Privacy by Design: Embedding privacy processes and technologies into every aspect of a system design/organizational operation, its management and personal data sharing protocols.

MANAGING THE CREATION AND SHARING OF PERSONAL DATA

Privacy Context

All types of organizations (across both the public and private sectors) generate personal data/information to operate. Personal data is defined as data that relates to an identified/identifiable living individual. Occasionally when someone dies, there may be remaining personal data considerations, if for example, the data has a connection to someone else who is still living, as might be the case where the data relates to an inheritable genetic disorder.

Personal data is valuable to organizations but is also a potential burden, as there is a requirement for organizations to manage and protect this data. If personal data is breached and there are negative consequences for individuals, then an organization may suffer legal consequences, fines, and reputational damage. In 2019, Banisar reported that 130 countries around the world had adopted, comprehensive data protection and privacy laws, with almost 40 other jurisdictions with pending bills or initiative. In many pieces of legislation, it is recognized that some personal data has greater sensitivities and consequences for an individual if compromised. In European law this is termed “special category data," which is identified as data on racial or ethnic origin, political opinions, trade union membership, religious or philosophical beliefs, biometric data, genetic data, health data, sex life, and sexual orientation. In the USA, the FTC (Federal Trade Commission) has broad authority to enforce data protection regulations and protect data privacy, but there is no federal data privacy law or central data protection authority. Most regulations are at the state level, where there are differences in the definition of personal data. It is to be noted that in terms of data relating to an identifiable individual, it is becoming increasingly difficult to assume data is de-identified or anonymized. Today a mass of available/open data and searching mechanisms mean that it is increasingly possible to reidentify people.

Across the globe, a patchwork of legislation intersects with data protection and privacy laws, e.g., laws covering freedom of information, human rights, libel, security services, etc. The balances between differing rights are weighted differently by countries. For example, citizens in Canada and Europe traditionally place a higher value on privacy laws while countries such as the USA have traditionally placed a greater emphasis on information access. However, the picture is complex and constantly shifting. For example, some states in the USA have made the decision to ban facial recognition software. In addition, each nation is influenced by localized case law. Critically, where an archive is located in the world will influence its immediate operations. However, as archives increasingly share and store information globally, there is a need to engage with personal data considerations at an international level.

Privacy by Design Frameworks

All organizations must build frameworks that embed privacy protections into all their processes, i.e. "privacy by design". This concept of privacy by design was first advocated for by Ann Cavoukian when she was the Information and Privacy Commissioner in Ontario, Canada. Cavoukian (2009) set out seven foundational principles for privacy by design:

1. Proactive not reactive; preventive not remedial
2. Privacy as the default setting
3. Privacy embedded into design
4. Full functionality – positive-sum, not zero-sum
5. End-to-end security – full lifecycle protection
6. Visibility and transparency – keep it open
7. Respect for user privacy – keep it user-centric

Cavoukian's principles have been adopted internationally. In 2018, the EU's key personal data law, the General Data Protection Regulation (GDPR), required privacy by design systems that would embed privacy considerations in planning new systems and processes, and document and provide accountability for the management of personal data. These include providing "Records of Processing Activities" (ROPA) and also "Privacy Impact Assessments" (PIA) which evaluate risks implicit in the processing of personal data and steps to mitigate these. Archives must ensure that their own operational processes build in privacy by design. Archives will process personal data for a range of reasons. For example, personal data will be collected from donors/depositors, people undertaking research and people getting copies of items.

Around the world standardized processes (which in some nations are a legal requirement) have emerged as a means to manage personal data. These framework tools include:

- Privacy and information management policies,
- Privacy impact assessment (the latter will normally build in assessment of how to ensure that each of the further points below are addressed),
- Limiting personal data collection,
- Data/information asset inventories,
- Risk registers,
- Process mapping/records of processing activities,
- Retention and disposition schedules and processes,
- Security controls,
- Access controls,
- Confidentiality agreements,
- Subject access procedures,
- Data sharing and data processing agreements,
- Privacy notices,

- Transfer protocols including global protocols,
- Information breach and near miss reporting structures, and
- Monitoring and audits.

Archiving Personal Data

In addition to Archives managing their operational records, there will be personal data within the archival collections themselves which require careful management from the point of acquisition through to cataloging and access. For example, decisions need to be made about the archive collections that are accepted, the information on acquisition which is recorded, what is logged onto a public archive catalogue and when items with personal data can be made available for public access. Sometimes an Archive will make the decision to release a record with redacted (i.e. removed) personal data or will limit the details included in the catalogue, e.g., with court records where the names of victims of sex offenses may be important but not publicized.

It is important to note that privacy by design does not preclude organizations capturing and retaining information for archival purposes. However, due to the risks and responsibilities that personal data can pose, many organizations may decide to dispose of personal data as soon as it is no longer required for operational purposes. While this potentially minimizes the risk that the confidentiality of personal data will be compromised, it is at odds with capturing a holistic information picture of society and organizational operations for posterity. It is therefore critical that archivists actively engage with the development of key policies and procedures nationally and across sectors to advocate for the importance of archives and thus the retention of personal data worth preserving. Where archivists are embedded within organizations, they must engage with data protection policies, records management policies, as well as retention and destruction strategies if they are to ensure that personal data is retained for archival purposes. Potentially, personal data's full value in appraisal has not always been recognized. As new tools for accessing, using and linking data are emerging, the value of personal data within archival collections must be further recognized.

In legislation across the world, the "public good" in the maintenance of archives has been recognized, including in the context of data protection laws. For example, the EU's General Data Protection Regulation (GDPR) (EU, 2018) legislates for personal data to be retained without the consent of those it concerns for archival purposes even though this purpose will not necessarily have been envisaged at the point of creation. At a national level, GDPR allows EU nation-states to legislate for archiving within their national context within GDPR parameters; for example, the UK's position on archiving personal data was set out in the UK Data Protection Act 2018 Part 2 chapter 3; Schedule 1 Part 1 and Schedule 2 Part 6. Additional UK guidance was further published in 2018 (Lomas et al). GDPR allows certain exemptions from data protection requirements for personal data held within archival collections and processed for archiving purposes. One example is an exemption from the EU's "right to be forgotten," which enables people to ask for their data to be destroyed/deleted/purged in certain circumstances. This right

does not apply to accessioned archives processed in the public interest. However, exemptions such as this apply to the archival collections but not to the Archive's own operational record sets.

An Archive should provide transparency on its processes including how it manages personal data. Even when an Archive has not had explicit consent to process personal data, it can nevertheless be clear on its processes, for example through the publication of its policies and potentially a personal data privacy notice. A privacy notice should apply to the organization's own records but in addition, archives must ensure their own processes and procedures are covered.

It is important to be clear on the particular privacy legislation which applies in the Archive's national context , but potentially taking into account any implications for where the Archive undertakes transactions internationally. Digital archives may be accessed internationally and become subject to different legislative regimes although the local context remains critical.

Critically, Archives can use certain tools to help consider how they manage personal data, whether it is data within their archival collections or relates to their own operations. In particular these including PIAs and ROPAs.

Balancing Competing Considerations

Archives will need to consider having policies and processes in place to manage risk and appropriately protect personal data. This sometimes involves weighing competing considerations. For example, who is asking for the data and will denial of access to data have consequences. If personal data is made available, could it result in damage or distress being caused to an individual/or multiple parties? If so, what is the level of damage or distress that is determined could potentially result for an individual(s) (e.g. is it mildly annoying or life changing)? What is the potential impact of withholding information for other parties? What is the potential impact of the release or decision to withhold the information for the organization concerned? Each of these factors needs to be weighed. Archives will often have some overarching policies and processes to deal with particular contextual considerations. A general policy is often assigning a cut-off date after which it is assumed a person is likely to have died. For example, many Archives assume a 100-year lifespan. This means that if a person was an adult in the archival record, the calculation assumes that the person was 16 years old at the time, which allows the record to be released after 84 years. However, people are living to a greater age nowadays and the date may need to be extended, especially if more sensitive information is involved. At the moment many Archives may undertake only limited checks to see if they can identify whether someone is deceased, but as this is getting easier, this expectation is potentially shifting, particularly where a name is added to an archival catalogue. For example, it would particularly compromise someone to a greater degree if their sexuality were added to a globally accessible catalogue given that certain countries outlaw LGBT+ rights. As such, the personal data content, person and context must be considered. For prominent people, additional checks are perhaps becoming more normal.

Key tools that help document and weigh different considerations are the PIA and the ROPA. As we are moving into increasingly litigious societies, documenting decision-making is likely to become more critical for organizations. Sometimes researchers will be given managed access to otherwise closed information in order to control the risks but ensure that critical research is undertaken at an early stage.

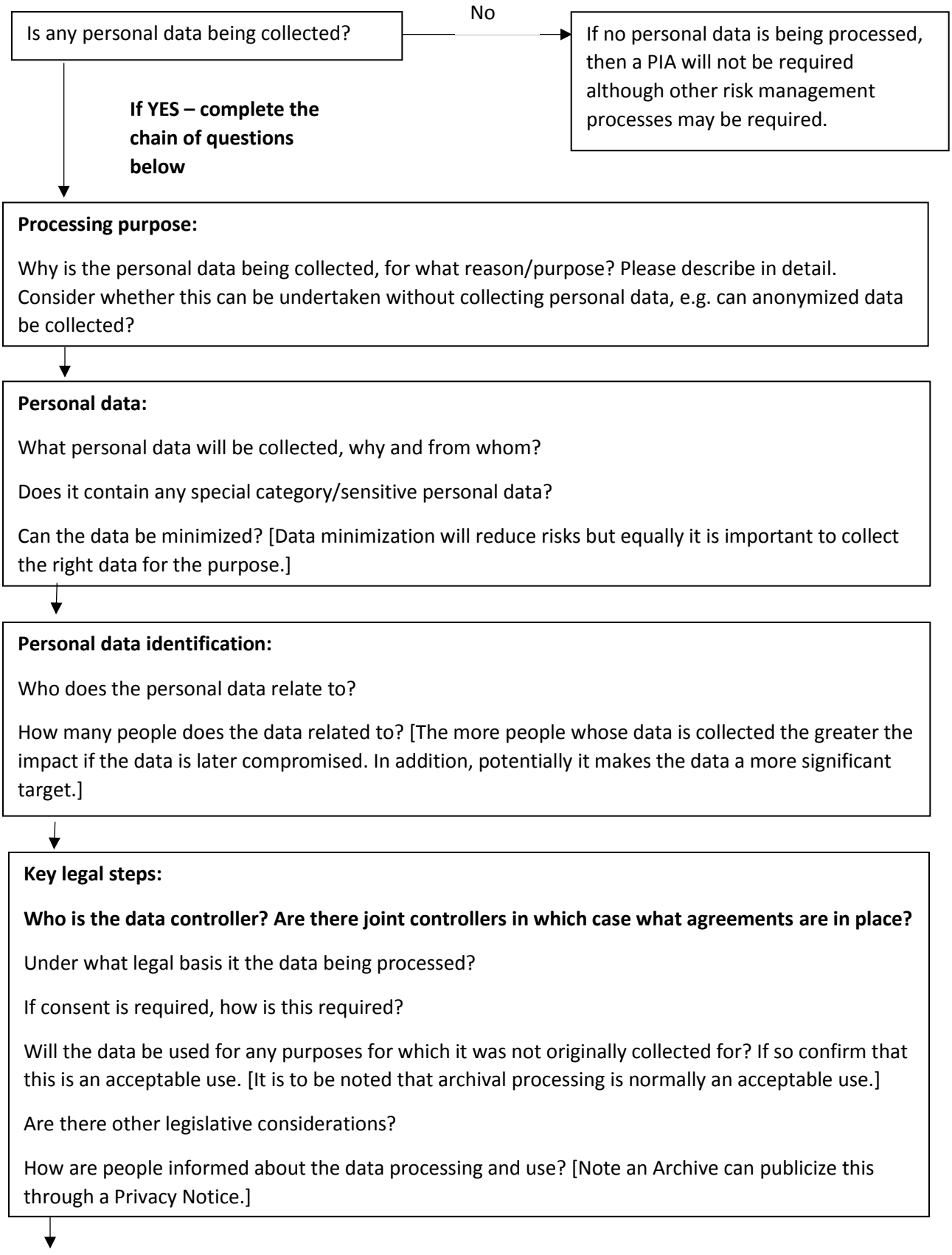
Privacy Impact Assessments (PIA)

Organizations need to build risk frameworks. One building block will often be the risk register. A risk register will determine an asset value and then the implications of that asset being compromised or not harnessed to access opportunities. This is another tool that assists an archive with managing risks. It would normally provide an overview of all organizational risks. It can consider key risks that relate to personal data management, including the implications of failing to protect personal data's confidentiality, integrity and availability. As such, the risk register does show how personal data risks form part of the bigger organizational risk picture.

A PIA is a very different risk tool from a risk register but will be another key part of the organization's risk management framework. The PIA will look in detail at the processing of personal data and any damage or distress that could be caused by inappropriate processing. It will describe the personal data under scrutiny, the purpose of processing, and how, why, where, and with whom it is shared. It will weigh the stakeholder considerations that need to be balanced. This involves assessing the potential risks to individuals and the measures that could be taken to mitigate any vulnerabilities. A PIA is recommended when an archive is implementing a new service or system, changing an aspect of the way it works, offering a service where there is a new legislative context, where a new archive is being acquired or where previously closed records are being opened. If any of these changes are planned and there are personal data considerations at play, then prior to the changes occurring, the PIA can assist with managing key steps and controlling the process appropriately. Figure 1 contains the key questions that need to be answered.

Critical data from a PIA can then be translated into a ROPA which can be monitored for changes. Where changes in the ROPA occur then the PIA should be revisited.

Figure 1: PIA Flowchart



Data processing:

Where and how will the personal data be stored?

Describe what security arrangements should be in place to protect the personal data?

How long will the data be retained? Are there any legal retention/accountability requirements?

How will the personal data be kept up-to-date and accurate?

How will the personal data be securely and completely deleted/purged/destroyed?

Data Access

Who will have access to the personal data and why including internal users, external user groups, contractors etc.?

What access controls/conditions will be in place?

What data sharing agreements are in place?

Additional data sharing/processing

Is the personal data shared externally?

Who receives the data?

What is the method of transfer?

For what purpose is it shared?

How is it secured by the recipient?

What is the legal basis for data sharing?

Is the data shared internationally? If so which countries and what is their legislative regime/requirements for managing personal data? Do the recipient countries have comparable personal data legislation in place?

If an aspect of the process is outsourced, what contractual requirements are in place (e.g. is there a data processing agreement in placement), how will the data be managed and secured, and who will have access.

Is a data sharing agreement in place?

Record of Processing Activity (ROPA)

In order to properly manage personal data, it is necessary to consider undertaking a process mapping exercise for each personal data asset considering what it is, how it is created and used through time including all those with whom it is shared. In addition, sometimes it is helpful to draw out a process map which helps visualize the flow of data. This information can then be developed into a Record of Processing Activity which in some countries is a legal requirement but in all contexts is beneficial in helping manage personal data.

Table 1 provides an example of a record of processing activity for an archival acquisition from an external source considering the accessions register data and associated files. Where the example data has been populated; it is assumed that this has been developed in an EU context where data protection legislation applies). It is to be noted that it is easier to manage this data in a spreadsheet or database.

Table 1: Example Archival Acquisition - Accessions Register and Files - Record of Processing Activity

ID	**PROCESSING QUESTIONS**	**PROCESSING [Example data has been populated]**
1.0	**Archival Acquisitions**	
	PROCESSING PURPOSE	
1.1	**WHY** is the personal data being collected, i.e. what is the overarching reason why you are processing this personal data?	To evidence the provenance and chain of custody of the acquisition, including proof or purchase, gift or loan, and to assist with ongoing management of the acquisition and holding rights and agreements.
	DATA SUBJECTS	
1.2	**WHO** are the data subjects whose personal data is being captured?	Archive/record creator's details. Donors, lenders and sellers, including individuals, families, employees, and purchasers that make the acquisition possible. Rights owners, including third parties. Rights may include copyright. Persons whose lives and activities are documented within the archival acquisition.
	INFORMATION ASSET OWNER	
	WHO is the information asset/personal data owner?	Archives Department
	WHAT type of personal data is collected?	Name, contact details, signature, rights, special requirements, relationships with creator, personal information about the creator and donor.
	WHAT personal data is collected that may need special consideration, e.g. is there any special category data/sensitive data, commercially confidential	A very wide range of personal data may be recorded with an acquisition, e.g. the contextual circumstances relating to a donor which may include health details. Further personal data may be taken from a donor/depositor/seller about the contents of the

	data, and legally privileged information?	archive being acquired [note some information may not be known until the acquisition is fully catalogued. Legal privilege [if for example there is information that relates to an ownership dispute].
	WHAT is the source of the personal data?	Donor/lender/seller
	WHAT is the legal context for processing the personal data?	Under the EU DP law the legal basis for processing the information is legitimate interests for a private archive or public interest for a public sector organization.
	WHEN and **HOW** is the personal data collected?	Ahead of the acquisition from the donor/lender/seller
	WHEN and **HOW** is the personal data captured?	Online system, file notes and signed acquisition forms and contracts
	How long is the data retained	Permanently
	LOCATION	
	WHERE is the personal data stored?	Acquisition file and CALM (stored on E Server)
	DATA SHARING and ACCESS	
	WHERE, WITH WHOM AND WHY is the personal data shared internally? If it is a global organization is any information shared across global boundaries?	Central registry
	WHERE AND WITH WHOM AND WHY is the personal data shared externally? Is any information shared across global boundaries?	With donor/lender/seller permission some basic details may be shared on the globally accessible catalogue. This provides important contextual information for researchers. Personal data from the archival collection itself will also be shared.

The Regulatory Context

Further reading on personal data and privacy by design is available at National Archives around the globe (e.g. Lomas et al, 2018). In addition, critical advice on legislation nationally and processes are available on the relevant regulators' sites. Regulators have a range of titles from Data Protection Registrars, Information Officers through to Privacy Commissioners. In a number of countries professionally qualified archivists have been assigned to these roles. This is an important space for the archive profession to engage with and influence to ensure a regulatory agenda continues to be delivered that recognizes the public interest in archival delivery through time.

~ Dr Elizabeth Lomas

References

Banisar, David, (2019) National Comprehensive Data Protection/Privacy Laws and Bills 2019 (November 30, 2019). Available at : https://ssrn.com/abstract=1951416 or http://dx.doi.org/10.2139/ssrn.1951416 (Accessed 27 June 2020)

Cavoukian, Ann. (2009) 7 Foundational Principles. Information and Privacy Commissioner of Ontario: Toronto. Available at: https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf (Accessed 27 June 2020).

Deloitte and Ryerson University. (n.d.) Privacy by Design: Setting a New Standard for Privacy Certification. Accessed July 28, 2020, https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-ers-privacy-by-design-brochure.PDF

EU (2016) The General Data Protection Regulation. 14 April 2016. Available at: http://www.eugdpr.org/ (Accessed 27 June 2020).

Lomas EJ, Abraham S, Todd M, Sexton A, Mitchell L, Sim mons J, Ellis M, Horton S, Huddleston D, Hutchinson J, Elliott J, McDonnell N, and Healy S. (2018) Guide to archiving personal data. The National Archives: London.
Available at: https://www.nationalarchives.gov.uk/documents/information-management/guide-to-archiving-personal-data.pdf (Accessed 27 June 2020).