Abstract – Collection institutions (Libraries, Archives, Galleries, and Museums) are responsible for storing and preserving large amounts of digital data, which can range from historical/public figure records, to state or country-wide events. The ingest process often requires sifting through large amounts of data which may not always be sorted or categorized from the source/donor. It is possible to discover information that was not intended to be disclosed should the donor not be privy to the existence of said material. This issue is typically handled by communicating with the donor, however, if they have no relation to what has been uncovered in the data, further steps may need to be taken. If the data belong to or are about someone living, that person may need to be contacted, depending on the nature of the data discovered. If the person of interest is no longer living, legally there would no issue disclosing all information uncovered, however, implications for living relatives must be considered should the disclosed information be potentially revealing or harmful to them. This can include hereditary health issues, political or religious views, and other sensitive information. There are significantly more variables to consider, such as public interest and defamation which can heavily impact the decision process following the discovery of sensitive data, all whilst guided, but not necessarily enforced by law. This remains somewhat of a gray area as the entities handling such data are often exempt from these laws and principles, making these decisions ethically and morally based more so than legally. In this article, the Australian laws and policies that surround privacy issues, defamation, and data relating to Aboriginal and Torres Strait Islander people and culture are explored. The aim is to raise awareness on potential issues that may arise in collection institutions as well as potential threats already sitting in storage and the laws and policies that may serve as guidelines to help overcome/mitigate such issues.

Keywords – access to information, defamation, privacy, sensitive information, Australian Law
Conference topics – Designing and Delivering Sustainable Digital Preservation; Exploring New Horizons

I. INTRODUCTION

Procedures for born-digital preservation have not yet been standardized among the many institutions forming such actions. Some institutions are progressive and are actively making advancements in born-digital preservation, whereas others are still in their infancy when it comes to preserving born-digital content. While digitization of hard-copy material is certainly part of digital preservation, as researched by LeClere [1], there are far more potential issues surrounding born-digital content. Although these issues are global, the laws which cover them are country specific. This paper is focused on the Australian jurisdiction. The main issues where Australian law may hinder the process are present during the ingest phase and the storage phase, specifically where access is made available. Examples of such issues will be discussed under the Ingest Scenarios section.

Although not always obligatory for all entities, laws and policies exist for good reason. Currently, the main entities performing digital preservation within Australia fall into this area, namely Libraries, Archives, Museums, and Universities. The material these entities store and make publicly available are exempt from the Privacy Act 1988 and the Australian Privacy Principles (APPs) within. As stated in the National Library of Australia privacy policy:

This policy sets out how the National Library of Australia (the Library) approaches and manages...
the Australian Privacy Principles (APPs) contained in Schedule 1 to the Privacy Act 1988 (Privacy Act).

The Privacy Act regulates how Commonwealth agencies such as the Library collect, store, use and disclose personal information, and how individuals can access or correct personal information the Library holds. It requires the Library to comply with the APPs and take reasonable steps to implement practices, procedures and systems to protect personal information.

The Privacy Act does not apply to library material held, man-aged and made accessible by the Library, whether published (such as books, journals, newspapers and websites) or un-published (oral history interviews, photographs and archival collections). [3]

However, the Privacy Act still applies to any personal user information collected from Library services as well as from all Library employees, temporary staff, volunteers and contractors.

The records held regarding Australian Aboriginal and Torres Strait Islander people have their own surrounding issues along with protocols to guide collection institutions through them. Extra care must be taken in order to maintain the customs of Indigenous peoples and to ensure the handling of their material is done accord-ing to their cultural needs. One must again emphasize, there may not be a definitive law regarding such actions, but collection institutions should feel ethically obliged to follow relevant protocols to comply with best practice. Being aware of existing laws and the issues which they aim to prevent, is a necessity for not only adopting best practice, but preparing for any future changes to privacy law.

II. SENSITIVE AND IDENTIFYING INFORMATION

Before understanding the laws that may affect digital preservation, it is important to understand the source of the issues and where they may arise. This understanding is crucial, as often the solution must come down to a judgment call, basing decisions on variables guided, but not often enforced, by Australian Law.

Sensitive and identifying information can be found on digital media by using forensic software tools which are freely available and easy to use, such as the BitCurator environment [4], bulk_extractor [5] and The Sleuth Kit (Autopsy) [6]. Material is often donated to collection institutions and this can lead to a range of issues. Libraries offer donor agreements which form a contractual agreement between library and donor, stipulating all conditions from both parties and how to handle the data once collected. These agreements may also pass ownership from the donor, removing them from any further say in the matter.

One issue is the discovery of sensitive data. In most scenarios the donor agreement will typically have instructions in place on how to handle this. However, there are scenarios where the solution is not so easily solved. Firstly, what data are classified as sensitive must be established along with what information can be used to identify an individual. In Part II-Division 1 of the Privacy Act [7], identifying information and sensitive information are defined as followed:

**Identifying information**
- Full name
- Alias/Previous name
- Date of birth
- Sex
- Last known address (including 2 previous)
- Name of current/past employer
- Driver’s license

**Sensitive Information**
- Racial or ethnic origin
- Political opinions/membership association
- Religious beliefs/affiliations
- Philosophical beliefs
- Membership of professional/trade association or union
- Sexual orientation or practices
- Criminal record
- Health/genetic information
- Biometric information/templates

Regarding the list of sensitive information, these data can be derived by online activity and how the user in question went about their daily activities on the device on which the donated material was created. Whilst there may not be an individual element that clearly species any of these elements, there may be definitive clues. Much of this information lies deep in a system, obscure, and difficult, if
not impossible to find by manual means (navigating directories without the assistance of a forensic tool). One tool that suits this need is bulk_extractor [5] which can be used to discover anywhere between thousands to millions lines of data deemed sensitive or personal. With this tool, online activity such as websites visited, which elements within that website were viewed, and any sub-pages visited are revealed. Emails, Face- book, Web browser searches, and much more can be derived and analyzed to establish information about the user.

For example, health or genetic information could potentially be established if the user frequently researched and visited websites on a health issue. Personal information could be revealed in emails. Religious beliefs and affiliations could also be revealed by online activity, contacts, and communications. Sexual orientation and practices are readily revealed should the user frequent pornographic websites. There is an abundant amount of data that are collected overtime, a digital footprint, something the average user typically will not put much effort into hiding. These data have much potential, both good and bad.

The following is a real-world example. The data have been taken from a real hard drive and processed through bulk_extractor. Any personal information has been redacted and the example has been carefully selected.

Bulk_extractor detected a high number of URL searches relating to job seeking:

“employsa.asn.au”, “Job+Search”, “Retail+Jobs”, “resumes”

Another discovery was visits to the McDonalds’ login page. It was also discovered that there was an official McDonalds’ email address assigned to the user. All this information together strongly suggests the user was employed by McDonalds. Correlated against other information and further investigation, it would not be farfetched to say one could establish which workplace the user was assigned to and how much further the investigation could go.

This example shows how individual elements, typically undetectable without the aid of forensic tools, combined with other data can reveal a lot about an individual, often sensitive and personal in nature.

Note that whilst numerous records handled by collection institutions are historic and often relating to a deceased person, their sensitive information may still affect any living family members.

This relates to health and genetic information. If the information collected indicates the deceased person had a medical condition that is inheritable, this reveals possible health information for their descendants [8].

While collection institutions must abide by the laws surrounding privacy with the consumer data they hold, e.g. account information for library users and staff, the collection material itself is exempt from such law. How-ever, this does not mean the laws should not be at least considered as guidelines, influencing policies and procedures for handling sensitive data within collection institutions. The State Library of NSW provided a ‘Sensitive Collections Material Policy’ in 2017 that addresses this with the opening statement as follows:

As part of the Library’s collections there is a significant number of records containing people’s personal information or, content that is considered culturally sensitive to Indigenous Australian peoples. Examples of these records include medical records, records of children in care, legal records and Indigenous cultural material. Library collection material is exempt from both the Privacy and Personal Information Project Act (1998) and Health Records and Information Privacy Act (2002), however in the spirit of this legislation and based on best practice considerations, the Library sees an ethical obligation to protect people’s personal and cultural information. Of equal importance to the Library is enabling individuals to seamlessly access information about themselves and their cultural heritage, especially those who have experienced institutional or other out-of-home care. In light of both of these considerations, this Policy outlines access guidelines to sensitive and private records held in the Library’s collections [9].

The policy goes on to address all instances of sensitive information and lists time restraints for each type of record. Using the privacy laws as guidelines for ethical obligations is something more
collection institutions should aim for as it provides a more trustworthy repository for people to commit to and prepares that institution for any future legal changes.

A. Ingest Scenarios

One of the key elements that must be identified is how the donated material relates to the donor and how they came into possession of it. There are many possibilities which change the severity of risk associated with handling such material.

Example 1 - The donated material belongs to and is data about the donor.
Example 2 - The donated material is of ancestry significance to the donor.
Example 3 - The donor has no relation and has discovered or purchased media in which the donated material was found (known material of significance to collection institution).

These examples relate to events prior to ingest as they would dictate how the donor agreement is written up. However, once the data have been collected and processed, further issues may arise as information can be discovered that was not intended nor covered specifically in the donor agreement. Even if the donor had searched through the material before handing it over, there is a chance they missed something. With training and the right tools, significant amounts of information can be uncovered on a system in obscure places, as well as rich amounts of metadata.

Following Example 1, once the donated material has been analyzed, should sensitive information be discovered, further decisions must be made based on what the sensitive information is. If this is covered in the donor agreement, then action should proceed as stated within the agreement. If the agreement does not cover the unsecured data and the donor is available, the donor needs to be involved as any decisions on how to proceed with the uncovered material. There are a few more variables that complicate this procedure. The information may in-criminate the donor and depending on the severity and nature of the discovery, law enforcement may need to be involved.

If this scenario were based on Example 2, this may lead to difficulties for living descendants, however, if no direct harm is caused by disclosing the information, legally there is nothing preventing it. The descendants may fight it and they may try to sue for defamation on behalf of their ancestor, or themselves. However, it should be noted that this is a gray area with an inconsistent history. This is discussed further in section III subsection B.

Another outcome, more likely to occur with Example 3, is the information discovered on donated material may be withheld from the public in their best interest. This may be relating to a public figure, loved and idolized by the country where the discovered material, whilst harmless, may alter how the public sees that figure. Alternatively, the information may need to be disclosed in the best interest of the public, commonly known as "Public Interest Disclosure" [10]. The donor would have likely signed all ownership of the material over to the collection institution as it has no relevance to them, meaning no further involvement from the donor is necessary in any decision making. There may be policies in place that help in handling such situations, but for many smaller institutions, this may be unprecedented which ultimately makes this an ethical and moral decision. It is situations like these that make this field difficult to develop definitive solutions for because no two cases will be the same, there are always gray areas and variables that complicate decision making.

III. Laws

As mentioned in the introduction, collection institutions such as national and state libraries and archives are exempt from privacy law regarding their collection material. However, it is important to familiarize oneself with the Privacy Act and the Australian Privacy Principles (APP) as well as determining if you are in fact an APP entity. The APP guidelines define an 'APP Entity' to be an organization or agency. The APP [8] define an organization to be:

An individual, a body corporate, a partnership, an unincorporated association, or a trust. This excludes organizations such as a small business operator, registered political party, state or territory authority, or a prescribed instrumentality of a state.
The APP defines Agencies as (but does not include State or Territory agencies):

- A minister, a department, a federal court, Australian Federal Police, a Norfolk Island agency, the nominated Australian Government Health Service (AGHS) company, an eligible hearing service provider, or a service operator under the Healthcare Identifiers Act 2010. Individuals may also fall under the agency category if they hold or perform duties of an office established by or under a Commonwealth enactment, or duties for the Governor-General, a Minister, as well as bodies established or appointed by them.

The APPs outline how personal information is handled, used, and managed by APP entities. This applies to most Australian and Norfolk Island Government agencies, private sector and not-for-profit organizations (with an annual turnover greater than A$3 million), private health service providers, and some small businesses. Small businesses (A$3 million or under) have responsibilities under the act if any of the following are true:

- Private sector health service providers, businesses that sell or purchase personal information, credit reporting bodies, contracted service providers for a Commonwealth contract, employee associations registered or recognized under the Fair Work Act 2009, businesses that have opted-in to the Privacy Act, businesses that are related to another business covered by the Act, or businesses prescribed by the Privacy Regulation 2013 [11].

Both the Privacy Act and the APPs are quite extensive, so each principle will not be discussed in detail, but the 13 APPs from the Privacy Act 1988, Schedule 1 are:

- **APP 1:** open and transparent management of personal information
- **APP 2:** anonymity and pseudonymity
- **APP 3:** collection of solicited personal information
- **APP 4:** dealing with unsolicited personal information
- **APP 5:** notification of the collection of personal information
- **APP 6:** use or disclosure of personal information
- **APP 7:** direct marketing
- **APP 8:** cross-border disclosure of personal information
- **APP 9:** adoption, use or disclosure of government related identifiers
- **APP 10:** quality of personal information
- **APP 11:** security of personal information
- **APP 12:** access to personal information
- **APP 13:** correction of personal information

Data security and privacy is always a current issue, ever changing, and highly desired. New Government Legislation Acts and policies are often being created, as are current ones being reviewed and amended as needed. Therefore, it is beneficial to be aware of such changes, for they may not be obligatory for your institution at the present time, but things can change. The European General Data Protection Regulation (GDPR) is a prime example as many would be aware by the policy updates from each service subscribed to. All Australian businesses need to comply if they have dealings in or with the European Union (EU). This includes having a branch in the EU, offering goods and services in the EU, and even if the business is monitoring individuals within the EU. The GDPR shares many requirements with the Privacy Act 1988, but there are additions that are not covered in the Act, one of which is the right to be forgotten [12]. Whilst compliance may not be mandatory, careful review of updated polices and requirements can lead to adopting best practices and better policies.

### Collection Institutions

There are a few circumstances in which collection institutions need to consider law. These include holding information, making it public, and how the information is being used. The main area of focus is the publicizing of information, as this is where the biggest potential threat lies. There are also risks surrounding the content held within collection institutions, however, there are restricted sections where this information is kept from the public. These sections require special access or permissions by the author or representatives. The National Library of Australia’s restricted area, known as the “Secure Room – Restricted” (SRR) is said to be almost as hard to access as a bank vault with its door shut [13]. Content is held within the SRR for various reasons, some of the main ones according to Gidney include:
• Secret/Sacred Indigenous material.
• Litigation – Ongoing court cases/upheld claims (defamation).
• Commercial in confidence .
• Refused classification (RC).
• Publication with significant/dangerous errors.

This list alone illustrates the need to carefully consider what information is made public, as you can imagine the potential risks involved, should this listed content not be made secure. Secure areas also serve as a holding place for original documents that may have had information omitted for publicly accessible versions. Gidney listed one such case where in 1997 Goodbye Jerusalem by Bob Ellis had a sentence omitted that made some offensive and damaging claims. Furthermore, on the topic of making information public, the disclosure of information marked “commercial in confidence” is forbidden without permission from the supplier. This includes any information that may result in damages to a party’s commercial interests, intellectual property, or trade secrets [14].

B. Defamation

Defamation is defined similarly from country to country, but one of the better definitions posted in an article in ‘The News Manual’, sourced from the British Defamation Act of 1992 is defined as:

The publication of any false imputation concerning a person, or a member of his family, whether living or dead, by which (a) the reputation of that person is likely to be injured or (b) he is likely to be injured in his profession or trade or (c) other persons are likely to be induced to shun, avoid, ridicule or despise him. Publication of defamatory matter can be by (a) spoken words or audible sound or (b) words intended to be read by sight or touch or (c) signs, signals, gestures or visible representations, and must be done to a person other than the person defamed. [15]

Prior to January 2006, defamation law varied across each state in Australia, but is now covered under the Uniform Defamation Law [16]. Furthermore, there was a distinction between libel and slander prior to the uniform law, however, the distinction was already disregarded in five jurisdictions and the rest of Australia followed with the introduction of the new law [17]. Regarding organizations and companies having the right to sue for defamation, this was possible under the old act, however, under the uniform law, if the corporation exceeds 10 employees, they cannot sue. This does not include not-for-profit organizations, and it does not include individuals within corporations of 10 or more employees if they are identified in the defamatory publication [18].

With all that in mind, it may seem unwise to publicize information, however, there are defenses against defamation claims and they are quite solid. First and foremost, ‘truth’ is the strongest defense, more so now under the uniform law as public interest is no longer a requirement needed to supplement the truth claim [18], [19]. As long as there is substantial evidence proving the information to be true, the defamation claim will not succeed. Should the claim be won, it may result in actions taken such as in the Goodbye Jerusalem case where the defamatory statement was omitted in the public version. The truth remains the strongest defense for collection institutions, however, it is void should ‘malice’ be proven, that is, if the information was published with ill-will or with harmful motives. It should also be noted, that should the published material be based on a deceased person, they cannot legally be represented in a defamatory case, even by family members. This of course can change should the published material cause harm for living family members, but they can only claim defamation on their own behalf, they cannot clear the name of their deceased family member [18].

The other defenses include: absolute privilege, qualified privilege, honest opinion, innocent dissemination (unintentional defamation), and triviality. For collection institutions, innocent dissemination is possible, but un-likely as items should be carefully reviewed before being published. Triviality may also prove to be a worthy defense, but the other defenses are not as relevant. Absolute privilege covers speech in parliament and court proceedings, meaning whatever is said and whatever motive

[1] Australian Federal politicians Peter Costello and Tony Abbott sued publisher Random House over Bob Ellis’s memoir Goodbye Jerusalem, which featured gossip falsely claiming that they had been ‘lured to the Liberal Party’ by a sexual liaison.
behind it cannot be used to sue for defamation. The reports of these proceedings are then protected by qualified privilege, however, only applicable if the report is honest, for the public, or the advancement of education [18].

IV. Aboriginal and Torres Strait Islander Material

Within Australian collection institutions, historical records are held containing information on Aboriginal and Torres Strait Islander affairs. There are unique policies and procedures for dealing with such records, one of which is commonly used in libraries called the Aboriginal and Torres Strait Islander Library, Information and Resource Network (ATSILIRN). The ATSILIRN protocols act as guidelines for librarians, archives, and all information services that interact with Aboriginal and Torres Strait Islander people or handle materials with such content [20].

The protocols were published in 199s by the Australian Library and Information Association (ALIA) and were then endorsed by ATSILIRN. Updates to the protocols took place in 200s and again in 2010, with 2012 being the latest revision. Once again, these serve only as guidelines, they are not definitive and must be interpreted and applied in context for each issue or situation the protocols may be needed. The protocols cover the following categories:

- Governance and management
- Content and perspectives
- Intellectual property
- Accessibility and use
- Description and classification
- Secret and sacred materials
- Offensive
- Staffing
- Developing professional practice
- Awareness of peoples and issues
- Copying and repatriation records
- The digital environment

Due to Indigenous protocol and sensitivities, some Aboriginal and Torres Strait Islander material must be locked in secure sections of collection institutions, an example of which can be found in the SRR of the NLA. Some of this material may also impose access restrictions and can only be accessed via special permissions such as content classified as ‘secret men’s’ or ‘secret women’s’ business, adding further conditional access [13].

In 2007, the National and State Libraries of Australasia (NLSA) developed a framework to guide National, State, and Territory libraries on how to approach Aboriginal and Torres Strait Islander library services and collections. However, this was superseded in 2014 with the ‘National position statement for Aboriginal and Torres Strait Islander library services and collections’ [21]. Within the position statement, it is made clear that the following policies/protocols are endorsed: The ATSILIRN, The United Nations Declaration on the Rights of Indigenous Peoples [22], and The National and State Libraries of Australasia Guidelines for Working with Community [23]. The standards that are promoted within the position state- ment include: Rights to be informed about collections relating to the people (culture, language, heritage).

The right to determine access and use of such material. Inclusion of Aboriginal and Torres Strait Islander peoples in all decision-making processes at all levels. Strategies to increase employment and retention of Aboriginal and Torres Strait Islander staff within the library and information sector. Strategies to strengthen cultural competency across the workforce, raising awareness and knowledge on issues for Aboriginal and Torres Strait Islander library users. Strategies to make usable copies of collection material to be returned to the rightful people to support cultural and language maintenance or revitalization.

In summary, the promoted standards aim to ensure rights are given to the people relating to the content, ensuring they have the rights to decide how content is handled and managed, to give the people a chance to be part of the process and to give back to the communities where possible.

Another important position statement from the NLSA is on Intellectual Property and how it differentiates Indigenous content and non-Indigenous content [24]. The World Intellectual Property Organization describes how intellectual property is expressed by Indigenous peoples with the following principles:
• Intellectual property is handed down, generationally (orally or by imitation).
• It reflects community cultural and social identify. It consists of characteristic elements of a community’s heritage.
• It can be produced by unknown authors or by communally recognized communities/individuals that have been granted the right, responsibility, or the permissions.
• It can often be created for spiritual/religious purposes and is something that constantly evolves within the community.

How Australian collection institutions handle Indigenous material and peoples is a good example of the importance of guidelines and protocols. While not bound by definitive law, we still must consider the affect our collected material can have on others, making this about ethically based, best practice decisions. This should be standard for all material, not just that of Australian Aboriginal and Torres Strait Islander content.

V. Conclusion

Whilst many institutions are yet to encounter issues such as those mentioned in this paper, it does not mean the potential for such issues to occur is not already present. Institutions are storing data, making selected content accessible, and giving it no further thought once processed regarding sensitive material. While some processing may be involved before and during ingest to discover such data, as well as having negotiated agreements with donors in the event such material is found, it may not be enough. Manually searching material or even using built in operating system search functions is not enough for the discovery of sensitive data. Tools exist, freely available, easy to use, and extremely thorough. Tools such as bulk_extractor [5] and The Sleuth Kit (Autopsy) [6] can be introduced into workflows to significantly increase the discovery of sensitive information.

Without a thorough investigation, sensitive information may be sitting in storage that could potentially be problematic. It may be useful information, important and vital to a collection, revealing information that was previously unknown. Hypothetically, should a disk image be created from computing system belonging to a historical figure and the collection institution wants to discover as much about that figure as they can, forensically analyzing the system will reveal what could not be seen prior. Hobbies, interests, past-time activities, social groups, and much more can be discovered. While these forensic methods are typically used to discover questionable and illegal content, they can also be used to find the opposite. Both outcomes should be the objective of every collection institution as they may be holding information crucial to an on-going or previously dismissed criminal investigation, or it may simply reveal fascinating new information about an entity within their collection.

The way Indigenous content and people are treated should be the exemplar of how all content and people should be treated. Whilst the protocols differ from culture to culture, the example is that we should be considering all aspects, all scenarios, and all potential issues. By doing so and by following guidelines, preventive practices can be adopted, rather than dealing with issues as they unfold. Admittedly, issues such as those discussed may never surface, depending on what type of digital material an institution is dealing with. However, it is wise to be prepared, especially given that the future will be primarily digital and we do not know how it is going to change, in turn, changing digital preservation.

If we only concern ourselves with the laws to which we are bound and not those from which we are exempt, then it limits our potential to see future issues, hidden threats, best practices, and to generally consider what is best for people. There is never a one-size-fits-all solution, every issue is unique and every guideline must be applied in context. Being aware is the first step to being prepared for any issues or changes in law that may affect collection institutions. We have discussed the laws that are applicable, emphasizing how they may serve as guidelines, we also gave insight into the issues that can arise in collection institutions, providing further awareness of current and future threat potential. One cannot prepare for something of which you are unaware of and it is much better to prevent, than fix, making awareness something to strive for.
REFERENCES


