Abstract

Digital preservation and digital forensics are two fields with differing goals that travel similar pathways which often converge. Each field does not necessarily acknowledge the other, but they are closely aligned and share similarities. Digital preservation has much to benefit from digital forensics; however, this is not to say digital forensics could not gain with respect to documentation and perspective with collaboration in mind.

One of the key differences is long-term preservation, where the material is cared for long after it has been processed versus forensic evidence gathered and used to prosecute, with no further regard once done so. The efforts that go into ensuring the preservation of digital objects are where the similarities between the two fields end. This results in forensic software being tailored to the specifics of the field, such as modern devices, specific data, and criminal prosecution. Perspective and purpose are important factors as they determine how the software is perceived and documented. This affects the adoptability of forensic software for memory institutions (galleries, libraries, archives, museums) as at face value, it does not cater to their needs, despite being beneficial.

In this thesis, the benefits of using forensic software for born-digital preservation are explored, as well as the risk to collections should data remain unprocessed via the suggested methods. Hidden data may already exist within storage collections, yet to be discovered and impossible to do so without the use of forensic software. These data, rightly named “sensitive data” have many implications. Sensitive data, whilst the key to criminal investigations, are also paramount to digital preservation as they can reveal ample amounts of new information. Australian law is explored regarding the risk of sensitive data discovery and the actions that follow. The policies in place are compared against those tailored towards Aboriginal and Torres Strait Islander material.

The scope has been narrowed down to Australian institutions, targeting State and National libraries whilst also considering archives, galleries, and museums, as these are the influential institutions, likely to lead smaller institutions by example. Australian institutions were investigated by the information publicly available and by communications, distributing a questionnaire to willing participants. Overseas institutions were investigated to form a comparison and to establish potential tools and methods that could be adopted. The data gathered from overseas institutions were derived by publicly available information and other
studies conducted. The main sources of data were derived from workflows as these allowed a visual representation of the processes and the tools used within collection institutions, revealing if and where forensics was being utilised.

It was evident the major collection institutions of Australia were at different maturity levels of digital preservation intake and performance. Some of the participants of the study identified the need for improvement regarding their workflows, whereas others had low demand and therefore did not see the need to make any changes. It was determined that some digital forensics was being utilised, but not to its full potential, and in most cases, was missing completely. The analysis of collection institutions and the benefits of digital forensics aim to increase awareness and provide workflow improvements to address issues surrounding sensitive data by incorporating forensic tools and methods.

The identification of maturity levels for digital preservation in Australian institutions has been established by the feedback provided via questionnaire and data gathered from public sources. This information, compared with other institutions, allows the establishment of an average baseline in terms of where the institutions are at with their digital preservation requirements/performance. Combined with an analysis of the data gathering capabilities of forensic software and the relevance to digital preservation allowed the contributions of this study to form. This includes real world experiments on donated material (hard drives), resulting in a plethora of data gathered with an extensive range in severity. The potential for sensitive data discovery was revealed as well as the ability to derive information about the users of the physical media.

The identification of gaps in digital preservation workflows, specifically the lack of transparency and undocumented handling of sensitive data discovery, has been reviewed and documented. Through this process, good practices were also identified, resulting in the discovery of exemplary workflow designs to help in determining how these gaps can be filled. The threats of sensitive data are explored as well as the potential law implications investigated that surround these issues. This includes examples of future and already existing threats that may reside in stored data that have not been processed assiduously using forensic software.

Amendments and enhancements to workflows to address sensitive data discovery are presented, enhancing digital preservation workflows with digital forensic tools and methods. This is not only to improve existing institutions, but to better enable up and coming
institutions to learn from others and learn current best practice. If institutions that are at a higher maturity level than others implement forensic processes into their workflows and visualise the steps taken, addressing all outcomes and decision making, then through this transparency, the peer institutions have benchmarks to follow and learn from.