Project Information			
Project Acronym	FIDO		
Project Title	Forensic Investigation of Digital Objects		
Start Date	February 2011	**End Date**	July 2011
Lead Institution	King's College London		
Project Director	Gareth Knight		
Project Manager & contact details	Gareth Knight Centre for e-Research, King's College London gareth.knight@kcl.ac.uk		
Project Web URL	http://fido.cerch.kcl.ac.uk		
Programme Name (and number)	Information Environment 09-11: Preservation Tools		
Programme Manager	Neil Grindley		

Document Name			
Document Title	Final Report		
Reporting Period			
Author(s) & project role	Gareth Knight, Project Manager		
Project team	Mike Bryant, Software Developer Kate O'Brien, Digital Records Officer Lindsay Ould, Information Manager and Digital Archivist		
Date	01 September 2012	**Filename**	
URL			
Access	☑ Project and JISC internal	☑ General dissemination	

Document History		
Version	**Date**	**Comments**
1.0	20 September 2011	First version
1.1	04 August 2012	Output and Results section updated to include presentations given following the project's completion
1.2	1 September 2012	Added reference to iJDC article being accepted

Contents

Acknowledgements

The project was funded by JISC under the Preservation Tools strand of the JISC Information Environment 2009-11 programme. Neil Grindley, the programme manager gave support and encouragement throughout the project's lifetime.

The FIDO project was a collaboration between the Centre for e-Research (CeRch) and the Archives and Information Management (AIM) Service at King's College London. CeRch is located within Information Services and Systems, and carries out R&D activities in information management, repositories, metadata, preservation, and e-research infrastructures, both internally for King's and for externally funded projects, building on experience gained when it hosted the Arts and Humanities Data Service. AIM is responsible for the College's archives, corporate records management, legal compliance, business continuity, and information resources. The latter includes collection development and management, the acquisition and cataloguing of library materials and information resources (including electronic resources), document delivery and interlibrary loans and special collections.

We would like to thank Patricia Methven, Sheila Anderson, and Mark Hedges for their contribution to the project management board and advice during the project.

Executive Summary

The Higher Education archive collects a broad range of materials, ranging from corporate business records to the personal collections of academics and notable individuals. Traditionally, information collected by the archives has been provided in analogue form. However, the growth of computing technology during the previous three decades has resulted in a large amount of born-digital material being produced. As their creators retire or pass away, an increasing amount of this digital information is being offered to the HE archive. For the Archivist or Digital Curator, the handling and processing of such personal data collections represents a challenge, requiring the development of new processes to ensure that information content held in diverse media types, file systems, data structures are located and transferred into a managed environment.

The Forensic Investigation of Digital Objects (FIDO) project applied a coordinated and strategic holistic approach to the management of digital material that bought together archival theory with more recently developed practices within digital forensics. It demonstrated how these complementary disciplines may be used to process two categories of digital equipment that are increasingly being provided by donors to the college archive:

1. Computer systems, such as Windows PCs and Apple Macs that are donated to the archives or remain in active use by college staff;
2. Digital media formats, such as floppy disks, CD-ROM, DVD-ROM, USB sticks, and SD solid state storage, among others.

The project demonstrated the feasibility of building a low-cost forensic environment based upon off-the-shelf hardware and free software which may be used by digital archivists and digital curators to acquire data held on diverse media types in a manner that ensures it is captured in its entirety and maintains its integrity. It went on to analyse working practices applied by forensic experts to acquire, examine, and analyse digital data, and developed a set of processes and procedures that could be applied to process personal digital collections.

1. Introduction

1.1. Background

The Higher Education archive collects a broad range of materials, ranging from corporate business records to the personal collections of academics and notable individuals. Traditionally, information collected by the archives has been provided in analogue form. However, the growth of computing technology has resulted in a large amount of born-digital material being produced. As their creators retire or pass away, an increasing amount of this digital information is being offered to the HE archive. For the Archivist or Digital Curator, the handling and processing of such personal data archives represents a challenge, requiring the development of new processes to ensure that information content held in diverse media types, file systems, and data structures are located and transferred into a managed environment.

Digital forensics emerged from the law enforcement community in the 1980s as a method to identify, acquire, analyse and report upon digital information that constitute evidence of a legal investigation. To ensure that digital evidence is trustworthy and admissible, investigators must be able to establish that the information is authentic in what it purports to be and who created it. Failure to establish these requirements may result in the breakdown of the criminal investigation, incurring financial and reputation costs. In recent years, considerable investment has been made to establish digital forensics as a discipline and flexible software tools have been developed to analyse the increasingly large and diverse types of digital information produced in an automated, non-invasive manner.

The Forensic Investigation of Digital Objects (FIDO) project applied digital forensic methods to enhance the working practices of an academic archive. During the investigation, it recognised the substantial overlap between digital forensic methods and long-standing archival theory and digital curation approaches, developing guidance material for use by archivists and digital curators to help them to understand forensic methods. It went on to make a set of recommendations for work that should be performed within the digital curation community to enable broader uptake of digital forensic tools and techniques.

1.2. Aims and Objectives

To address the challenge of applying forensic processes and procedures developed by law enforcement to an academic archive environment, the FIDO project established three key objectives:

1. Evaluate the suitability of digital forensic principles and practices to enable HE archives to meet organisational commitments and legal requirements for maintaining digital records;

2. Assess the effectiveness of using the chosen digital forensic tools set to identify, acquire, and analyse digital information held on digital media and computer systems in an archival environment;

3. Seek to embed digital forensics tools & techniques into the working practices of the KCL Archives & Information Management (AIM);

It was recognized that the success of the project hinged upon how effectively the digital forensics principles, techniques and tools proposed for use in the project met the requirements of staff working with digital media and computer systems

1.3. Environmental Context

The Archives & Information Management (AIM) Service performs several key functions within King's College London, collaborating with internal and external parties to acquire material of archival value, store it within a managed environment, and make it available for use. The acquisition policy for AIM describes its collection remit as follows:

> *"The College Archives seeks to acquire, preserve and make available all material of long-term, evidential and research interest that forms part of the College's heritage. These archives range from institutional records created by the College (and associated institutions) in the course of day-to-day business to the private papers of academics and others engaged in pioneering research, as well as papers of organisations and individuals that reflect research strengths."*
> *Archives & Corporate Records Service (ACRS): Acquisition Policy* http://www.kcl.ac.uk/content/1/c6/06/81/50/ACRSacquisition.pdf

Traditionally, the work of the College's archives has focused upon the acquisition and management of physical records - paper, cassette tapes, video tapes, and other realia – created for business and research purpose. However, the growth of computing technology during the previous three decades has resulted in a large amount of born-digital material being produced, which are now being offered to the HE archive.

To equip the College Archives with the skills and expertise necessary to curate and preserve digital data in the long-term, the Centre for e-Research and Archives & Information Management Service collaborated on a JISC-funded project, titled PEKin (Preservation Exemplar at King's) to evaluate current practices, policies and procedures relating to information management and embed best practice. As an outcome of this work, the AIM Service was provided with a data management system, and a set of revised policies, procedures and guidelines to support the management process.

Although an OAIS compliant data management system was produced, it was recognised that there remained gaps in the service. Most notably, processes for acquiring, analysing and appraising digital records performed during the Pre-Ingest phase remained ad hoc and reflected practices developed for physical records. These processes were medium-agnostic, applicable to both physical and digital records, but did not address the distinct characteristics of digital material. Specific issues raised by archival staff and management include:

1. Significant effort must be made by the depositor in the initial stages to prepare the data for deposit, which may potentially act as a barrier for submission;

2. Acquisition processes emphasise the acquisition of the physical media only; processes for creating copies of digital data were undocumented and applied inconsistently;

3. Processes for analysing and appraising digital data were labour-intensive, requiring the allocation of a significant amount of time to process large collections; there were also concerns that valuable data may be overlooked by the archivist, due to the inconsistent use of different analysis methods

4. Advice provided by staff did not take into account the distinct requirements of managing digital material

The College Archives possess procedures and policies to assist staff to process data collections and transfer them into a digital archive for curation and preservation. However, these procedures were written with an assumption that the archive will be handling research data collections and digital business records, rather than Personal Digital Archives. It was

therefore seen as necessary that the characteristics of PDAs were understand and appropriate methods established.

1.4. Challenges posed by Personal Digital Archives

The Personal Digital Archive was recognised as a type of digital collection that is likely to be problematic to process by Archives staff. The term may be applied to any collection of digital items "*within an individual's control that have been stored and maintained by the individual*" (Cushing, 2010), but it is more commonly used to refer to digital collections that have been created in a non-business environment. It may comprise one or more types of digital storage device, each of which may contain data intended to fulfil a number of purposes.

A PDA is often created as a by-product of activities performed by the owner (e.g. a user purchases an Apple Macintosh to enable them to use photo editing software) and will "*grow organically*" over time as new content is added or removed (McKemmish 2005, p20). Studies on information retention practices by Kaye et al (2006) and Kirk and Sellen (2008) explore the motivation for performing information archiving within a work and home environment, identifying several reasons that digital material is held within a Personal Digital Archive: (1) to enable easy retrieval for access and use at a later date; (2) to build an unified body of work that reflect the person's career trajectory and provide evidence of achievements; (3) to enable access and use by others; (4) fear of losing irreplaceable information, and (5) as evidence of identity (Kaye et al, 2006) By contrast, a research data collections or business records, which may consist of a finite set of files intended to fulfil a distinct purpose.

The archival challenges posed by Personal Digital Archives stem from the method of construction and the form in which they are provided to a data archive. These may be understood as a set of stages through which an investigator must progress to the next, higher level (Figure 1). Once the investigator has resolved the hardware and software challenges, they will possess digital information of value to their investigation.

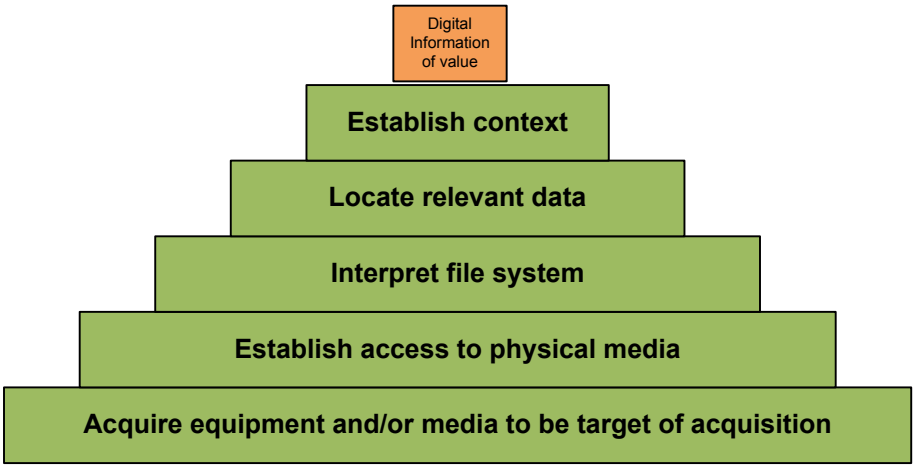

Figure 1: Data retrieval pyramid

Specific challenges that must be addressed at each stage:

1. *Acquire equipment and/or media to be the target of acquisition*
   A Personal Data Archive may encapsulate data stored on several storage devices[1] held in different locations. The challenge for the investigator is to determine the items that

[1]E.g. portable media such as 3.5-inch floppy disk, CD-ROM, as well as solid-state media installed in a laptop, digital camera, phone, and tablet devices.

should be acquired.

2. *Establish access to physical media*
   The digital media selected by the creator to store digital information will be affected by the options available at the time. Contemporary data may be stored on one or more of several devices, including hard disk (PATA, SATA, SCSI or USB), solid state media (Secure Digital, CompactFlash, MultiMediaCard), optical media (Blu-Ray, DVD, CD-ROM), and/or device-specific storage (e.g. solid state media embedded within a digital camera). In addition, consideration should be given to data stored on obsolete media formats that may be deposited (e.g. 3.5-inch, 5.25-inch, 3-inch floppy disks, ZIP100, ZIP250, ZIP750, Jaz 1GB). The challenge for the investigator is to establish an appropriate method of accessing disparate media types and applying consistent processing tools and workflows for their acquisition and analysis.

3. *Interpret file system*
   A file system acts as a method of storing data on digital media for later retrieval. Common file systems used by contemporary operating systems include NTFS and FAT32 for MS Windows, HFS and HFS+ for Apple MacOS, and ext2-4 in the Linux OS. The challenge for the investigator is to identify the file system in use and determine the most effective method of interpreting its structure[2].

4. *Locate relevant data*
   Large capacity digital media may contain thousands of files, obtained from different sources, including operating system files, software application executables and libraries, log records, internet browser cache, temp files, as well as user created data. The challenge for the investigator is to locate digital information of value within the digital 'haystack'.

5. *Establish context*
   Data may be organized and labelled according to the user's ad hoc needs and/or in accordance with file system requirements with little or no consideration that they would be examined by others at a later date. The challenge for the investigator is to establish the semantic context embedded within the organisation structure and ensure it is transferred into a managed environment.

To address each requirement, new processes and procedures must be developed for the archive to obtain digital information and transfer it into a managed environment in a manner that maintains all relevant data attributes, minimises the risk of accidental change, and is documented appropriately.

[2] E.g. mount the filesystem directly or perform file carving

2. Review of forensic Investigation Models

2.1. Introduction

A forensic Investigation is a process through which facts associated with an area of interest are located, examined, and interpreted. Facts gathered may be used to produce a new hypothesis or to support/disprove one that has previously been developed (Kent et al, 2006, 3-1).

Several forensic models have been created and published during the past twenty years that seek to frame the investigation process. These models build upon the principles of forensic science, information technology, and knowledge management, but differ in the composition of these elements, level of prescriptiveness, degree of detail, and terminology in use. One of the simplest forensic investigation models was published by the National Institute of Standards and Technology (NIST) in 2006. This describes a set of four high-level, implementation-independent processes – Acquisition, Examination, Analysis, and Reporting - that are commonly performed in a forensic investigation, as well as a conceptual model for understanding the interpretation process (Kent et al, 2006, 3-1). Forensic investigation models produced by other authors adopt a more granular, practice-based approach. The work of the First Digital Forensics Research Workshop (DFRWS) has been particularly influential in the field, providing a framework for structuring a forensic investigation (Palmer, 2001). In the initial phase, the investigator seeks to preserve the “crime scene”, the state of the computing environment in the form that it was found, and acquire a copy for analysis. They subsequently identified data of value to their investigation, interpreting the contextual role it performed, and record the result of their investigation. Each stage is described using practical examples developed with recognition of the techniques available at the time. Further work by Reith, Carr & Gunsch (2002), Carrier & Spafford (2003), Politt (2004), Ruibin et al (2005), and Agarwall & Gupta (2011) build upon the DFRWS model, revising it to take into account new forensic analysis techniques and other scenarios.

Although intended for use by law enforcement, each forensic model offers features that may be used to inform Digital Curators and Archivists on the methods and techniques that may be applied to analyse digital data. To gain a better understanding of the models, the project examined each in turn and synthesised the components onto a common framework. (Figure 1)

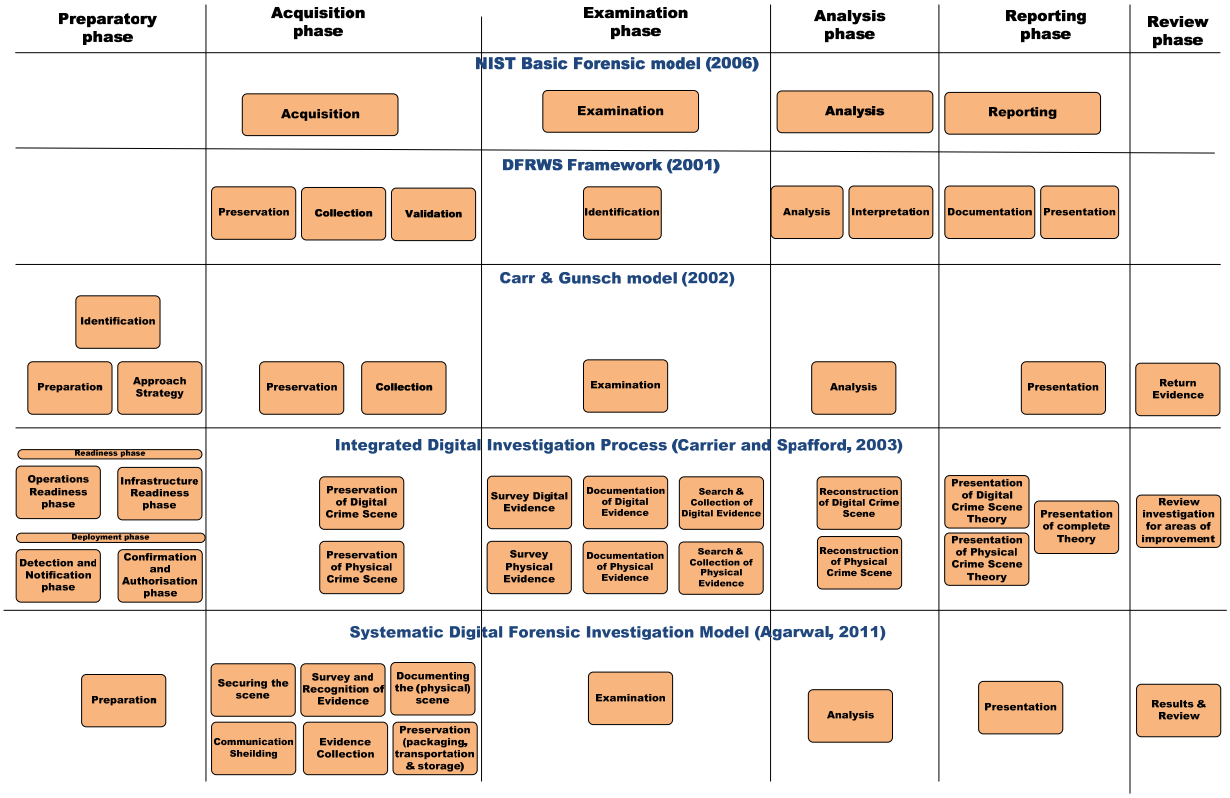

Figure 2: Aggregate forensic model

The synthesis revealed commonalities in the activities performed by forensic practitioners to analyse digital media. There were also a number of features distinct to one or two models that were considered worthy of wider application. The IDIP model (Carrier and Spafford, 2003) and SDFIM model (Agarwal & Gupta, 2011), for example, strengthen the relationship between digital and physical evidence, exploring how evidence collected in one medium may inform the investigation process used in the other medium – a workflow that may be helpful for archivists working with hybrid collections. Additionally, Carr & Gunsch (2002) provide a 'return evidence' activity that could be applied by an archivist wishing to return deposit media to its original owner. These models also establish much-needed feedback loops that allow Data Examination and Information Analysis activities to be repeated several times, as-and-when new questions are raised.

2.2. Aggregate Forensic Investigation model

By analysing the forensic models in combination and merging the functions, it is possible to identify six core stages of a forensic investigation, each of which incorporates a number of methods.

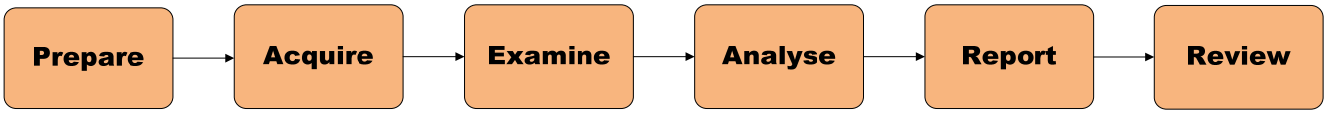

Figure 3: Aggregate forensic investigation model

1. *Prepare:* A set of activities necessary to recognise the incident, establish the environmental conditions in which the investigation is required, permissions that must be obtained, and select an appropriate strategy (including use of specific tools and techniques) to be applied (Agarwal, 2011; Carrier and Spafford, 2003,Carr and Gunsch, 2002).

2. *Acquire:* Data related to a specific event or topic of interest is identified, labelled, recorded, and collected. This stage will cover activities necessary to isolate, secure and preserve the state of physical and digital evidence (e.g. preventing people from using the digital device or allowing other electromagnetic devices to be used within an affected radius), followed by the creation of a digital copy for later examination.

3. *Examine:* Techniques are applied to perform an in-depth systematic examination of acquired data to identify and locate information of potential relevance to the investigation.

4. *Analyse:* Information contained within the extracted data is manually analysed by an investigator and evaluated for relevance in addressing questions raised during the investigation. New questions may be raised during this phase that requires the Examination activity to be repeated several times.

5. *Report:* The results of the investigation activity are documented and presented for consideration, on conclusion of the investigation. The report will include details of actions performed, knowledge gained, and steps that must/should be taken

6. *Review:* The experience of performing the investigation is reviewed to identify improvements that could be made to existing processes (Agarwal & Gupta, 2011; Carrier & Spafford, 2003) and action performed to store the evidence in an appropriate environment for later consultation and/or return to the owner (Carr and Gunsch, 2002).

The broad investigation model may be applied to the Pre-Ingest phase of an OAIS-compliant archive, formalizing the activities necessary to locate digital information and transfer it into a managed environment for curation and preservation. The work performed by the FIDO project focused upon the technological components of stages 1-3. Subsequent stages to appraise and report upon the investigation process were considered to be part of the normal operation of a digital archive and, as a result, out of scope.

3. Implementation

3.1. Setup of Digital Forensic environment

The first task for a digital archive is to build the forensic environment that will be used to acquire, examine, and analyse digital resources. A forensic 'lab' consists of a set of hardware devices and software tools that have been evaluated by an authorised party and selected for use in a computer based investigation (Ghirardin, 2009). The form that a forensic environment will take is influenced by a number of factors, including the type of media/equipment to be handled, volume of work, and number of staff, funding available, and characteristics of the physical space. In the case of an academic archive, a forensic lab is likely to be small, consisting of one or two computer systems and a number of peripherals. These components may be obtained using one of three different methods:

1. A ready-built forensic system tailored to the institution's needs is purchased from a commercial provider;
2. A bespoke system is built using off-the-shelf components;
3. A commercial system is purchased, which is subsequently extended and enhanced to fit the needs of the institution.

The FIDO project adopted the second approach to the setup of a forensic environment, utilising common hardware and free/open source software, where possible. The approach enabled the project team to develop experience with the use of forensic tools and resources, while minimising the initial investment costs[3]. This differs from the approach adopted by other academic institutions, such as Stanford University and the Bodleian Library, which have taken the third approach, purchasing a FRED system and extending it to fit specific needs (Kirschenbaum, Ovenden, & Redwine, 2010; Olson, 2010).

3.1.1. Forensic workstation hardware

A budget of £1300 was allocated to the project for the purchase of hardware and software necessary to build a single forensic workstation capable of accessioning personal digital archives. This was used to purchase four components:

1. *Personal Computer:* A standard Intel PC was purchased to serve as the forensic workstation for use by archival staff. The project purchased a low cost machine (Zoostorm dual-core Pentium tower PC) fitted with a 3.20GHz Intel Pentium dual core CPU, 2GB DDR3 RAM, Super multi DVD-RW optical drive, and 500GB hard drive for the purpose.

2. *Write Blocker:* A plug-through device fitted between the computer and storage device that monitors and blocks write operations to the media. Many write blocker variants exist, intended for use with one or more interfaces (USB, SCSI, SATA, Firewire). The purchase costs of a write blocker are high, with costs varying between £200-1000 at the time of investigation. The project selected a WiebeTech USB WriteBlocker, one of the cheapest write blocker devices on the market at the time.

3. *Floppy disk controller:* The Kryoflux USB-based floppy controller enables digital media, such as 5.25 and 3.5 inch floppy disks, to be acquired at the bit level. The disk image may be stored as a raw data stream or encoded in one of several platform/emulator-specific formats (e.g. IPF or ADF for Amiga disks, ST for Atari ST). The Kryoflux was

[3] The purchase cost of a Forensic Recovery of Evidence Device (FRED) system was around £4000/$6000. By comparison, the cost of a x64 system will vary between £200-1500, depending upon hardware specification.

purchased under a 'Small Academic Edition' licence, intended for institutions that have less than six employees or will process less than 100 disks per year. Additional features, such as an enclosure and hardware upgrade protection are omitted from this bundle.

4. *Hard drive enclosure:* A hard drive enclosure enables 3.5- 2.5 inch PATA/IDE and SATA hard drives to be connected via a USB interface. It is used to provide quick access to hard disks that have previously been housed within a depositor's computer, without the need to dismantle the forensic machine.

A total amount of £1273.86[4] was spent establishing a basic forensic workstation. The hardware serves as the basis for a low cost forensic workstation that addresses current needs for accessioning personal digital archives. However, it was recognised that additional investment will be required at a later date to accession digital collections on other computing devices and digital media types.

3.1.2. Forensic workstation software

A large component of the work performed by the FIDO project focused upon the identification, analysis, evaluation, and selection of forensic software tools for use in an archival environment. To assess the suitability of the forensic tools for use, the evaluation exercise was framed around two broad questions:

a. *Functionality:* Does the software tool provide the required functionality and conform to relevant standards within the designated environment?

b. *Usability:* Can the software tool be used by archive staff within an academic environment?

The need to consider functionality and usability may appear to be obvious when selecting software tools. However, these factors take upon increased importance when performing knowledge transfer, utilising expertise gathered in one domain to inform the development of another. Although the tools may be appropriate for use by law enforcement and the legal profession, presumptions made in their construction may make it difficult for them to be used by digital curators and archivists, who will approach an investigation with different questions in mind, and will apply a different knowledge base and expertise to the task. To evaluate these tools, software testing was framed around three activities:

1. Identify software tools that may be used for a specific activities within a forensic investigation
2. Develop experience of use of software tool to develop experience
3. Trial use of software tool with the archival team

The FIDO project found that many of the OSS digital forensic tools were unsuitable for use by archivists in their current form, requiring technical expertise to operate that many archival staff did not possess and produced output that requires further manipulation to be of use for allocated tasks (see 3.4.2). As a result, a combination of commercial and open source software was adopted for the forensic system.

[4] At the time of purchase, the PC was priced at £203.99 and a 19-inchTFT monitor was priced at £77.99 on the MISCO website, the WiebeTech USB WriteBlocker was purchased from an Amazon reseller for £179.97 + £4.59 shipping, the floppy controller was purchased from Kryoflux Products and Services Limited (UK) for 899 Euros (787.320 GBP at the time of purchase) under a Small Academic Edition licence, and the 'Max Value 3.5 inch USB Hard Drive Enclosure' was purchased from Amazon for £20. A 3.5-to-2.5 IDE connector cable and power adaptor was sourced from local supplies.

3.1.2.1. Testing environment

The choice of operating system selected for use on a forensic workstation has wide-ranging implications influencing the software tools that may be utilised, the training that must be provided, and the overall experience of use. Software developers (individuals or businesses which produce forensic software) commonly prioritise a specific platform utilised by their target community. Commercial developers, such as Access Data Group LLC, Guidance Software Inc, and Passmark Software produce software for use on Microsoft Windows, whereas software developers producing open source forensic software frequently prioritise Linux-based platforms (although efforts are made to make the software available on other operating systems).

In the initial phase, forensic software was tested using a combination of Linux Ubuntu[5], Microsoft Windows[6], and a number of Linux-based live CD distributions[7]. The live discs were found to be particularly useful, providing access to a number of pre-configured forensic applications that could be tested using sample data. Installed versions of Linux Ubuntu and Microsoft Windows operating systems were utilised for forensic tools that required additional configuration (e.g. Timescanner and Log2Timeline) or to tailor elements of the forensic software[8] (an approach adopted when testing Autopsy and OCFA)

The testing phase revealed several Linux based OSS forensic tools that may be utilised to acquire, examine, and analyse digital data within a curatorial environment (see 3.4). Following identification, the project manager worked with archival staff to trial these tools and evaluate their usability. It quickly became evident that the host operating system itself was a barrier to uptake, being considered confusing and difficult to use by staff more familiar with a Microsoft Windows environment. As a result of this discovery, the project chose to standardise upon the Microsoft Windows platform for the production system, prioritising forensic tools that could be executed within this environment.

3.1.2.2. Production Environment

The forensic workstation provided to the archives department was installed with Microsoft Windows as the primary operating system. Several applications were installed on the system, including OSForensics (for use when analysing digital media and locating data of value), LibreOffice, the Kryoflux analysis tools and drivers, McAfee virus checkers (institutional licence), and VirtualBox. An Ubuntu virtual machine was created for use, to cover the eventuality that Linux-based forensic tools may be needed at a later date.

[5] Ubuntu 10.10 (Maverick Meerkat) was used for the first few months of the project, later upgraded to Ubuntu 11.04 (Natty Narwhal)

[6] Microsoft XP was used as a basis in the initial phase of the project. However, this was replaced with Microsoft Windows 7 following the transfer of the machine to the archives.

[7] A Live Disc is an item of digital media (CD, DVD) that contains an operating system that may be booted and used without installation or making changes to other operating systems installed on the machine. Live discs examined during the testing phase include BackTrack (http://www.backtrack-linux.org), CAINE (http://caine-live.net), DEFT Linux (http://www.deftlinux.net), and PlainSight (http://www.plainsight.info).

[8] Live discs remove the need to install software applications. However, configuration changes and activities performed will be lost when the machine is rebooted. An operating system running in memory will often be slower than one installed on disc, which will have a negative impact upon performance speed of software applications.

3.2. Phase 1: Preparation

3.2.1. Introduction

The Preparation phase of an archival-forensic investigation encapsulates a set of activities necessary to setup the investigation, performing actions such as establishing the deposit scenario, and selecting an appropriate strategy to use. The preparatory activities for obtaining a PDA are broadly similar to the process currently applied to negotiate for use of research data from funded researchers, but will incorporate additional decision points specific to the forensic investigation.

3.2.2. Establish scenarios for the application of forensic tools

The project identified five scenarios in which AIM would encounter complex digital archives that require the application of forensic tools.

1. **Scenario 1: Staff retirement**
   A long-serving staff member retires, leaving behind the research that they accumulated over the years.

   Example: A long-serving staff member retires, leaving behind their research. Data is held on their desktop computer, a number of external drives, the department network drive, and email server. In addition, their office contains several filing cabinets of papers. Data is also held on network drives and the email server. The archives service wish to capture a copy of the data held in these different locations before they are lost.

2. **Scenario 2: Staff move to other institutions**
   A staff member moves to another job, leaving behind their computer for repurposing. Their manager wishes to confirm that all work-related data has been transferred to the network drive for reference by others, which prompts them to contact the archives for advice.

   Example: A colleague of a recently departed senior manager gets in touch: she is concerned that there may be useful information, and perhaps important documents, held only within his email system, which will be deleted in a few weeks.

   Example: A long-serving KCL employee has recently left for a new job, leaving behind their work PC for repurposing. It is recognised that they amassed a considerable amount of digital information related to the operation of the department, some of which they created, while other was obtained from elsewhere. Although much of the data is held on network store, there is concern that unique information may be held on the internal hard disk and within the email archive of the individual.

3. **Scenario 3: Department re-structuring**
   One or more departments are re-structured to fit within the changed scope of the institution and employees are transferred into different groups.

   Example: A department is transferred from the Information Services group into an academic school and administrative staff are reallocated. The archives service wish to capture a record of work performed by the department before it is re-organised.

4. **Scenario 4: A retired academic wishes to donate their personal archive**
   A retired professor wishes to dispose of work produced over their lifetime and contacts the archives service to ask if they would be interested in acquiring the material.

   Example: An eminent retired KCL professor is planning to emigrate to New Zealand to be near her grandchildren and wishes to deposit her lifetime's work with the archives. The personal archive contains several filing cabinets of papers and a number of desktop PCs that were bought at different times over the previous two decades. Each machine contains digital information related to the professor's research and teaching activities, as well as personal information, such as personal correspondence and financial management documents. The computers have been used by other family members in the past and may contain other information created/obtained by them

5. **Scenario 5: Deposit of an academic's hybrid archive following their death**
   The executor or family member of a notable academic wishes to deposit their personal research collection within an appropriate archive. The personal archive contains several filing cabinets of papers and a number of desktop PCs that were bought at different times over the previous two decades. Each machine contains digital information related to the professor's research and teaching activities, as well as personal information, such as personal correspondence and financial management documents. The computers have been used by other family members in the past and may contain information created/obtained by them.

   Example: The literary executors of a military historian wishes to deposit his personal research collection with an appropriate archive. The historian died twelve months previous, leaving four shoe boxes of floppy disks, a PC of unknown manufacture, plus extensive printed proofs, off-prints of journal articles and assorted press cuttings. The executors are also in negotiation with a publisher with regard to the historian's last, unpublished work.

   Example: The daughter of a retired Lieutenant General believes that the laptop purchased by her father following his retirement contains material of potential value. She believes that he used the machine to write his memoirs a few years ago and may have exchanged his emails with former colleagues regarding relating to his service with NATO forces in Kosovo, 1999, and subsequent war crimes trials, some of which are still on-going. She does not know how to access the machine itself, does not possess any passwords, and has not seen the material itself. However, she is anxious to find out if there is anything of significance and to see it safely preserved.

3.2.3. Initiate Contact

The initiate contact activity performed by a digital archive is likely to be similar to the process currently applied when negotiating with funded researchers to obtain research data. Contact would be initiated by the archive (e.g. making an enquiry regarding the existence of specific work) or the depositor (e.g. the retiree, their family or estate) and basic information provided on the type and amount of data contained within the digital collection.

3.2.4. Establish agreement

In the second phase, the archive and depositor negotiate the conditions of deposit and the transfer method. Negotiation must take into account the additional complexities introduced by the forensic process. Key issues that an archive may wish to consider during the initial negotiation stage include the level of analysis authorised by the depositor (which will have

implications for the forensic examination method to be applied) and the type of material that they are willing to make available for access and use.

3.2.5. Determine transport method

Finally, the method of transporting the digital archive to the digital archive must be taken into account. Digital media may be personally delivered by the depositor or transported using the postal service. In other circumstances, the depositor may wish to retain the original digital media or the media may be considered too fragile to transport, necessitating the need for the archivist to visit the depositor and obtain a copy of the data. In these circumstances, the investigator will need to provide relevant tools (e.g. external USB hard disk, boot disc).

3.3. Phase 2: Acquisition

3.3.1. Introduction

Acquisition is commonly defined as the act of obtaining possession or control of an object. For an archive handling physical records, acquisition will involve the negotiation and transport of a box of papers or other items, whereas a forensic investigator will be more concerned with the process of preserving a crime scene. In both cases, the challenge is to obtain the information in a manner that maintains its context, while avoiding actions that will result in its modification, corruption, or destruction.

Traditionally, a Digital Curator/Archivist wishing to obtain data for inclusion within a digital archive will perform a manual examination of files held on digital media, appraise the content of each, and transfer a subset of relevance to their investigation. Although effective in many circumstances, the process can be time-consuming to perform for high capacity media and does not guarantee that all relevant digital data will be located, particularly if time constraints are placed upon the investigation (e.g. they have 1 hour to examine the disk before having to return it to the creator). There is also a risk that the examination process will result in data being changed – the creation/modification/access date may be updated when accessed and content may be converted to a later format version when rendered using a specific software application.

To enable the state of digital media to be acquired in a complete, unaltered form, digital forensic practice within law enforcement has focused upon the capture of a mass storage device or computer memory at the bit level for storage within a disk image[9].

Several reasons may be identified for choosing to capture a complete image of physical media, as opposed to simply copying the files:

1 The investigator is able to state that there is a *reasonable* probability that they have acquired all data held on the drive, including hidden/deleted data that is invisible to the end user[10];

2 The investigator is less dependent upon the continued availability of the source media, which may potentially fail or begin to corrupt data through continued use;

3 The creation of a secondary copy of the data that has been replicated to multiple devices enables the investigator to analyse the data, without the risk that the analysis may cause inadvertent, unrecoverable change to the only copy;

4 The acquisition of a disk image enables an investigator to perform their analysis away from the original creation environment;

5 The analysis of a disk image within a forensic environment enables the use of methods and tools that are not available or feasible in the original creating environment;

Disk imaging is commonly performed in the IT industry as a method of data backup, in order to enable quick restoration in the event of media failure or data transfer, while the forensic

[9] A disk image is a set of one or more files that, in combination, contain the content and structure of a mass storage device

[10] An investigator can only claim that it is reasonably probable that they have captured all relevant data, due to the different methods available to hide data.

community use it as a basis for analysis. The case for disk imaging in the digital archives community is less established, but has the potential to incorporate each of the outlined factors.

3.3.2. Acquisition and its implication for later stage

The acquisition stage performs a key role within the forensic workflow, providing the digital data that will be the target of analysis. Decisions made during the stage will have wide-ranging implications upon subsequent stages. To ensure that the data captured during this stage is fit for purpose, it is helpful to consider the role of the disk image in enabling the archive to meet its objectives. Specific questions related to the performance of forensic imaging and creation of disk images include:

1. What imaging format should be used to encode and store disk images?
2. What information should be created to support the disk image?
3. What software tools should be used to acquire disk images?

The issues raised in these questions are inter-related: the answer to one question will have an impact upon the approach adopted to address other questions.

3.3.3. Selection of disk image format

First, the project sought to determine the encoding format in which the disk image would be stored. A large number of encoding formats exist capable of storing media within a file. Many have been designed for the storage of specific media types (e.g. ADF and IPF for Amiga floppy disks), whereas other types of disk image are intended to be applicable to a broad range of media, such as the “raw” DD format.

To select an appropriate encoding format, it is necessary to determine a set of evaluation criteria and apply it to the disk imaging formats available. An investigation of the topic found that relatively little work had been performed in this area, the only notable example being that produced by Garfinkel et al (2006), which hightlights the importance of extensibility, licence status, compression support, and data location as factors that require consideration. To determine the image format to use within the KCL Archives, the project drew upon similar work performed by Todd (2009) on the topic of preservation file formats and the aforementioned paper by Garfinkel et al, to identify eight factors that should be taken into account when selecting a disk image format:

No	Name	Description
1	Adoption	Extent to which the format is in widespread use within the forensic community and elsewhere
2	Software independence	Extent to which the format is independent of specific hardware and software
3	Disclosure	Extent to which the file format specification is in the public domain
4	Metadata support	Extent to which descriptive information is supported within the format
5	Licence status	Licence associated with the format, which may affect the degree of disclosure and adoption.
6	Level of fixity analysis supported	The level at which fixity information may be recorded about and within the object
7	Support for split files	The ability to split a large disk image into smaller sections of an arbitrary size for storage on disc or other media
8	Compression support:	The ability to compress the data image to reduce

		storage space. Compression support is useful but it not considered mandatory.

The evaluation criteria was applied to three disk imaging formats recognised as being in common use within the forensic community and well supported by disk imaging software:

1. *Raw/DD:* A type of disk image created by the DD Unix command and other software tools. Raw images contain a bit copy of a source device, without any attempt made to identify or interpret the filesystem or files held on the disk[11].

2. *Advanced Forensics Format (AFF):* An extensible open format developed by Simson Garfinkel and Basis Technology. The AFF format is comprised of two layers: a Data-Storage layer that contains the disk information and a Disk Representation layer that may be used to associate metadata with specific segments of the disk image.

3. *Expert Witness Forensics (EWF):* A proprietary disk image format used by Guidance Software in the EnCase software tool[12]. It is widely considered to be the de facto standard for forensic disk images, due to the popularity of EnCase within the law enforcement community. It is also supported by a number of open source tools, via the LibEWF library.

On the basis of the evaluation, the FIDO project designated AFF as the preferred format in which to acquire disk images, due to the degree of disclosure and metadata extensibility.

3.3.4. Selection of forensic imaging tool

A broad range of forensic imaging tools exist, each of which is capable of acquiring bit-level data and write it to an image file. The project team began by reading discussion papers, articles, and case studies to identify software applications that were considered to be helpful within the forensic community. This was accompanied by hands-on testing to gain practical experience of forensic imaging tools available.

Following initial experimentation, the project sought to establish criteria for evaluating each imaging tool and select one (or more) that would be suitable for use by archivists. It was recognised that the NIST's forensic imaging tool assessment criteria may be used as a basis for evaluating the functionality provided by forensic software. However, there was a need for additional criteria to establish their suitability for use by archivists who, in most cases, had limited technical knowledge. Specifically, four additional requirements were identified:

1. *Operable within a 'live environment':* Forensic practice commonly focuses upon use of third party boot media to launch a basic operating environment on the source machine, to minimise the risk of accidental data change. Therefore, it is essential that forensic imaging software is able to be configured and used within the constraints of an operating system that runs in memory (e.g. a FreeDOS command line prompt, Linux LiveCD).

2. *Metadata support:* Facility to manually/automatically capture descriptive information about the disk at the point of acquisition and/or provenance information about the acquisition process

3. *Image format support:* Ability to store the acquired disk in the preferred forensic image format

[11] As a result, it is a misnomer to describe Raw as a disk imaging format.
[12] See http://www.forensicswiki.org/wiki/Encase_image_file_format for further information

4. *Usability:* Ability to configure and apply the software tool without knowledge of command line tools or specialist knowledge

Several disk imaging tools were trialled, including DD, Dc3dd, dcfldd, FTK Imager, Guymager, Kryoflux Imager, and OSFClone. The latter four proved to be easiest to use, providing user interfaces and user feedback that could be understood by archival staff. For the capture of hard disks and other high capacity storage media, OSFClone was selected for use: it contains an basic Linux operating system and integrated imaging software that can be launched from a CD-ROM or USB device; and can be easily configured to acquire an image and write it to an NTFS formatted USB disk. Kryoflux Imager (http://www.kryoflux.com/) was selected for use for the capture of floppy disks, due to the associated hardware's ability to access media types that are incompatible with the Intel PC's floppy disk controller.

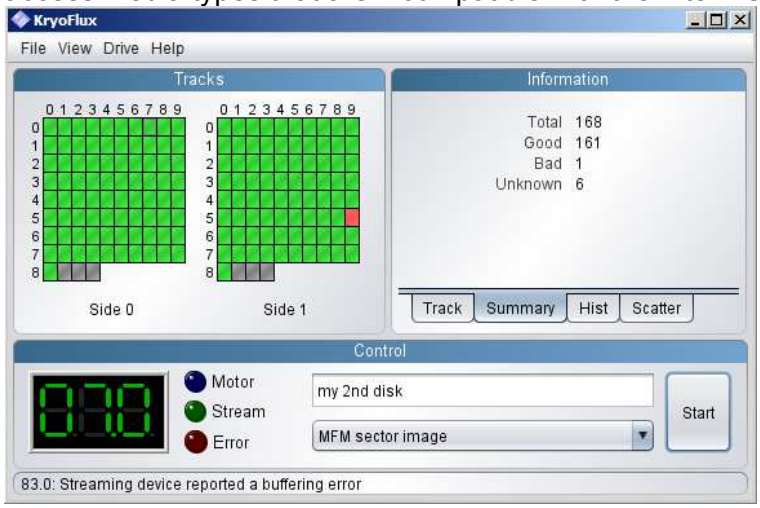

Figure 4: Kryoflux Imager screenshot

3.3.5. Documenting the disk image

The technical composition of each disk image was documented using Fiwalk (http://afflib.org/software/fiwalk), an open source file system walking tool developed by Simson L. Farfinkel that uses the Sleuth Kit API to interpret disk images. Fiwalk records the XML data structure of a disk image, including information on partition and data files held in each sector. The FiWalk metadata format outputs four types of information:

1. Provenance metadata on the software used to generate the XML output, including the version number, build environment, libraries used, and compilation date.

2. Disk-level technical metadata, describing the number of sectors, sector size, etc.

3. Partition-level technical metadata on each partition contained in the disk image. Relevant information will include partition offset (from the start of the device), file system in use, and other pertinent information.

4. File-level technical metadata on each active and inactive files held on the disk, including information such as filename, file size, creation/access/modification date (dependent upon OS), and disk sector location (Figure 4).

```
<fileobject>
  <filename>MININT/system32/REC_EXIT.GIF</filename>
  <partition>1</partition>
  <id>3098</id>
  <name_type>r</name_type>
  <filesize>668</filesize>
  <alloc>1</alloc>
  <used>1</used>
  <inode>64001644</inode>
  <meta_type>1</meta_type>
  <mode>511</mode>
  <nlink>1</nlink>
  <uid>0</uid>
  <gid>0</gid>
  <mtime>10505784442</mtime>
  <atime>1127170800</atime>
  <crtime>1127210390</crtime>
  <libmagic>GIF image data, version 89a, 40 x 88</libmagic>
  <byte_runs>
   <run file_offset='0' fs_offset='2073636864' img_offset='2073636864' len='668'/>
  </byte_runs>
  <hashdigest type='md5'>014e7737bd4eda3111b0bff53a4d23e2</hashdigest>
  <hashdigest type='sha1'>1f4702a14e3717dce8e28ae1e0797d24d69c21e1</hashdigest>
</fileobject>
```

Figure 5: File-level metadata extracted by Sleuthkit and formatted by fiwalk

The disk inventory held on disk was recognised as useful for an investigator, providing insight into its structure and content. However, in its current form, the information is difficult to analyse and interpret. Future work should focus upon the development of tools capable of interpreting the FIWalk structure and presenting its content in a visual form.

3.3.6. Bringing it together: Developing an acquisition workflow

The final component focussed upon the creation of a set of procedures, describing how digital media may be acquired using the selected forensic tools and formats. Guidance was developed to address five scenarios:

1. *Forensic imaging of an internal hard disk installed in a x86/x64 computer that offers USB connectivity*
   A boot disk (floppy disk, CD-ROM) is used to launch the owner's machine. Forensic imaging software is used to capture a bit copy of the hard disk and write it to an external USB hard disk.

2. *Forensic imaging of a hard disk removed from the owner's computer:*
   A 2.5" or 3.5" ATA/IDE or SATA hard disk is removed from the owner's computer and installed within a USB drive enclosure. Forensic imaging software is used on the Ingest machine to acquire a bit copy of the hard disk and write it to storage.

3. *Forensic imaging of 3.5 or 5.25 inch floppy disk media:*
   One or more 3.5/5.25 inch floppy disks are accessed using the Kryoflux floppy disk drive connected to the Ingest machine and a bit copy of each is created.

4. *Forensic imaging or copying of data held on solid state media*
   One or more solid state devices are connected to the Forensic Ingest machine (via USB or appropriate media reader) and copied or forensically imaged.

5. *Copy data held on digital media*
   In the event that forensic imaging is impractical or unnecessary to perform, Archival staff may copy data held on digital media to other media for examination and analysis.

To enable Archival staff to determine the most effective method of acquiring digital media, a decision model (Figure 5) was produced to guide the process.

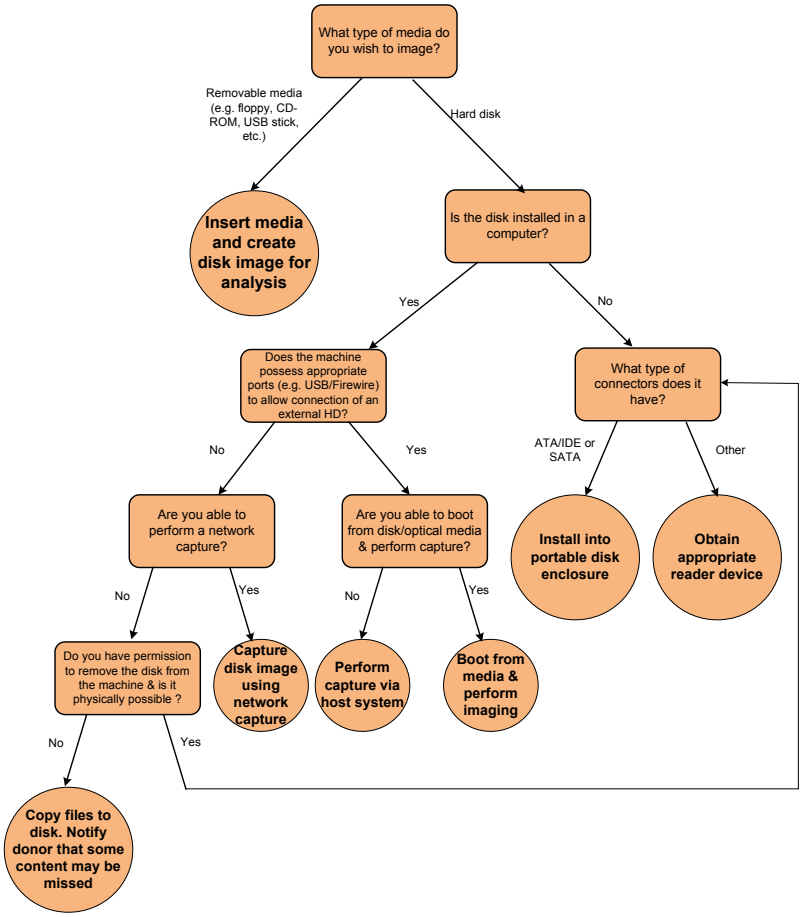

Figure 6: Decision tree for deciding most effective method of connecting digital media for imaging

3.3.7. Post-project work

It is recognised that many types of digital media and computer system are not covered in the aforementioned scenarios. New processes must be developed to acquire disk images from devices such as tablets, phones, Apple Macs, and various types of legacy hardware[13], among others.

[13] A researcher working in the 1980s and 1990s may, for example, have stored their research on 3-inch floppy disk, a Iomega ZIP 100/250/750 cartridge, Iomega 1GB Jazz disk, or an obsolete hard disk or digital tape format.

3.4. Phase 3: Examination

3.4.1. Introduction

During the Examination phase, an in-depth, systematic analysis is performed on the acquired data to identify and locate information that matches one or more search parameters. The objectives of a Digital Curator, Archivist, and Forensic Investigator will be broadly similar during the stage – each will wish to locate digital information of relevance to their investigation. Differences begin to emerge when the type of information to be located is taken into account. An Archivist or Digital Curator is likely to be interested in user-created data, which represents the intellectual output of the author, whereas a law enforcement officer may take a holistic approach, analyzing activities performed upon the machine.

Digital media frequently contain thousands of files, including executables and libraries associated with the installed operating system and software applications, data files that document the use of the machine[14], and user-created data, all of which must be analysed. The application of custom organisational structures and labels by the user may also result in files being held in unexpected locations. Traditionally, a Digital Archivist wishing to locate digital data of relevance to their investigation would apply a number of manual search activities. For example, a file search may be performed to locate files that possess a specific filename (e.g. *report*, *photograph*), are encoded in designated formats (e.g. PDF, Doc, JPG), or contain certain key words (e.g. “war studies”, “Kuwait”, etc.). Although effective when handling small digital collections, the process can be time-consuming to perform upon large collections and is not guaranteed to provide every file of relevance.

To enable law enforcement to locate relevant information on digital media, several forensic methods have been developed, which may be applied by Digital Archivists. These include the following:

- *Identifying content according to its source:* Hash filtering builds upon techniques commonly used to monitor bit integrity – a hash sum (e.g. MD5, SHA-1) is generated for one or more files held on disk and compared to a dataset of hash sums that have originated from a known third-party source (e.g. files distributed with Microsoft Windows, Adobe Photoshop, or other software packages). Data classification may be used as a basis to identify data files obtained from a specific source (or alternatively, an unknown source) that require further investigation.

- *Identify content based upon its timestamp:* Digital timestamps held on the digital media may be analysed to gain insight into the activities performed by a user and the files created, modified and accessed over a specific time period.

- *Identifying files according to its location:* Many types of digital media store data in locations that are inaccessible when examined using common disk navigation tools. Examples include hidden partitions that have been rendered inaccessible to the host OS, and Unallocated/Inactive Space that holds data deleted by the user which has not been overwritten. By utilising methods such as file system undelete and file carving, an investigator may be able to recover data for use in the investigation.

The forensic community has made significant investment into the creation of proof-of-concept and production-ready software tools capable of automating the examination and identification process. Although it is feasible to apply each of the methods in turn, this may be time-consuming to perform upon a large disk. The investigator must decide upon the

[14] For example, log records, internet browser cache, and temp files

examination methods that should/can be used and the order in which they will be applied (e.g. the archive is allowed to analyse files available on disk, but is not authorised to recover deleted data or analyse user logs or the browser cache).

3.4.2. Analysis of Case Management tools

Forensic practitioners working with OSS tools utilise several methods to access digital media/disk images and examine data files contained within. A wide-ranging investigation of forensic tools was performed to identify candidates capable of analysing a disk's encoding structure and locating relevant data. The project team began by reading discussion papers, articles, and case studies to identify software applications that were considered to be helpful by the forensic community. The analysis revealed the existence of several Case Management and other forensic framework tools published using an open source licence or made available for free that were capable of performing the stated examination methods within a single, integrated environment. Hands-on testing was performed on several tools to build an understanding of the capabilities and limitations of each.

- *Autopsy Forensic Browser (v2):* The Autopsy Forensic Browser (http://www.sleuthkit.org/autopsy/) provides a browser-based interface to the Sleuthkit software compendium, enabling analysis of active and deleted data and creation of timelines, among other features. However, the default UI is extremely limited, lacking support for file requestors/dialog boxes[15] and visualisation[16]. Forensic hashing (to distinguish between 'good' and 'bad' files) is supported through a third-party add-on, but is difficult to use from a workflow perspective, requiring the user to examine external text files.

- *Autopsy Forensic Browser (v3):* The third version of Autopsy is a Java-based reimplementation of the tool that may be executed locally within a desktop environment. The project team was not able to perform a detailed analysis of the tool (it was published in August 2011, one month after the project's conclusion). Early testing found that it was unstable when handling large (e.g. 10-50GB) disk images, but its modular design and usable interface represented a promising start.

- *Digital Forensic Framework:* DFF (http://www.digital-forensic.org/) is an open source, cross platform forensic tool for analysing digital media and disk images. The user interface is built using the Qt framework, enabling it to appear as a native application irrespective of the operating system (Mac, MS Windows, Linux) on which it is executed. DFF supports file browse & search, file carving, and timeline visualisation functionality, but lacked support for forensic hashing at the time of testing.

- *OSForensics:* OSForensics (http://www.osforensics.com/) is a commercial case management tool developed by Passmark Software. During the project's funding period, OSForensic was published as a time-limited open beta. Subsequently, the full version was made available for-cost and a cut-down version was released for free.

- *Open Computer Forensics Architecture:* OCFA is an open-source framework created by the Dutch national police (http://sourceforge.net/apps/trac/ocfa/wiki) for use in analysing digital media. Similar to Autopsy v2, OCFA's functionality is provided through integration with a number of open source tools, (including The Sleuth Kit, Scalpel, Photorec[17], libmagic, and others), which are configured using a browser-based front-end.

[15] The user must copy the full path+filename of the disk image into the appropriate text box
[16] The user must export the text-based timeline and utilise a suitable third party tool
[17] See http://www.cgsecurity.org/wiki/PhotoRec and http://www.cgsecurity.org/wiki/File_Formats_Recovered_By_PhotoRec

- *PTK Forensics:* PTK (http://ptk.dflabs.com/) is a commercial tool developed by DFLabs that builds upon the Sleuthkit tool compendium. The full version supports features such as forensic hashing, timeline generation, data carving, file signature analysis, and file extension mismatch checking. However, many of these features are unavailable in the free, non-commercial version. The user interface was found to be easier to use by non-technical users in comparison to Autopsy and OCPA.

The project also tested forensic tools provided on various forensic live CDs, such as CAINE and DEFT Linux.

During the third stage of testing, the project manager worked with archival staff to trial each tool and determine if it was useful and usable. It quickly became evident that many of the open source forensic tools were unsuitable for use by archivists in their current form, utilising technological terms that they did not fully understand and required technical expertise that many did not possess. The browser-based interfaces of Autopsy Forensic Browser v2 and Open Computer Forensic Architecture (OCFA), for example, were considered to be difficult to configure and use, providing inadequate feedback upon the analysis process. Digital Forensic Framework (DFF) and PTK Forensics were considered to be more usable, addressing some of their needs (user friendly graphical interface and flexible search interface). However, the archival staff did not believe that these tools offered sufficient benefits to justify their use[18], preferring to use a Windows Explorer interface to perform their analysis.

3.4.3. Methodological implications for the project

The archivists' reaction to the open source forensic tools forced the project team to re-evaluate the methodology adopted by the project. Although the project's aim was to evaluate open source forensic software and introduce relevant tools into the archival environment, the latter was no longer considered to be feasible, unless resources were allocated to perform additional development work. To address the issue, the project adopted a dual approach:

a. To ensure that the AIM Service were provided with a workable forensic solution, the team broadened its investigation to explore the use of commercial forensic software that could be executed within a Microsoft Windows environment.

b. To meet the strategic objectives of the project, the team would continue to explore the use of open source and free software tools that operated in a Linux and Windows environment.

As a result of the former, the project selected OSFClone and OSForensics as the preferred forensic tools for use within the Archives service. These tools provide graphical interfaces and support documentation that were considered to be simple to understand by archival staff. The use of these tools, describing the actions that may be performed to acquire, analyse and examine a disk image are documented in the Ingest Handbook for Digital Media (Knight, 2011d). As an output of the latter approach, the project was able to identify a number of powerful open source forensic tools that may be enhanced and extended through future work to better serve the needs of the digital curation and digital archive communities. These tools are documented within this report, as well as accompanying papers produced during the life of the project.

---

[18] Future enhancements to DFF and Autopsy v3 may provide further benefits, offering the functionality integration provided by commercial tools

3.4.4. Forensic hashing to identify content according to its source

Forensic hashing may be performed using a number of open source forensic tools[19], providing the investigator with information on the (potential) origin of files held on disk. The FIDO project built a forensic hashing workflow using the following tools:

- *The SleuthKit (TSK):* a compendium of open source forensic tools and scripts developed by Brian Carrier
- *Unix File:* A characterization tool that uses the magic number contained in header information to identify known file formats.
- The National Software Reference Library (NSRL) dataset

A Perl script called 'Sorter', included with The SleuthKit, was applied to classify files held on disk (including those located in unallocated space) as known/unknown. A small amount of configuration is required to set up the forensic hashing workflow – variables in the Perl script must be updated to refer to Linux/Windows tools and the NSRL dataset must be converted into a tab-delimited format that can be interpreted by TSK[20]. However, these tasks do not require extensive Perl or scripting knowledge.

Forensic hashing may be performed at the media level or file level. The application of media level hashing is useful if the investigator wishes to perform a holistic analysis of the disk, but can take some time to perform[21]. File level hashing is effective for establishing the provenance of a small number of files, but is dependent upon the investigator selecting a subset of files and initiating the hash lookup. Sorter is limited to media-level analysis only, which imposes restrictions upon the investigation workflow that may be applied. By contrast, commercial tools, such as FTK and OSForensics support file level hash lookup, making them more effective for use in analysing a small number of files within a larger collection.

On completion of the media analysis, Sorter will output a set of HTML documents, indicating the files those listed in the NSRL dataset and those that do not. For an Archivist wishing to locate user-created data, the latter is likely to be of interest, potentially containing data unique to the target machine (Figure 7).

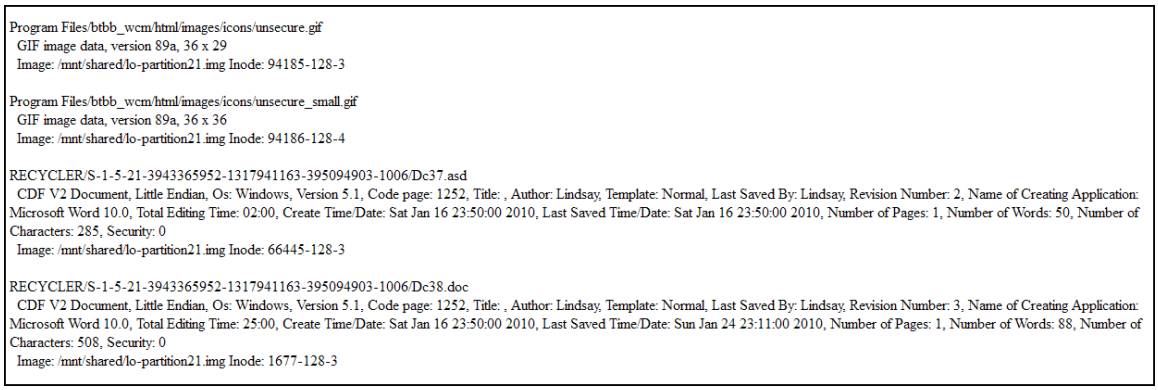

Figure 7: Sorter output listing unknown files

[19] Autopsy 2 supports hash filtering though integration of a third party plugin. Autopsy has been rewritten in Java for version 3 and, at the time of writing, does not support hash filtering.
[20] NIST currently make the NSRL dataset available as four ISO, each of which contains a set of comma-separated text files identifying Known Good and Known Bad files. The set of text files must be merged into a single (extremely large) file and subsequently converted into a tab-delimited format (using HFind) that may be interpreted by Sorter.
[21] For large disks, the forensic hashing process can take several hours, or possibly days, to perform.

Forensic hashing provides a useful starting point to identify files held on disk that may require curation and preservation. However, the lack of support for file level analysis in open source forensic tools is a significant limitation that restricts the investigation workflow that may be applied. The time required to process an entire disk image is likely to be unacceptable to an investigator wishing to establish the origins of a small number of files. To enable greater uptake of forensic hashing methods, further work is required to implement file-level hashing as part of case management tools, such as Autopsy v3 or Digital Forensic Framework.

3.4.5. Timeline analysis to identify content according to its timestamp

A second discovery method that may be used to locate relevant data is to perform a temporal analysis. In recent years the forensic community has focused upon the development of Super Timelines, which bring together temporal information from multiple sources - Windows Registry entries, Internet browser history and cache files, email exchanges, data files, as well metadata embedded within files – for analysis. Development work related to the creation of Super Timelines is, arguably, at a more advanced stage within the open source forensics community. The Perl-based TimeScanner and Log2Timeline tools have been particularly influential in the area, providing a common framework to develop modules for parsing individual log formats and outputting the concatenated information to a standardized format[22] for further analysis. These tools provide input modules for processing several common applications in succession[23] and classify temporal information into common fields. By comparison, commercial case management tools, such as FTK and OSForensics provide analysis functionality for individual components (Windows registry, browser history, and cache files), but do not currently provide a holistic representation of this information.

The setup of TimeScanner is likely to (ironically) take a significant amount of time, requiring the installation and configuration of a number of tools[24]. However, once configured and directed toward a mounted drive (TimeScanner cannot process image files), it will parse temporal information held in many different locations and formats[25], normalise the information[26], and output it to an open, structured format (TSK MacTime, SIMILE XML, BeeDocs, CSV, tab-delimited) for analysis within an appropriate software package.

3.4.6. Recovering content held in inaccessible locations on disk

Data recovery is a third discovery method that may be applied, for the purpose of locating 'lost' data that the user has removed or 'moved[27]' to another device. A file that is deleted by the user will not be immediately removed from disk. Instead the operating system re-classifies the sectors on which it is located from 'allocated' to 'unallocated'. Data held on these disk sectors may be recovered for a time until the sectors are overwritten with new data[28].

[22] Supported formats include TSK MacTime, SIMILE XML, BeeDocs, CSV, and tab-delimited

[23] A full list of Input Modules is provided at http://log2timeline.net/

[24] The following packages are required: The SleuthKit (TSK) command tools (FLS, specifically), Log2Timeline, MAKE, RegTime, and a large number of Perl modules.

[25] Examples include Internet browser history and cache files, email mail folders, log files, databases, as well as embedded metadata and file attributes.

[26] Each temporal event possesses 17 elements (some of which may be unpopulated). Date and time information extracted from different sources is normalised to a standard format. However, information in other metadata elements remains as-is, resulting in many different terms being used to describe the same event type (e.g. createDate, MediaCreateDate).

[27] A file move is comprised of a copy and delete operation. As a result, data transferred to another disk will remain on the source media in unallocated space.

[28] The time period that information will be available is subject to a number of factors, including the amount of space available for use, size of the data to be recovered, and disk usage. For large files that occupied multiple

Data recovery techniques utilise knowledge of an operating system to retrieve data that is not available through the file system. Several forensic techniques exist that may be used to recover difficult-to-access or deleted information:

1 *File system undelete:* Data files that have been deleted (e.g. via the Windows Recycle Bin) may be recovered using 'undelete' tools. Undelete tools use a file system pointer contained in the Directory entry to identify and recover files. However, they are only effective for a short time when the pointer continues to exist and cannot be used to recover data that has been overwritten in part or is located on a corrupted file system.

2 *File Carving:* File Carving[29] refers to a process by which raw data held on media is analysed and patterns sought in its structure that indicate the presence of specific content types. If a recognizable data structure is identified, such as the header and footer of a JPEG image, the data file is "carved" for analysis (ForensicWiki, nd). File carving may be used to identify and recover information fragments within a partially deleted file or recover data from a file system that has been corrupted by mechanical failure or virus attack.

3 *Text extraction:* A third option is to analyse raw data held on media and extract alphanumeric text for storage as a separate file. Text output may be used by an investigator as a simple method of determining the textual information that exists on disk, enabling them to establish whether the disk contains information created by an author on a specific topic[30].

As a data recovery method, File Carving offers the most versatility, enabling an investigator to analyse a disk and extract many types of content (text, images, sound) encoded in different file types. File carving tools use several techniques, alone or in combination, to identify data files. Header/Footer carving and its variants are the widely supported carving method within forensic carving tools. However, its use is limited to specific use cases – file formats that possess a large header (e.g. PNG images) held on a disk with low fragmentation. Erroneous results will be produced in other circumstances, such as when searching for formats that possess little or no header/footer (e.g. ZIP or text files), or when searching for data held on fragmented drives where file content are stored in non-contiguous locations. Alternative carving methods, such as statistical and block-based carving, offer the potential to provide better support for fragmented disks, but are limited to proof-of-concept implementations at the time of writing.

3.4.6.1. File carving using open source tools

File carving may be performed using a number of open source and free forensic tools. Specific software examined by the FIDO project include the following:

- *Foremost:* Foremost (http://foremost.sourceforge.net/) is a command line tool that uses header-footer, header-maximum size, and internal data structure methods to identify start and end points of a file to be carved

- *Magic Rescue:* Magic Rescue (http://www.itu.dk/~jobr/magicrescue/) is a command line tool that uses pattern matching to identify known file types. A byte sequence is analysed for hexadecimal values that match those stored in a 'Recipe' configuration file. Recipes for new formats may be added by the user.

---

sectors, there is an increased likelihood that at least some fragment of the original content will remain, allowing an investigator to recover some of the content.

[29] Synonyms include Data Carving, or simply Carving

[30] Forensic tools, such as Bulk Extractor, may be used to extract specific information, such as email addresses and web sites visited.

- *PhotoRec:* PhotoRec (http://www.cgsecurity.org/wiki/PhotoRec): is a data carving tool that uses pattern matching to identify headers of known file types[31].

- *RecoverJPEG:* RecoverJPEG (http://www.rfc1149.net/devel/recoverjpeg.html) is a command line based carving tool that uses header/footer analyse to identify JPEG and Quicktime files.

- *Scalpel:* Scalpel (http://www.digitalforensicssolutions.com/Scalpel/) is a cross-platform file carver that uses header-footer definitions to locate file types contained within a disk image or raw device. The tool was originally based upon Foremost, but has been subsequently rewritten to utilise carving algorithms to improve the speed and accuracy of carving activities.

- *SFDumper*: SFDumper (http://sfdumper.sourceforge.net/) is a Bash script that combines the functionality of Sleuthkit, Foremost and a number of other tools to identify and carve recognised file types.

- *SleuthKit DLS:* A Unix tool contained in SleuthKit that may be used for header/footer data carving .

A number of controlled experiments were performed using header/footer carving to extract different file types (JPEG, PNG, AVI, WMV, and Ascii text). These demonstrated the variable performance of each tool, depending upon the file type to be extracted and the level of file fragmentation. PhotoRec and Scalpel were found to be the most reliable for identifying and carving data files from a disk image. However, limited time availability meant that it was not possible to perform extensive testing on a wide range of file types. Further work is required to trial data carving tools, building upon reports produced in the digital forensic community.
.

[31] The PhotoRec website claims that the tool recognises more than 200 file types (http://www.cgsecurity.org/wiki/File_Formats_Recovered_By_PhotoRec)

3.4.6.2. Integration of forensic carving tools

To enable the carving tools to be applied by archival staff, the project's software developer produced a Java front-end that could be used to configure the various parameters of Scalpel and Magic Rescue through a graphical interface.

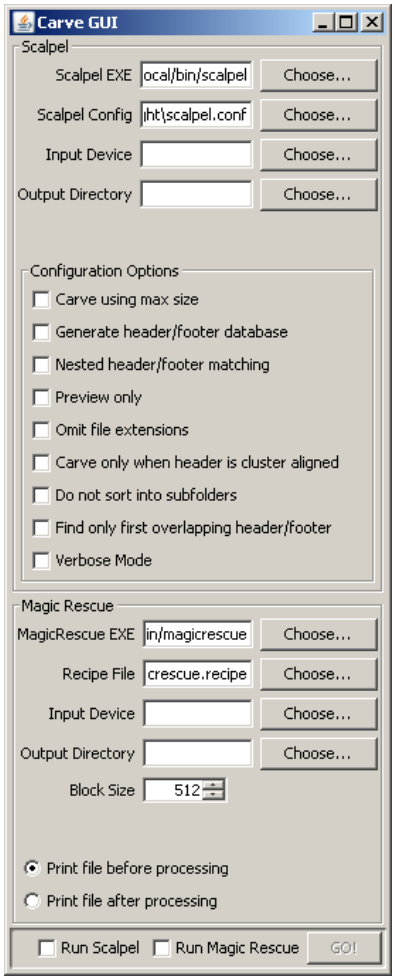

Figure 8: Java Carve GUI

Further development work is required to integrate the functionality offered by the various tools and provide output in a consistent form.

3.4.7. Bringing it all together: Developing a structured approach to investigation

The final component of the work focused upon the creation of a set of procedures that may be applied to examine digital media using the selected forensic tools and formats. The project team wished to avoid being too prescriptive in the workflow it advocated for analysing a disk and locating relevant data. Each disk will be provided under different conditions and contain varying types of data of value, requiring a combination of different techniques to be applied. To guide the process, a decision tree was developed that could be used to scope requirements and tailor the investigation to meet their needs. In the initial phase, the Investigator is encouraged to consider the context in which the disk was used, asking questions such *as "How did the creator use the disk?", "Was it in a home or work setting?"* and *"What type of data are they likely to have created?"* To address these questions, the

investigator may perform an unstructured analysis of the disk, examining the various directories and files to build an understanding of the content. Once this initial investigation has been completed, the investigator will have a better understanding of the content held on disk and the type of research questions that they should address[32]. Figure 8 illustrates the decision tree provided to Archival staff for selecting an appropriate examination method.

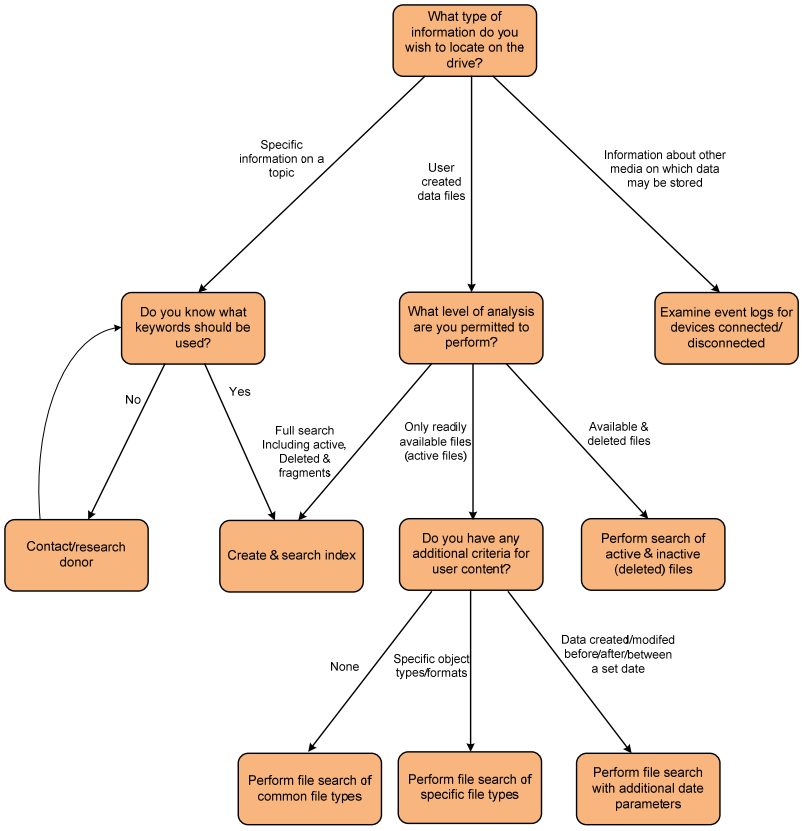

Figure 9: Decision tree for use when selecting an appropriate search method

This decision-based approach was considered to be helpful to staff members who were unsure of the approach that they should adopt to structure their investigation.

3.7. Post-project work

The project was unable to achieve its aims of introducing open source forensic software tools into the working practices of archivists. However, it identified a number of open source forensic tools that could, with further development, be modified to fit curatorial needs. A key recommendation of the project is that the Digital Curation community should actively engage/collaborate with developers of open source forensic tools to ensure that they support curation standards and enable them to be used more easily by less technical users.

[32] e.g. “Does the media contain information on topic X?”, “What files were created between X and Y?” Where do I find a password for file X?”

4. Outcomes

1. **Evaluate the suitability of digital forensic principles and practices to enable HE archives to meet organisational commitments and legal requirements for maintaining digital records;**

   The work performed by Forensic Investigators and Archivists are broadly similar: both communities are focused upon the task of acquiring, analysing, and appraising data in a manner that maintains its integrity and authenticity; and both are governed by codes of practice and standards that serve as a broad framework to evaluate the suitability of different methods and success of their implementation. However, differences in the language and methods used by these communities have traditionally served to limit the potential for knowledge transfer.

   The project demonstrated that the methods and techniques developed by forensic practitioners may be used to enhance the operation of a HE archive, if work is performed to tailor components to fit the needs of staff with different skillsets. The review of forensic models identified a number of flexible workflows that could be applied by archivists and curators to work with digital media. As a result of work performed to map and synthesis the different approaches, an integrated model was produced that could be used to inform best practice in the area and guide the investigation process. This was supported by the terminology mapping report which provides a point of reference for archival staff wishing to understand forensic principles and how they compare to existing archival concepts.

2. **Assess the effectiveness of using the chosen digital forensic tools set to identify, acquire, and analyse digital information held on digital media and computer systems in an archival environment;**

   The open source forensic tools evaluated by the project were considered to be unsuitable for use by archival staff in their current form, requiring familiarity with command line parameters and a level of technical knowledge that many staff members did not possess (see 3.4.3). As a result of this finding, the project modified its approach: it would continue to identify open source tools that could be used for a forensic investigation, but would provide the AIM Service with a forensic workstation that contained forensic software available under different licences, such as commercial and freeware, that could be easily used by non-technical staff.

   The project was able to identify a broad range of open source tools that could be adopted for use within a digital curation/archival environment, including Digital Forensic Framework, TimeScanner, PhotoRec, and Scalpel. These tools may be refined and integrated within a toolkit for use in an archival environment at a later date.

3. **Seek to embed digital forensics tools & techniques into the working practices of the KCL Archives & Information Management (AIM);**

   The FIDO project was successful in improving knowledge of digital forensic methods within the KCL Archives & Information Management service, providing staff with a broad understanding of the concepts that underpin digital forensics and the necessary expertise to begin to handle real-world collections. Training was provided through a number of hands-on sessions for groups of various sizes (the most notable being the Digital Forensics for Archivists training event), as well as the provision of documentation describing disk imaging and data location activities.

5. Outputs and Results

The primary outputs of the project are:

- A Glossary of terms that compares forensic concepts to their archival equivalent (O'Brien et al, 2011c)
- A set of reports that describe the principles of disk imaging (Knight, 2011a), forensic file identification (Knight, 2011b), and file carving (Knight, 2011c).
- An Ingest Handbook for Digital Media that describes the issues that should be considered and steps that may be taken to examine and analyse digital information held on digital media (Knight, 2011d)
- A discussion paper that provides an introduction to Document and Media Exploitation (Knight, 2011e)
- Training material, including presentations, practical exercises and handouts
- A 'Java Carve' graphical user interface for Scalpel and Magic Rescue
- Two Fact sheet on topics related to digital forensics (O'Brien, 2011a, O'Brien, 2011b)

In addition, project staff gave presentations, workshops and papers at various events, and had fruitful discussions with others as the project progressed which helped shape the work and relate it to other current activities.

The creation of the archival-forensic mapping exercise proved to be an interesting and useful exercise for the staff working on the project, prompting them to re-evaluate the archival principles that they had been taught and the methods in which they can be applied to the digital domain. Significant overlap was recognized in the approach advocated by forensic and archival practitioners, even though the medium in which they frequently work and the terminology that they use is different. The process of mapping terminology proved to be contentious, causing disagreement among various staff members. Although commonalties were identified in the concepts used, it was difficult to make a 1:1 mapping between forensic and archival terms. In a number of examples, terms used in one domain would overlap with two or more terms used in the other domain, or would refer to processes at different levels of abstraction (forensic concepts were often more practice-led than their archival equivalent). The mapping of archival-forensic terminology is likely to be an area for discussion and debate across the archival, digital curation, and forensic communities for a number of years.

The three working papers on forensic disk imaging, file identification, and file carving provided the archives service with an overview of the techniques available to acquire and analyse artefacts within the digital domain and the forensic methods and tools that can be applied. Each paper provides an introduction to the topic and practical guidance on the application of selected forensic tools that may be used within a curatorial environment. The working papers contributed to the development of the Ingest Handbook, Fact Sheets and associated papers written by the project team on the subject.

The low-cost forensic workstation built by the project has been passed onto the AIM service for use in accessioning born digital and hybrid collections. The Ingest Handbook for Digital Media outlines a semi-structured, decision-based model to be applied by archival staff to acquire, examine and analyse data in a manner that minimises the risk of accidental data change or corruption during the accessioning process. Although intended for use within KCL, much of the guidance is institution and application independent, enabling them to be adopted and adapted by researchers working in other environments.

Finally, as an extension of the project, the project manager produced a set of training materials, including Powerpoint slides and handouts, that can be used to teach forensic practices within an archival setting.

An article describing the work performed in the FIDO project has been submitted and accepted for publication in Vol 7, No 2 edition of International Journal of Digital Curation (iJDC).

Presentations on the FIDO project were given at the following events:

- Knight, G. (2011). *Digital Forensics in the Archive: Using open source & free software to capture and curate archival digital records*. Digital Preservation Coalition workshop. Digital Forensics for Preservation workshop. The Oxford Centre, Oxford. June 28 2011

- Knight, G. (2011). *Watching the Detectives: Using digital forensic techniques to investigate the digital persona*. Anatomy Museum, King's College London, 8th November 2011

- Knight, G. (2012). *New tricks for analysing old data: Digital forensics and the implications for the digital archive*. Brith Gof Seminar/workshop. National Library of Wales, Aberystwyth. 30th May 2012

In addition, a two hour lesson on Digital Forensics and its role in Digital Curation was provided to students enrolled on the 'Digital Preservation: Theory & Practice' module on March 28th 2012, as part of the MA Digital Asset Management course at King's College London.

Digital Forensics for Archivists training event

On August 16th 2011, the FIDO project ran a 'Digital forensics for archivists' training event for staff working within the Archives & Information Management (AIM) service and IT Systems at King's College London. The aim was to provide them with an understanding of digital forensic methods and tools and an overview of the work perfomed by FIDO.

Title	Description	Presenter
Introduction and Archiving Scenarios	Introduction to digital forensics and comparison to archival appraisal processes; description of scenarios where digital forensics techniques may be helpful	Lindsay Ould
Archival-Forensic Vocabulary Mapping	A description of the concepts that underpin digital forensics and outline of their archival equivalent	Kate O'Brien
Applying Digital Forensic techniques to AIM	An outline of the forensic model and forensic techniques applied by the FIDO project	Gareth Knight
Discussion session	Event participants were separated into four breakout groups and asked to discuss the practical and ethical issues to be considered when handling data obtained	Lindsay Ould (lead)

	in four different scenarios.	
Hands on session	Practical exercise to analyse a 60gb disk image, identify data of potential value and determine the owner's data handling practices using OSForensic.	Gareth Knight (lead)
Feedback and Summary	Broad discussion of the practical sessions	Gareth Knight and Lindsay Ould

The sessions were well received by archival staff, many of whom should significant interest in using the forensic techniques and tools in their own work. A full write-up of the training session may be found at the following URLs:
http://www.kcl.ac.uk/innovation/groups/cerch/about/archive/news/2011/digitalforensics.aspx
http://fido.cerch.kcl.ac.uk/?page_id=53

6. Implications

Forensic Model
The synthesised Forensic Model, constructed by mapping several existing models onto a common framework and incorporating archive specific activities, outlines a broad structure for processing digital media using forensic methods within an archival setting. This is supplemented by a number of decision trees, intended to assist archivists when deciding upon the approach that they should adopt. Value may be gained by developing the model further, identifying pre-ingest activities performed in a range of academic environments. Further consideration should also be given to the decision associated with each stage.

Selection criteria for imaging format
The disk image format selection criteria outlined in this report represents the first (known) attempt within the digital curation community to specify a set of factors for evaluating the suitability of image formats for curation and preservation purposes. This list may be extended and enhanced by other institutions, by refining and extending criteria, and tailoring the vocabulary in use. Further work may also be performed to map the selection criteria onto a formalised decision model, such as the PLATO Decision Tree for preservation planning.

Forensic tools
The FIDO project identified a number of open source and free forensic tools that offer functionality not currently provided by commercial solutions and are currently underused within the curation community, due to the difficulty in setting up and configuring them. Work is required to extend these tools to meet the needs of archival and curatorial staff. These tools may be integrated within broader system architectures, such as those offered by Fedora Commons, ePrints, Curator's Workbench, and others.

7. Recommendations

- Development work should be performed to integrate the functionality provided by disparate forensic tools and enable them to be used by a non-technical audience.

- File format registries and datasets should be updated to enable identification of disk image formats.

- The digital curation community should actively engage/collaborate with software developers responsible for digital forensic software to ensure that these tools incorporate features useful for managing disk images in a curation environment. Particular areas for

development should include: extending disk imaging software to enable the capture of user/machine-provided descriptive metadata and audit trails, for storage in appropriate metadata formats (e.g. simple/qualified Dublin Core; PREMIS Events, Open Provenance Model).

- Forensic hashing techniques provide a useful starting point for locating unknown digital files that may represent user-created content. However, the generated output frequently lists a large number of files that will be superfluous to the investigator's needs. Development work should be performed to combine forensic hashing with alternative analysis techniques, such as hash de-duplication and name identification, to further categorise unknown files into sub-types.

- There is a need for further case studies that explore how forensic investigation methods may be integrated into the operation of digital archives in the academic and, more broadly, cultural heritage sector.

- As an extension of the above, further research is required to explore the policy implications of storing disk images for the short or long-term. Building upon the work performed in the FIDO project, thought should be given to the selection criteria for choosing disk imaging formats and implications for the collection and retention policies, among others.

- Research should be performed to establish the most effective method of linking different manifestations of an object, when one is contained within a disk/memory image. A possible avenue for investigation is to use RDF metadata contained within an AFF4-encoded disk image.

References

Agarwal, A & Gupta, M. 2011. “Systematic Digital Forensic Investigation Model”. *Proceedings of the 5th National Conference; INDIACom-2011. Computing For Nation Development*, March 10 – 11, 2011. ISSN 0973-7529 ISBN 978-93-80544-00-7

Amber L. Cushing, 2010. “Highlighting the archives perspective in the personal digital archiving discussion”, *Library Hi Tech,* Vol. 28 Iss: 2 pp. 301 – 312. Accessed September 04, 2011: http://dx.doi.org/10.1108/07378831011047695

Bancroft, J. O’Brien, K, McCarthy, E. Smith, L. 2010. “Archives & Information Management: Acquisition Guidelines". Internal document

Carrier, B. & Spafford, E.H. 2003. “Getting Physical with the Digital Investigation Process”. "International Journal of Digital Evidence Fall 2003, Volume 2, Issue 2. Accessed September 04, 2011: https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2003-29.pdf

CCSDS. 2002. Recommendation for Space Data System Standards. Reference Model for An Open Archival Information System (OAIS). CCSDS 650.0-B-1 Blue Book. January 2002. Accessed September 04, 2011:
http://public.ccsds.org/publications/archive/650x0b1.pdf

Cushing, A. L. 2010. “Highlighting the archives perspective in the personal digital archiving discussion”. *Library Hi Tech. Volume 28 issue 2*. 301 – 312. Emerald Group Publishing Limited. Accessed September 04, 2011: http://first.emeraldinsight.com/journal.htm?issn=0737-8831&volume=28&issue=2&ft=1&article=2380280210

Palmer, G. 2001. A Road Map for Digital Forensic Research: Report From the First Digital Forensic Research Workshop (DFRWS). DFRWS Technical Report. Utica, New York. August 7-8, 2001
http://www.dfrws.org/2001/dfrws-rm-final.pdf

ForensicWiki (n.d.). *File Carving*. Retrieved May 21 2012 from:
http://www.forensicswiki.org/wiki/File_Carving

Garfinkel, S.L. 2006. Forensic feature extraction and cross-drive analysis. Digital Investigations. 2006. S71 – S81. Accessed September 04, 2011: http://simson.net/clips/academic/2006.DFRWS.pdf

Ghirardin, A. 2009. Building a low cost, highly scalable, enterprise level computer forensics lab. Accessed September 04, 2011: (http://web.mclink.it/MC8247/Building-CF-Labs.pdf):

Kaye, J. J., Vertesi, J., Avery, S., Dafoe, D., David, S., Onaga, L., Rosero, I., & Pinch, T. 2006. “To have and to hold: exploring the personal archive”. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 275–284). New York: ACM Press. Accessed September 04, 2011: http://jofish.com/writing/tohaveandtohold.pdf

Kent, K. Chevalier, S. Grance, T. & Dang, H. 2006. “Guide to Integrating Forensic Techniques into Incident Response: Recommendations of the National Institute of Standards and Technology”. National Institute of Standards and Technology Special Publication 800-86. Accessed September 04, 2011: http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf

King, G.L. 2006. “Forensics Plan Guide”. *SANS Institute*. Accessed September 04, 2011: http://computer-forensics.sans.org/community/papers/gcfa/forensic-investigation-plan-cookbook_283

King’s College London. N.d. “Archives & Corporate Records Service (ACRS): Acquisition Policy”. Accessed September 04, 2011: http://www.kcl.ac.uk/content/1/c6/06/81/50/ACRSacquisition.pdf

Kirk, D. & Sellen, A. 2008. “On human remains: Excavating the home archive”. *Microsoft Technical Report.* Accessed September 04, 2011: http://research.microsoft.com/apps/pubs/default.aspx?id=70595

Kirschenbaum, M.G., Ovenden, R. & Redwine, G. 2010. “Digital Forensics and Born-Digital Content in Cultural Heritage Collections”. *Council on Library and Information Resources*. Accessed September 04, 2011: http://mith.umd.edu/wp-content/uploads/whitepaper_borndigital.pdf

Knight, G. 2011a. “Forensic Disk Imaging Report. Accessed http://fido.cerch.kcl.ac.uk/

Knight, G. 2011b. “Forensic File Carving Report”. Accessed http://fido.cerch.kcl.ac.uk/

Knight, G. 2011c. “Forensic Hashing Report”. Accessed http://fido.cerch.kcl.ac.uk/

Knight, G. 2011d. “Ingest Handbook for Digital Media”. Accessed http://fido.cerch.kcl.ac.uk/

Knight, G. 2011e. “Document and Media Exploitation: What is it and how can it be applied to an academic environment?”. Accessed http://fido.cerch.kcl.ac.uk/

Knight, G. 2012. “The Forensic Curator: Digital Forensics as a solution to addressing the curatorial challenges posed by Personal Digital Archives*”. International Journal of Digital Curation*. Forthcoming

National Institute of Standards and Technology (NIST), (n.d.) Computer Forensic Tool Testing: Disk Imaging
http://www.cftt.nist.gov/disk_imaging.htm

National Institute of Standards and Technology. 2004. “Digital Data Acquisition Tool Specification. Draft 1 for Public Review of Version 4.0”. Accessed September 04, 2011: http://www.cftt.nist.gov/Pub-Draft-1-DDA-Require.pdf

O’Brien, K. 2011a. “Fact sheet for donors of digital archives”. Forthcoming

O’Brien, K. 2011b. “Fact sheet for archivists: Lessons learned from the FIDO project”. Forthcoming

O’Brien, K. Ould, L and Knight, G. 2011c. “Forensic-Archival Terminology Mapping”. Forthcoming

Olson, M. 2010. “Digital Forensics @ Stanford Libraries: Why we have FRED and why you don't need one?” Accessed September 04, 2011: http://www.rbms.info/conferences/preconfdocs/2010/SeminarIOlson.pdf

Pollitt, M. 1995. “Computer Forensics: an Approach to Evidence in Cyberspace”. Proceedings (of the National Information Systems Security Conference, Vol. II, pp 487-491.

Baltimore, MD. 1995. Accessed September 04, 2011: http://www.digitalevidencepro.com/Resources/Approach.pdf

Pollitt, M. 2004. “Six Blind Men from Indostan”, Digital Forensic Research Workshop 2004, Baltimore, MD. http://www.dfrws.org/2004/bios/day1/D1-Pollitt-Keynote.ppt

Pollitt, M.M. 2007. “An Ad Hoc Review of Digital Forensic Models”, in *Proceeding of the Second International Workshop on Systematic Approaches to Digital Forensic Engineering* (SADFE’07), Washington, USA.

Reith, M., Carr C. and Gunsch, G. 2002. “An Examination of Digital Forensic Models”. *International Journal of Digital Evidence*. Fall 2002 Volume 1, Issue 3. Accessed September 04, 2011: http://www.utica.edu/academic/institutes/ecii/ijde/articles.cfm?action=article&id=A04A40DC-A6F6-F2C1-98F94F16AF57232D

Ruibin, G et al. 2005. “Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework”. *International Journal of Digital Evidence, Volume 4, Number 1, Spring 2005*. Accessed September 04, 2011: http://www.informatik.uni-trier.de/~ley/db/journals/ijde/ijde4.html

Sue McKemmish, 2005. ”Traces: Document, record, archive, archives”, in: Sue McKemmish, Michael Piggott, Barbara Reed and Frank Upward (eds.), Archives: Recordkeeping in Society (Charles Sturt University, Wagga Wagga 2005) 1-20.

Todd, M. 2009. File Formats for Preservation. Digital Preservation Coalition Report. *55*. Accessed September 04, 2011: http://www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation

Whittaker, S. & Hirschberg, J. 2001. “The character, value, and management of personal paper archives”. *ACM Transactions on Computer-Human Interaction, 8*(2), 150–170. – Accessed September 04, 2011: http://dis.shef.ac.uk/stevewhittaker/move-paper_final_revised.pdf

Williams, P., Dean, K., Rowlands, I., & John, J. L. 2008. “Digital lives: Report of interviews with the creators of personal digital collections”. *Ariadne, 55*. Accessed September 04, 2011: http://www.ariadne.ac.uk/issue55/williams-et-al/