DVD-ROMs, discs durs amb les

[40] *Data management plans and DMPTool*. Disponible a: <http://www.cdlib.org/services/uc3/dmp/index.html> [data de consulta: 11 abr. 2013]

[41] *Plans for data management and sharing of the products of research*. Disponible a: <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp> [data de consulta: 11 abr. 2013]

dades 'en brut' fruit dels treballs efectuats al projecte per tal de garantir la seva preservació. Només a la Universitat de Barcelona hi ha més de 700 projectes de recerca actius, així que és fàcil preveure que la gestió d'un volum de dades considerable pot representar un greu problema. Per tant, la unitat forense digital haurà d'aplicar un *workflow* semblant al descrit anteriorment, però amb algunes diferències.

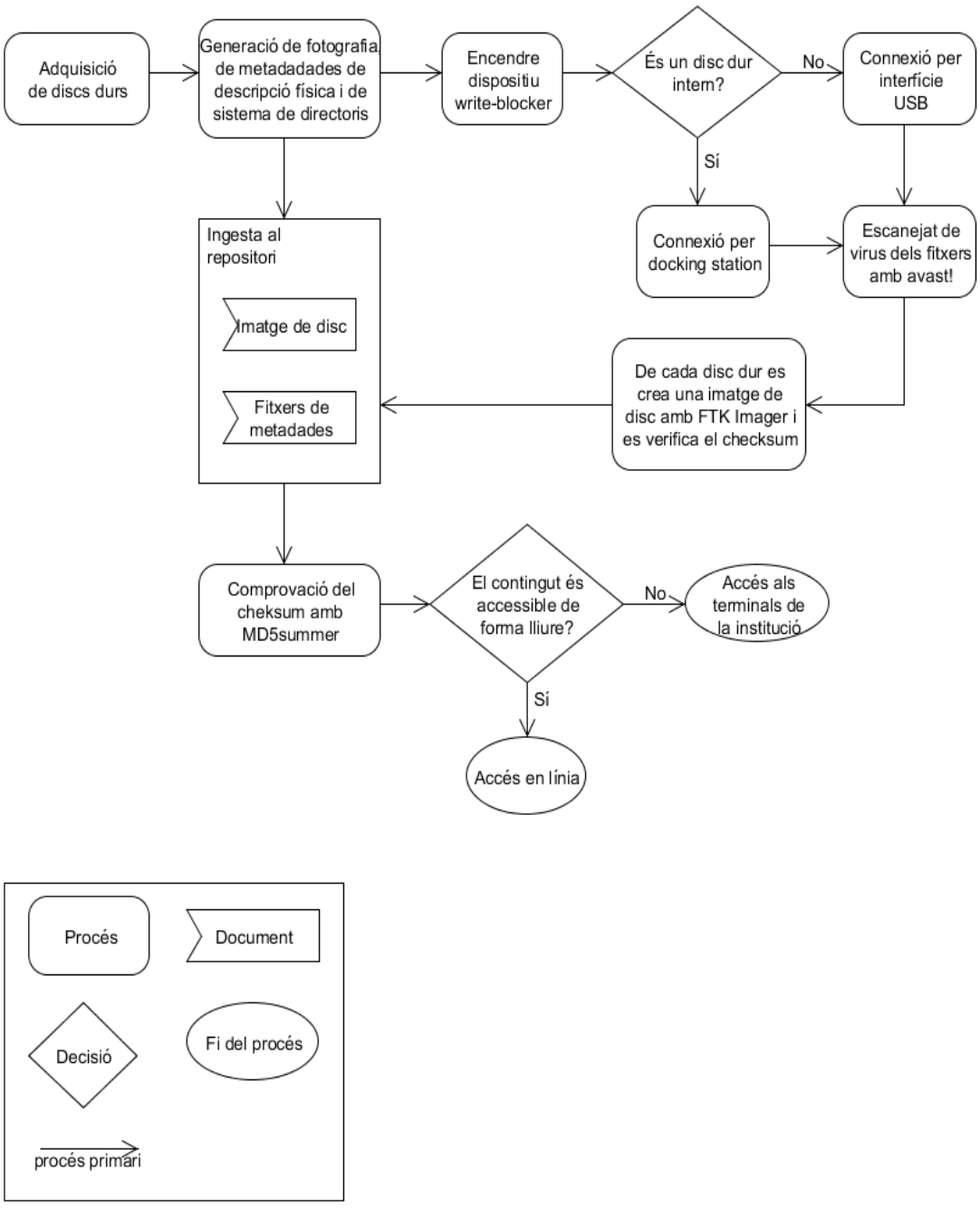


Figura 71. *Workflow* del procediment de treball amb discs durs amb *data management plans*

La diferència més gran és que la preservació es faria només amb les imatges de disc i no es faria selecció de fitxers. Els processos de verificació i comprovació del *checksum* serien els mateixos i la institució facilitarà l'accés a les dades en funció de la seva política de difusió.

5.5.3. Altres

El procediment amb la resta de suports físics seria molt semblant a la de l'emprat per als disquets, ja que la capacitat d'emmagatzematge de dispositius USB o CD-ROM/DVD-ROMs és baixa comparada a la dels discs durs. Algunes diferències que s'haurien de tenir en compte serien:

- **Memòria USB**. La seva entrada de dades es faria directament per la interfície USB, però sempre amb un *write-blocker* connectat per evitar l'escriptura de dades.
- **CD-ROM/DVD-ROM**. Són suports de només lectura, per tant no caldria l'ús de *write-blocker*. Si no es disposés d'un ordinador amb aquest tipus d'unitat, es pot connectar una unitat externa per la interfície USB.

6. PROPOSTA DE PLA D'ACCIÓ

Per tal de posar en marxa la unitat d'anàlisi forense digital, s'ha calculat el pressupost necessari (en dos escenaris), les diferents fases d'implementació, en quin calendari es pot aplicar el pla, com es faria el seguiment i l'avaluació, i finalment quines condicions s'aplicaran a la difusió del projecte.

6.1. PRESSUPOST

S'han calculat els pressupostos en funció dels dos escenaris dels requeriments específics de la unitat:

- Un de **nivell bàsic**, on s'ha d'adquirir un ordinador amb accessoris addicionals per fer anàlisi forense i no cal fer inversió en programari, ja que s'utilitzaria programari lliure.
- Un de **nivell avançat**, on s'ha d'adquirir programari i una estació de treball optimitzats per a aquesta tasca. S'han considerat diferents pressupostos en funció a dos tipus de maquinari forense (FRED i FRED SR), de programari forense (FTK i EnCase) i la formació corresponent al programari comercial, que es pot ampliar a altres cursos ja esmenats a *5.3.2. Formació*).

En ambdós casos s'han afegit les despeses necessàries per a la formació del personal, la seva remuneració i les despeses corrents i no corrents.

Tots els imports estan en euros. En alguns casos s'han hagut de calcular a partir del canvi de divises de dòlars nord-americans a euros degut a la no disponibilitat dels accessoris a la Unió Europea, com en el cas de la disquetera TEAC FD-55GFR. Per calcular el canvi de divises, s'ha utilitzat el web xe.com en data de 26 d'abril de 2013. També s'han desglossat els imports sense IVA i amb IVA. Per últim, no s'ha contemplat al pressupost l'adquisició d'ordinadors 'Rosetta' ni de la targeta controladora FC5025 degut a, en el primer cas, la dificultat de localitzar-ne als llocs de venda habituals i a, en el segon cas, que és impossible la seva compra a la Unió Europea.

El pressupost quedaria desglossat de la següent manera:

- **Maquinari i programari**. Els preus s'han obtingut dels següents distribuïdors i botigues en línia, en data de 30 d'abril de 2013 (no s'ha inclòs el programari lliure, ja que és de cost zero):
  - Insectra Technology Services[42]. Empresa distribuïdora amb seu a Madrid de maquinari i programari forense.
  - The Software Preservation Society. Distribuïdor de la targeta Kryoflux.
  - Ebay[43]. Dins aquesta famosa pàgina de subhastes *on-line* és possible trobar fàcilment disqueteres de 3 ½, de 5 ¼ i unitats Zip externes.
  - PC Componentes[44]. D'aquesta botiga en línia s'han tret els pressupostos per als components d'ordinador i el *docking station*.
- **Formació**. S'ha inclòs la formació corresponent estrictament a programari forense i els desplaçaments i dietes corresponents, necessaris per a aquesta formació.
- **Despeses corrents**. Consum d'electricitat, internet i fungibles necessaris.
- **Despeses no corrents**. S'han calculat els costos d'amortització del material, que és quant val la renovació de la instal·lació inicial (nou maquinari, nova formació, etc.) passat cinc anys, que equival a un 20%.
- **Personal**. Els costos s'han calculat en el supòsit de la contractació directa per part de la institució d'un cap de projecte i un enginyer informàtic en el règim comú (que són contractes que no generen IVA), amb els percentatges corresponents de càrregues a l'Estat, i contemplant la unitat d'anàlisi forense com una instal·lació permanent. Per aquest motiu, els imports de les despeses de funcionament són anuals, ja que el personal seria fix i de dedicació exclusiva. Dins els costos s'inclou també el cost total que ha de suportar la institució que contracti el personal, i no només els sous bruts dels treballadors.

[42] Pàgina web disponible a: <http://www.insectraforensics.com> [data de consulta: 30 abr. 2013]

[43] Pàgina web disponible a: <http://www.ebay.es/> [data de consulta: 30 abr. 2013]

[44] Pàgina web disponible a: <http://www.pccomponentes.com/> [data de consulta: 30 abr. 2013]

6.1.1. Maquinari (nivell bàsic)

Taula 3. Pressupost per a accessoris

ACCESSORIS	IMPORT	IVA (21%)	TOTAL
Write-blocker Tableau T35u Forensic Bridge	305,41	81,18	386,59
KryoFlux Personal Edition Premium	99,35	26,41	125,76
Disquetera de 5 1/4 interna TEAC FD-55GFR	102,66	27,29	129,95
Docking station 3 ½ i 2 ½ polzades IDE i SATA Startech	50,91	13,53	64,44
Disquetera de 3 ½ externa Freecom	31,37	8,34	39,71
Disquetera Zip externa Iomega 750 MB	34,72	9,23	43,95
Disquetera de 3 ½ interna Sony MPF920-E 3.5"	9,32	2,48	11,80
TOTAL COST ACCESORIS	**633,74**	**168,46**	**802,20**

Taula 4. Pressupost per a ordinador de nivell bàsic

COMPONENTS	IMPORT	IVA (21%)	TOTAL
Processador Intel Core I7-3820 3.60 GHz Box Socket 2011	204,61	54,39	259,00
Disc dur Hitachi Deskstar 7K1000 4 TB	152,47	40,53	193,00
Placa base Asrock X79 Extreme4-M	151,68	40,32	192,00
Sistema operatiu Windows 7 Professional 64 bits OEM Service Pack 1	98,75	26,25	125,00
Disc dur Seagate Barracuda 7200.14 3 TB GB SATA3	92,43	24,57	117,00
Torre Cooler Master CM-690 II Advanced USB 3.0	73,47	19,53	93,00
Monitor ASUS VH168D 15.6" LED	59,21	15,74	74,95
Font d'alimentació Aerocool Strike-X Power 600 W	46,61	12,39	59,00
Muntatge	35,55	9,45	45,00
Targeta gràfica Asus Geforce En210 Silent 1 GB GDDR3	22,12	5,88	28,00
Memòria RAM Corsair Value Select DDR3 1333 PC-10600 2GB	15,60	4,15	19,75
Gravadora/lectora CD-ROM/DVD-ROM Samsung SH-224BB 24X OEM	13,03	3,47	16,50
Ratolí Logitech PS/2	7,86	2,09	9,95
Teclat HP Standard PS/2	7,07	1,88	8,95
TOTAL COST ORDINADOR	**980,46**	**260,64**	**1241,10**

6.1.2. Maquinari i programari d'anàlisi forense

Taula 5. Pressupost per a programari d'anàlisi forense

PROGRAMARI	IMPORT	IVA (21%)	TOTAL
AccessData FTK	2.719,50	571,10	3.290,60
EnCase Forensic v7.06 (inclou servei de manteniment per 1 any)	3.237,25	860,54	4.097,79

Taula 6. Pressupost per a maquinari d'anàlisi forense

MAQUINARI	IMPORT	IVA (21%)	TOTAL
Unitat FRED estàndard	5.443,27	1446,95	6.890,22
Unitat FRED SR	10.701,70	2247,36	12.949,06

6.1.3. Formació

Taula 7. Pressupost per a formació d'anàlisi forense

CURS	IMPORT	IVA (21%)	TOTAL
AccessData 5 Day BootCamp	1.902,06	399,43	2.301,49
EnCase Computer Forensics I	1.902,06	399,43	2.301,49
EnCase Computer Forensics II	1.902,06	399,43	2.301,49

6.1.4. Cost de personal

Taula 8. Cost del cap de projecte

CONCEPTE	PERCENTATGE	IMPORT
Treballador		
Sou brut anual		44.000,00
Deducció de SS del treballador	4,70%	2.068,00
Retenció d'IRPF	23,00%	10.120,00
Atur	1,55%	682,00
Formació professional	0,10%	44,00
Sou net anual		31.086,00
Empresa		
Contingències comuns SS de l'empresa	23,60%	9.701,58
Atur	7,05%	2.898,14
FOGASA	0,20%	82,22
Formació professional	0,70%	287,76
Accidents de treball i malalties professionals	1,65%	678,29
Cost anual cap de projecte		57.647,99

Taula 9. Cost de l'enginyer informàtic

CONCEPTE	PERCENTATGE	IMPORT
Treballador		
Sou brut anual		36.700,00
Deducció de SS del treballador	4,70%	1.724,90
Retenció d'IRPF	20,00%	7.340,00
Atur	1,55%	568,85
Formació professional	0,10%	36,70
Sou net anual		27.029,55
Empresa		
Contingències comuns SS de l'empresa	23,60%	8.661,20
Atur	7,05%	2.587,35
FOGASA	0,20%	73,40
Formació professional	0,70%	256,90
Accidents de treball i malalties professionals	1,65%	605,55
Cost anual enginyer informàtic		48.884,40

6.1.5. Pressupostos totals (nivells bàsic, avançat amb unitat FRED i amb unitat FRED SR)

Taula 10. Costos d'instal·lació i de funcionament (nivell bàsic)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials			
Maquinari (accessoris)	633,74	133,09	766,83
Maquinari (2 ordinadors)	1.960,92	411,79	2.372,71
Mobiliari	1.331,00	279,51	1.610,51
Altres depeses d'instal·lació (pintura, aire condicionat…)	3.000,00	630,00	3.630,00
Connexions informàtiques	800,00	168,00	968,00
Total	**7.725,66**	**1.622,39**	**9.348,05**
Despeses de funcionament (anuals)			
A. Despeses corrents			
Electricitat	600,00	126,00	726,00
Internet	400,00	84,00	484,00
Fungibles (paper, tòner…)	600,00	126,00	726,00
Total	**1.600,00**	**336,00**	**1.936,00**
B. Despeses no corrents			
Amortització de material (20%)			1.869,61
Total			1.869,61
C. Despeses de personal			
Cap de projecte			57.647,99
Enginyer informàtic			48.884,40
Total			**106.532,39**

Taula 11. Pressupost total (nivell bàsic)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials	7.725,66	1.622,39	9.348,05
Despeses de funcionament (anuals)			
A. Despeses corrents	1.600,00	336,00	1.936,00
B. Despeses no corrents			1.869,61
C. Despeses de personal			106.532,39
Total	**9.325,66**	**1.958,39**	**119.686,05**

Taula 12. Costos d'instal·lació i de funcionament (nivell avançat - FRED)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials			
Maquinari (1 unitat FRED)	5.443,27	1.446,95	6.890,22
Programari (FTK)	2.719,50	571,10	3.290,60
Programari (EnCase)	3.237,25	679,82	3.917,07
Formació (FTK)	1.902,06	399,43	2.301,49
Formació (EnCase)	3.804,12	798,87	4.602,99
Desplaçaments i dietes	5.000,00	1.050,00	6.050,00
Mobiliari	1.331,00	279,51	1.610,51
Altres depeses d'instal·lació (pintura, aire condicionat…)	3.000,00	630,00	3.630,00
Connexions informàtiques	800,00	168,00	968,00
Total	**27.237,20**	**6.023,68**	**33.260,88**
Despeses de funcionament (anuals)			
A. Despeses corrents			
Electricitat	600,00	126,00	726,00
Internet	400,00	84,00	484,00
Fungibles (paper, tòner…)	600,00	126,00	726,00
Total	**1.600,00**	**336,00**	**1.936,00**
B. Despeses no corrents			
Amortització de material (20%)			6.652,18
Total			**6.652,18**
C. Despeses de personal			
Cap de projecte			57.647,99
Enginyer informàtic			48.884,40
Total			**106.532,39**

Taula 13. Pressupost total (nivell avançat - FRED)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials	27.237,20	6.023,68	33.260,88
Despeses de funcionament (anuals)			
A. Despeses corrents	1.600,00	336,00	1.936,00
B. Despeses no corrents			6.652,18
C. Despeses de personal			106.532,39
Total	**28.837,20**	**6.359,68**	**148.381,45**

Taula 14. Costos d'instal·lació i de funcionament (nivell avançat - FRED SR)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials			
Maquinari (1 unitat FRED SR)	10.701,70	2.247,36	12.949,06
Programari (FTK)	2.719,50	571,10	3.290,60
Programari (EnCase)	3.237,25	679,82	3.917,07
Formació (FTK)	1.902,06	399,43	2.301,49
Formació (EnCase)	3.804,12	798,87	4.602,99
Desplaçaments i dietes	5.000,00	1.050,00	6.050,00
Mobiliari	1.331,00	279,51	1.610,51
Altres depeses d'instal·lació (pintura, aire condicionat…)	3.000,00	630,00	3.630,00
Connexions informàtiques	800,00	168,00	968,00
Total	**32.495,63**	**6.824,08**	**39.319,71**
Despeses de funcionament (anuals)			
A. Despeses corrents			
Electricitat	600,00	126,00	726,00
Internet	400,00	84,00	484,00
Fungibles (paper, tòner…)	600,00	126,00	726,00
Total	**1.600,00**	**336,00**	**1.936,00**
B. Despeses no corrents			
Amortització de material (20%)			7.863,94
Total			**7.863,94**
C. Despeses de personal			
Cap de projecte			57.647,99
Enginyer informàtic			48.884,40
Total			**106.532,39**

Taula 15. Pressupost total (nivell avançat - FRED SR)

CONCEPTE	IMPORT	IVA (21%)	TOTAL
Despeses inicials	32.495,63	6.824,08	39.319,71
Despeses de funcionament (anuals)			
A. Despeses corrents	1.600,00	336,00	1.936,00
B. Despeses no corrents			7.863,94
C. Despeses de personal			106.532,39
Total	**34.095,63**	**7.160,08**	**155.652,04**

6.2. IMPLEMENTACIÓ

Dins de la planificació de la implementació s'han especificat quines són les tasques necessàries per a la posada en marxa de la unitat digital forense en cinc fases diferents, que inclouen la definició de l'estratègia que vulgui utilitzar la institució, la formació del personal, les tasques de condicionament de l'espai, la instal·lació tècnica amb els tests necessaris i el procés d'introducció de continguts al repositori institucional.

FASE 1. KICK OFF

- **Objectiu**: comunicació de la proposta de la unitat digital forense als responsables de la institució i presentació de tot personal que operarà a la unitat.
- **Durada**: 1 dia
- **Tasques:**
    - Presentació del projecte a la institució
    - Presentació de l'equip
    - Comunicació dels avantatges de les tècniques de l'anàlisi digital forense dins els objectius de preservació del patrimoni cultural
    - Presentació del calendari i de les fases d'implementació
- **Lliurables:**
    - Informe detallat sobre l'anàlisi digital forense aplicat a les biblioteques i arxius que inclou un *benchmarking* de cinc unitats ja existents
- **Personal**:
    - Cap de projecte
    - Enginyer informàtic

FASE 2. ESTRATÈGIA

- **Objectiu**: analitzar els objectius de la institució en funció del seu fons de material *born digital* actual i el que preveu obtenir en el futur.
- **Durada**: 1 setmana
- **Tasques:**
  - Reunió/entrevista amb el responsable de la institució
- **Lliurables:**
  - Informe d'objectius de la institució
- **Personal**:
  - Cap de projecte

FASE 3. PRELIMINARS DE LA UNITAT

- **Objectiu**: preparar al personal destinat a la unitat per poder executar les tasques d'anàlisi digital forense i tenir un entorn de treball preparat per a les necessitats de la unitat
- **Durada**: 2 setmanes
- **Tasques:**
  - Desplaçaments per rebre la formació en anàlisi digital forense (en el cas d'un escenari de nivell avançat)
  - Formació en anàlisi digital forense
  - Condicionament de l'espai que ocuparà la unitat amb el material necessari (mobiliari, pintura, prestatgeries, connexions a Internet, endolls, etc.)
- **Lliurables:**
  - Certificat de formació en anàlisi digital forense (en el cas d'un escenari de nivell avançat)
  - Entorn de treball preparat
- **Personal**:
  - Cap de projecte
  - Enginyer informàtic
  - Personal extern per al condicionament de la unitat

FASE 4. IMPLEMENTACIÓ TECNOLÒGICA

- **Objectiu**: posar a punt el programari i el maquinari per tal que pugui analitzar, recuperar i preservar material *born digital*
- **Durada**: 6 setmanes
- **Tasques:**
  - Instal·lació del maquinari i del programari
  - Test del programari necessari, dins l'ordre lògic dels *workflows* establerts
  - Test d'integració de continguts *born digital* al repositori institucional
  - Elaboració de procediments de treball
- **Lliurables:**
  - Informe de resultats del maquinari i del programari
  - Guies de treball amb *workflows* dissenyats per a les diferents casuístiques
- **Personal**:
  - Cap de projecte
  - Enginyer informàtic

FASE 5. INGESTA AL REPOSITORI

- **Objectiu**: comprovar la integració del contingut *born digital* dins el repositori institucional
- **Durada**: 2 setmanes
- **Tasques:**
  - Comprovació dels *checksums* dels continguts incorporats al repositori per garantir la integritat de les dades
  - Revisió dels continguts incorporats al repositori
  - Validació dels continguts
- **Personal**:
  - Cap de projecte
  - Enginyer informàtic

6.3. CALENDARI

La durada prevista per a tenir enllestida la unitat d'anàlisi forense seria de dotze setmanes aproximadament. S'ha elaborat el següent calendari amb el detall de cada fase i les seves tasques corresponents:

FASES I TASQUES	SETMANES											
	1	2	3	4	5	6	7	8	9	10	11	12
Fase 1. Kick off												
Fase 2. Estratègia												
Definició d'objectius												
Fase 3. Preliminars de la unitat												
Formació en programari forense												
Condicionament de l'espai												
Fase 4. Implementació tecnològica												
Instal·lació del maquinari												
Instal·lació del programari												
Test creació d'imatges de disc												
Test generació de *checksums*												
Test de suites d'anàlisi forense digital												
Test d'identificació de fitxers												
Test de validació de fitxers												
Test d'introducció de continguts												
Elaboració de guies de treball												
Fase 5. Ingesta al repositori												
Comprovació de *checksums*												
Revisió de continguts al repositori												
Validació												

Figura 72. Diagrama de Gantt amb la durada de cada una de les fases

6.4. SEGUIMENT I AVALUACIÓ

Per tal de fer un seguiment de qualitat de les tasques en quant el funcionament de la unitat, es poden fer dos tipus de controls:

- **Control a nivell general**. Un any després de la posada en marxa de la unitat, es pot avaluar si s'han assolit els objectius amb la pràctica d'una sèrie d'indicadors:
    - El programari i maquinari donen solucions als objectius plantejats per la unitat?
    - S'hauria d'ampliar l'abast de suports acceptats?
    - És necessari revisar els processos de treball?
    - La formació del personal és suficient per poder assolir els objectius?
    - L'equipament de la unitat és adequat per a les activitats de la unitat?

  Els instruments de seguiment i avaluació poden ser molt diversos, com:
    - Informes anuals.
    - Plans de treball.
    - Enquestes.
    - Anàlisi fet per empreses externes.
    - Entrevistes.
    - Reunions de *focus group*.

  Si aquestes qüestions es resolen negativament, caldrà ampliar o reduir les prestacions de la unitat i al cap d'un any fer una nova avaluació per comprovar l'eficàcia de les mesures.

- **Control continuat sobre actuacions concretes.** Quan la unitat rebi els materials *born digital* per preservar, s'haurien de fer valoracions de forma periòdica (mensual o anualment) i treure conclusions amb la creació d'estadístiques. Per exemple, si la unitat rep una donació de 100 disquets, s'haurà de determinar:
    - Grau d'incidències:
        - S'han pogut rescatar continguts malmesos?
        - Hi havia algun disquet que no es va poder recuperar?

- Grau de productivitat:
  - El temps que es triga des de l'arribada del material fins que s'integra al repositori és adequat?
  - El programari és efectiu per reduir el temps de producció o s'hauria de crear un de nou?
- Grau d'eficàcia de les ingestes:
  - Triga massa temps el repositori en processar les imatges?
  - Les imatges es corrompen quan es pugen al repositori?

Un cop hagi acabat l'actuació, mitjançant l'informe de resultats es podrà fer una valoració vers els resultats de la unitat i així prendre les mesures oportunes.

6.5. POLÍTICA DE DIFUSIÓ

Es poden contemplar dues vies de difusió del projecte: que sigui exclusivament d'ús intern per part de la institució o bé permetre que sigui obert el seu ús per a tercers.

El més adient seria la segona opció, ja que actualment no existeixen unitats d'anàlisi forense digital a l'àmbit territorial català i es podrien beneficiar biblioteques i arxius del voltant per recuperar continguts *born digital* i incorporar-los a les seves col·leccions. Al CRAI de la Universitat de Barcelona, existeix el Taller de Restauració que s'ocupa de restaurar llibres antics i també accepta comandes d'altres institucions i de particulars. Per altra banda, es podria oferir el servei a canvi d'una tarifa en funció del tipus de suport que generaria ingressos que ajudarien a l'amortització de les despeses d'instal·lació de la unitat.

7. CONCLUSIONS

Com ja s'ha exposat, la problemàtica que presenten els suports amb material *born digital* s'ha de considerar dins els centres responsables del patrimoni cultural. A institucions anglosaxones ja s'han pres mesures per preservar aquests continguts, que en alguns casos ha obligat a crear grups de treball específics i aplicar solucions en funció de la política de cada centre. Però, realment és viable i justificable aquesta instal·lació dins la Biblioteca de Catalunya (o altres centres semblants de Catalunya)? A tall il·lustratiu, aquests són els suports amb material *born digital* presents als catàlegs de la Biblioteca de Catalunya i de la Universitat de Barcelona (a data de 6 de juny de 2013):

Taula 16. Suports amb material *born digital* a la BC i UB (xifres en nombre de registres dels catàlegs)

CENTRE	Disquets	CD-ROM	DVD-ROM	Memòries USB
Biblioteca de Catalunya	171	838	20	14
CRAI de la UB	76	476	11	4

És especialment destacable l'augment de memòries USB, que està substituint el suport DVD-ROM d'algunes publicacions oficials de la Generalitat de Catalunya. Si afegim això a la quantitat important de disquets als fons dels centres, es pot concloure que una unitat forense digital és necessària per evitar la pèrdua de continguts als discs magnètics i també per preservar memòries USB, que ja han desplaçat als discs òptics com suport d'emmagatzematge de dades. Tampoc s'han d'obviar els CD-ROMs i DVD-ROMs, que també estan subjectes al deteriorament però en menor mesura que els disquets.

Sobre la viabilitat econòmica de la unitat a la Biblioteca de Catalunya, existeixen diverses opcions que es poden considerar. Certament, la reducció progressiva del pressupost és un llast important (al 2009 fou d'onze milions d'euros, mentre que pel 2012 es va quedar en poc més de vuit milions), però si es parteix d'un pressupost de nivell bàsic i s'ajusten les despeses de personal amb una reducció de dedicació horària

d'un 50%, el cost inicial suposaria poc més de nou mil euros i les despeses anuals serien de una mica menys de 60.000. Recordem, a més, que la unitat donaria servei a altres institucions i això reduiria una mica els costos.

Altra opció seria no contractar nou personal i aprofitar els professionals del Grup de Preservació Digital de la BC, el qual està format actualment per:

- Karibel Pérez i Ramon Novoa (Àrea de Tecnologia de la Informació)
- Paquita Navarro (Unitat de Digitalització)
- Margarida Ullate (Unitat de Sonors i Audiovisuals)
- Sergi Font (projecte Google)
- Ciro Llueca (PADICAT)
- Eugènia Serra (Coordinació General)

Aquest grup té molta experiència en projectes de preservació digital (com per exemple el projecte COFRE, que està orientat a la creació d'un repositori) però igualment haurien de rebre una formació específica per tal de fer servir satisfactòriament el maquinari i programari d'anàlisi forense digital.

A tall de conclusió, es pot dir que una unitat d'anàlisi digital forense no és una opció, sinó que és totalment necessària per evitar la pèrdua de materials *born digital* que, malauradament, és possible que molts ja s'hagin perdut per sempre. Aquest treball, per tant, es podrà dir que ha aconseguit el seu objectiu quan es trobi operativa una unitat a la Biblioteca de Catalunya i doni servei a altres institucions.

GLOSSARI

AES. *Advanced Encryption Standard* (estàndard d'encriptació avançada). Especificació per a encriptació de dades electròniques establert pel NIST (National Institute of Standards and Technology). Existeixen tres tipus de clau de xifrat: 128, 192 i 256 bits.

Arxiu híbrid. Col·lecció documental formada per materials tradicionals en paper i també per artefactes digitals (ordinadors i suports d'emmagatzematge de dades).

Arxiver digital. Professional amb un ampli coneixement de les funcions i tasques d'arxivística i versat en eines de preservació digital per tal de gestionar material *born digital.*

Born digital. Terme que s'aplica a materials creats en forma digital.

Checksum. Funció que detecta canvis accidentals en una seqüència de dades per protegir la integritat dels mateixos i verificar que no hagi canvis. El resultat és un valor *hash* que identifica unívocament les dades. Hi ha de diferents tipus en funció del seva complexitat, com MD5, SHA-1, SHA-2, etc.

CLV. *Constant Linear Velocity* (línia de velocitat constant). Mètode constant de lectura i escriptura en un disquet o CD-ROM.

Densitat. Disposició lògica i física de les dades emmagatzemades a un disquet.

Emulació. Recreació de l'entorn en què es va crear originalment un objecte digital. Un exemple és l'emulació del sistema informàtic d'Apple Macintosh que va desenvolupar l'Emory University.

FM. *Frequency Modulation* (freqüència modulada) i també anomenat BMC (biphase mark code). Mètode per codificar dades digitals als primers disquets de simple densitat, amb una taxa de transferència de dades de 250 KB per segon.

GCR. *Group code recording* (gravació de codi en grup). Mètodes de codificació per a diferents suports magnètics. En el cas dels disquets, fou un mètode desenvolupat per a l'ordinador Apple II.

Hash. Vegeu *Checksum*

Imatge de disc o imatge forense. Representació en forma de bits d'un suport amb contingut digital (com un disquet o un CD-ROM) sector per sector. En principi és un sol fitxer que representa tot el suport. La seva mida és variable: per a un CD-ROM seria d' aproximadament 700 MB, però per a un disc dur pot tenir una mida de 4 GB, 8 GB, 40 GB, etc.

MFM. Modified Frequency Modulation (freqüència modulada modificada). Mètode per codificar dades digitals a suports magnètics com els disquets, amb una taxa de transferència de dades de 500 KB per segon.

Migració. Transferència de dades a entorns informàtics més actuals. Això pot incloure la conversió de recursos d'un format de fitxer a un altre (per exemple, la conversió d'un fitxer Word a un PDF) o d'un sistema operatiu a un altre (per exemple, de Windows a Linux).

Refreshing. Transferència de dades d'un suport físic a un altre sense cap tipus de canvi o alteració de dades. Per exemple, passar dades d'un disquet a un disc dur o d'un CD-ROM a un DVD-ROM.

Sistema de fitxers. Mètode per emmagatzemar i organitzar fitxers d'ordinador i les dades que contenen per tal de facilitar-ne la localització i accés.

Workflow. Descripció d'una seqüència d'operacions d'una persona, un grup de persones, una organització o un o més mecanismes.

Write-blocker. Dispositiu que evita qualsevol alteració dins un suport amb contingut digital, ja que actua com a un bloquejador d'escriptura. Existeixen programaris que actuen com a *write-blockers*, però el maquinari és més fiable.

BIBLIOGRAFIA

Llibre

AIMS Work Group. *AIMS born-digital collections: an inter-institutional model for stewardship.* AIMS Work Group, 2012. Disponible en línia a: <http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final_A4.pdf> [consulta: 15 febr. 2013]

John, Jeremy Leighton. *Digital forensics and preservation.* Great Britain : Digital Preservation Coalition, 2012. 60 p. ISSN 2048-7916. Disponible en línia a: <http://www.dpconline.org/component/docman/doc_download/810-dpctw12-03pdf> [consulta: 9 febr. 2013]

Kirschenbaum, Matthew G. ; Ovenden, Richard ; Redwine, Gabriela. *Digital forensics and born-digital content in cultural heritage collections.* Washington, DC : Council on Library and Information Resources, 2010. VIII, 93 p. (CLIR publication ; 149). ISBN 978-1-932326-37-6. Disponible en línia a: <http://www.clir.org/pubs/reports/pub149/reports/pub149/pub149.pdf> [consulta: 9 febr. 2013]

Lee, Christopher A. *I, digital: personal collections in the digital era.* Chicago : Society of American Archivists, cop. 2011. 379 p. ISBN 1-931666-38-5.

Article

Doherty, Sean. "Product review: Encase Forensic 7" [en línia]. A: *LTN.* 29 gen. 2013**.** Disponible en línia a: <http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202584495563&Product_Review_Encase_Forensic_7/> [consulta: 5 maig 2013]

Li-Madeo, Carolyn. "Digital archives and born-digital materials" [en línia]. A: *Antelope as document.* 30 oct. 2012**.** Disponible en línia a: <http://antelopeasdocument.wordpress.com/2012/10/30/digital-archives-and-born-digital-materials/> [consulta: 23 febr. 2013]

Loftus, Mary J. "The Author's desktop" [en línia]. A: *Emory Magazine.* Winter 2010**.** Disponible en línia a: <http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html> [consulta: 8 març 2013]

Pozo, Nicholas del ; Elford, Douglas ; Pearson, David. "Prometheus: managing the ingest of media carriers". A: *Proceedings of DigCCurr 2009*, Digital Curation Practice, Promise and Prospects, University of North Carolina at Chapel Hill, North Carolina 2009**.** Disponible en línia a: <http://www.slideshare.net/natlibraryofaustralia/prometheus-13399586> [consulta: 23 febr. 2013]

Gengenbach, Martin J. "The way we do it here: mapping digital forensics workflows in collecting institutions." A Master's Paper for the M.S. in L.S degree. August, 2012. <http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf> [consulta: 23 febr. 2013]

Presentació

An introduction to the futureArch project and BEAM. **Cliff, Pete.** Oxford : s.n., 23rd April 2009. Digital Repositories Workshop: Tools and Infrastructure. Disponible en línia a: <http://www.slideshare.net/pixelatedpete/pete-cliff-drw0409#Introduction%20to%20the%20futureArch%20project%20and%20BEAM> [consulta: 10 febr. 2013]

Defining the role of digital archivist. **Smith, Kari R.** MIT Libraries, 2013. Disponible en línia a: <http://libraries.mit.edu/archives/digital-archives/blog/presentation_DefiningDARole.pdf> [consulta: 30 març 2013]

Hybrid teams for hybrid archives: collaboration and born-digital archives. **Carroll, Laura L ; Farr, Erika.** RMBS, 2010. Disponible en línia a: <http://www.rbms.info/conferences/preconfdocs/2010/SeminarICarroll_Farr.pdf> [consulta: 24 febr. 2013]

Informe

Elford, Douglas et al. "Media Matters: developing processes for preserving digital objects on physical carriers at the National Library of Australia". A: World Library and Information Congress. 74th IFLA General Conference and Council, 10-14 August 2008, Québec 2008. <http://archive.ifla.org/IV/ifla74/papers/084-Webb-en.pdf> [consulta: 28 març 2013]

John, Jeremy Leighton. "Adapting existing technologies for digitally archiving personal lives: digital forensics, ancestral computing, and evolutionary perspectives and tools." 2008. <http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf> [consulta: 23 febr. 2013]

LeClaire, Yvonne. "The Forensic process examined: creating cases for classroom use". 2012. <http://cs.lewisu.edu/mathcs/msis/projects/msis595_YvonneLeClaire.pdf> [consulta: 2 maig 2013]

Meeks, Elijah. "Robert Creeley e-mail correspodence network". Disponible en línia a: <https://dhs.stanford.edu/visualization/robert-creeley-e-mail-correspodence-network/> [consulta: 23 febr. 2013]

McMillon, Matthew. "Building a low cost forensics workstation". Disponible en línia a: <http://www.sans.org/reading_room/whitepapers/incident/building-cost-forensics-workstation_895> [consulta: 29 març 2013]

Pérez, Karibel; Serra, Eugènia. "Repositori de preservació digital de la Biblioteca de Catalunya: informe descriptiu i de situació." 2010. Disponible en línea a: <http://www.recercat.cat/handle/2072/97251> [consulta: 16 març 2013]

Pàgina web

AcessData [en línia] <http://www.accessdata.com> [consulta: 17 abril 2013]

Biblioteca de Catalunya [en línia] <http://www.bnc.cat/> [consulta: 15 març 2013]

BitCurator [en línia] <http://www.bitcurator.net> [consulta: 29 març 2013]

Device Side Data [en línia] <http://www.deviceside.com> [consulta: 2 març 2013]

Digital Intelligence [en línia] <http://www.digitalintelligence.com/> [consulta: 15 abril 2013]

futureArch [en línia] <http://futurearchives.blogspot.com.es> [consulta: 10 febr. 2013]

Guidance Software [en línia] <http://www.guidancesoftware.com/> [consulta: 17 abril 2013]

Kryoflux [en línia] <http://www.kryoflux.com> [consulta: 2 març 2013]

ÍNDEX DE FIGURES I TAULES

Figures

Figura 1. Disquet de 5 ¼ polzades (Apple) 6
Figura 2. Disquet de 3 ½ polzades 7
Figura 3. Discs durs SATA i IDE 7
Figura 4. Disc Zip Iomega 8
Figura 5. CD-ROM 8
Figura 6. Memòria USB de 4 GB (Sandisk) 8
Figura 7. Disc dur USB extern (Western Digital) 9
Figura 8. Disquet de 8 polzades 10
Figura 9. Disquets de 2, de 2,8 i de 3 polzades (Fuji Film, Smith Corona i Maxell) 11
Figura 10. Disc LS-120 (Imation) 11
Figura 11. Discs magnetoòptics (Sony, Fujitsu i Olympus) 12
Figura 12. Ordinador ZX Spectrum amb lector/gravador de casset incorporat 12
Figura 13. Targetes de memòria (Sandisk, Olympus i Sony) 12
Figura 14. Targetes de mòbil SIM 13
Figura 15. Memòries RAM 13
Figura 16. Targeta perforada (IBM) 14
Figura 17. Cinta magnètica (3M) 14
Figura 18. Disquet de la col·lecció Stephen Jay Gould 22
Figura 19. Diagrama de la xarxa de correu electrònic de Robert Creeley 23
Figura 20. Text migrat al format PDF (arxiu de Barbara Castle) 26
Figura 21. Text original en format LocoScript de l'ordinador Amstrad PCW (arxiu de Barbara Castle) 27
Figura 22. Disquet de 3 ½ polzades corresponent a l'arxiu de la impremta Clutag 27
Figura 23. Creació de la imatge d'un disc dur mitjançant una unitat FRED 28
Figura 24. Unitat personalitzada per a la duplicació de CD-ROMs a la NLA 30
Figura 25. Llistat de tasques a Prometheus 31
Figura 26. Pantalla d'inici de sessió del repositori Rescue 34
Figura 27. Ordinadors donats per Salman Rushdie 35

Figura 28. Emulació del sistema operatiu del Macintosh Performa 5400 36
Figura 29. Base de dades de l'arxiu digital de Salman Rushdie..... 36
Figura 30. Torre d'ordinador amb compartiments externs 39
Figura 31. Unitat FRED SR..... 41
Figura 32. Creació d'imatge de disc amb FTK Imager 46
Figura 33. Informe de la imatge amb *hash* MD5 i SHA1 en FTK Imager 47
Figura 34. *Evidence Tree* en FTK Imager..... 47
Figura 35. Opció de propietats de fitxer en FTK Imager..... 48
Figura 36. Visualització de la llista de fitxers en FTK Imager..... 48
Figura 37. Visor hexadecimal a FTK Imager..... 49
Figura 38. Creació d'imatge de disc amb DTC 51
Figura 39. Lectura de pistes de disquet amb DTC 52
Figura 40. Selecció de fitxers i carpetes a MD5summer 54
Figura 41. Generació de *checksums* a MD5summer 54
Figura 42. Generació de *checksum* SHA1 amb HashX 56
Figura 43. Visualització de fitxers orfes amb el text recuperat a Autopsy..... 58
Figura 44. Visualització de tipus de fitxers a Autopsy 59
Figura 45. Opcions de visualització de dades a FTK 60
Figura 46. Visualitzador de l'explorador a FTK..... 62
Figura 47. Visualització general i del correu electrònic a FTK..... 63
Figura 48. Visualització dels gràfics a FTK..... 63
Figura 49. Opcions per obrir dispositius locals a EnCase Forensic..... 65
Figura 50. Opcions d'EnCase Forensic per processar fitxers..... 66
Figura 51. Sistema de cerca a EnCase Forensic..... 67
Figura 52. Reconeixement de formats de fitxer a DROID 69
Figura 53. Informació d'un fitxer JPG a JHOVE..... 70
Figura 54. Informació d'un fitxer PDF a JHOVE 71
Figura 55. Informació d'un fitxer WAV a JHOVE 72
Figura 56. Ordinador 'Rosetta' amb unitats de disquet instal·lades..... 73
Figura 57. Unitat FRED..... 74
Figura 58. Part anterior de la targeta Kryoflux..... 77
Figura 59. Part posterior de la targeta Kryoflux..... 77

Figura 60. Part anterior de la targeta FC5025 79
Figura 61. Part posterior de la targeta FC5025 79
Figura 62. Part frontal de la unitat de disquet de 5 ¼ TEAC FD-55GFR..... 81
Figura 63. Part posterior de la unitat de disquet de 5 ¼ TEAC FD-55GFR..... 82
Figura 64. Unitat de disquet de 3 ½ USB (Freecom)..... 82
Figura 65. Unitat de disquet de 3 ½ interna (Sony)..... 83
Figura 66. Unitat Zip externa i disquet Zip de 750 MB (Iomega) 83
Figura 67. *Docking station* amb connexió SATA superior i connexió IDE frontal (StarTech)..... 84
Figura 68. Disc dur (Maxtor) connectat a un *write-blocker* (Tableau) 85
Figura 69. *Workflow* del procediment de treball amb disquets 94
Figura 70. *Workflow* del procediment de treball amb discs durs amb arxius personals 96
Figura 71. *Workflow* del procediment de treball amb discs durs amb *data management plans*..... 98
Figura 72. Diagrama de Gantt amb la durada de cada una de les fases 111

Taules

Taula 1. Comparativa de *checksums* 56
Taula 2. Maquinari i programari forense a les unitats ja existents..... 91
Taula 3. Pressupost per a accessoris 102
Taula 4. Pressupost per a ordinador de nivell bàsic..... 102
Taula 5. Pressupost per a programari d'anàlisi forense..... 103
Taula 6. Pressupost per a maquinari d'anàlisi forense..... 103
Taula 7. Pressupost per a formació d'anàlisi forense..... 103
Taula 8. Cost del cap de projecte 104
Taula 9. Cost de l'enginyer informàtic..... 104
Taula 10. Costos d'instal·lació i de funcionament (nivell bàsic)..... 105
Taula 11. Pressupost total (nivell bàsic) 105
Taula 12. Costos d'instal·lació i de funcionament (nivell avançat - FRED)..... 106
Taula 13. Pressupost total (nivell avançat - FRED) 106

Taula 14. Costos d'instal·lació i de funcionament (nivell avançat - FRED SR)..... 107

Taula 15. Pressupost total (nivell avançat - FRED SR) 107

Taula 16. Suports amb material *born digital* a la BC i UB (xifres en nombre de registres dels catàlegs) 114

ANNEX. FORMULARI DE DONACIÓ

Dades personals del donant

Nom	
Cognoms	
NIF/Passaport	
Adreça	
Població	
Codi postal	
Telèfon	
Mòbil	
Adreça electrònica	

Tipus de suport:

☐ Disquets (3 ½ polzades, 5 ¼ polzades i Zip)

☐ Discs durs interns

☐ Discs durs externs amb connexió USB

☐ Discs òptics (CD-ROM i DVD-ROM)

☐ Memòries USB

Tipus de contingut:

☐ Documents

☐ Fotografies

☐ Vídeo

☐ Àudio

☐ Correu electrònic

Sistema operatiu que es va utilitzar originalment:

☐ Windows

☐ MS-DOS

☐ OS X

☐ Linux

☐ Altres (especificar):

Data d'últim ús del suport:

Programes que es van utilitzar per crear el contingut (processadors de text, editors d'àudio, etc.):

Signatura del donant:	Data: