**Summary of Legislations:**

**Privacy Act 1988**

# The Privacy Act 1988 is an Australian law dealing with privacy. It was introduced to promote and protect the privacy of individuals and to regulate how Australian Government agencies and organisations with an annual turnover of more than $3 million, and others, handle personal information.

**GDPR**

The General Data Protection Regulation (EU) 2016/679 (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA areas. The GDPR's primary aim is to give individuals control over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU. Superseding the Data Protection Directive 95/46/EC, the regulation contains provisions and requirements related to the processing of personal data of individuals (formally called *data subjects* in the GDPR) who are located in the EEA, and applies to any enterprise—regardless of its location and the data subjects' citizenship or residence—that is processing the personal information of individuals inside the EEA.

**HIPAA**

The Health Insurance Portability and Accountability Act of 1996 is a United States federal statute. It was created primarily to modernize the flow of healthcare information, stipulate how personally identifiable information maintained by the healthcare and healthcare insurance industries should be protected from fraud and theft, and address limitations on healthcare insurance coverage.

**FDA Part 11**

* FDA 21 CFR Part 11 compliance dictates that those companies who use electronic systems for document and signature control must provide assurance that the electronic documents are authentic. The regulations all stipulate the necessity of the confidentiality of electronic records.
* FDA 21 CFR Part 11 compliance dictates that all system users have the training necessary to perform their assigned tasks and projects.
* FDA 21 CFR Part 11 compliance dictates that electronic records that are signed must contain a name, the signature meaning and the date/time of signing.
* FDA 21 CFR Part 11 compliance dictates that signatures whether electronic or handwritten be linked to their respective records.
* 21 CFR compliance dictates that an FDA-regulated company's electronic system must be validated according to the FDA's validation standards.

**Summary of “Ten Simple Rules for Digital Data Storage”**

**Introduction**

* The nature of scientific data has changed dramatically with the rapid pace of technological progression. This, in-turn, has led to the need for increased coordination amongst data collectors and institutions.
* “data” can now mean “anything from petabytes of information stored in professionally maintained databases, to spreadsheets on a single computer, to handwritten tables in lab notebooks on shelves.”
* “All remain important, but data curation practices must continue to keep pace with the changes brought about by new forms of data and new data collection and storage practices.”
* Proper storage of data is a pre-requisite to data sharing.

**Rule 1: Anticipate How Your Data Will Be Used**

* Have a clear roadmap of what to expect before acquisition starts:
  + How will the raw data be received? Are they delivered by a machine or software, or typed in?
  + What is the format expected by the software used for analysis?
  + Is there a community standard format for this type of data?
  + How much data will be collected, and over what period of time?
* Estimate storage volume needed to store the data.
* Have a plan for what metadata will be collected and how it will be maintained.

Rule 2: Know Your Use Case

* Well defined use cases make data storage easier:
  + Should the raw data be archived (Rule 3)?
  + Should the data used for analysis be prepared once or re-generated from the raw data each time (and what difference would this choice make for storage, computing requirements, and reproducibility)?
  + Can manual corrections be avoided in favor of programmatic or self-documenting approaches (e.g., Jupyter notebook or R markdown)?
  + How will changes to the data be tracked, and where will these tracked changes be logged?
  + Will the final data be released, and if so, in what format?
  + Are there restrictions or privacy concerns associated with the data (e.g., survey results with personally identifiable information [PII], threatened species, or confidential business information)?
  + Will institutional validation be required prior to releasing the data?
  + Does the funding agency mandate data deposition in a publicly available archive, and if so, when, where, and under what license?
  + Does the target journal mandate data deposition?
* Knowing the what, when and how of the data’s use will make creating a reliable roadmap easier.

Rule 3: Keep Raw Data Raw

* If only derived data are stored, it can be difficult for other researchers to confirm analytical results, assess the validity of statistical models, or directly compare findings across studies.
* Therefore, data should always be kept in raw format whenever possible. This ensures transparency in analysis; having the data stored and archived in their original state gives a common point of reference for derivative analyses.
* The spirit of this rule is that data should be as ‘pure’ as possible when stored. Derivations should be documented.

Rule 4: Store Data in Open Formats

* Maximises accessibility and long-term value; do this by storing data in formats that have freely available specifications.
* The key idea is that accessing data should not require proprietary software, hardware, or a license.
* Examples of open data formats include comma-separated values (CSV) for tabular data, hierarchical data format (HDF) and NetCDF for hierarchically structured scientific data, portable network graphics (PNG) for images, KML (or other Open Geospatial Consortium [OGC] format) for spatial data, and extensible markup language (XML) for documents.

Rule 5: Data Should Be Structured for Analysis

* For data to make sense, it should be structured in a way that makes use, interpretation, and analysis easy.
* Axiom: “write code for humans, write data for computers”.
* Interoperability is facilitated when variable names are mapped to existing data standards.

Rule 6: Data Should Be Uniquely Identifiable

* For the purpose of reproducibility, data used in a scientific publication should be uniquely identifiable.
* Unique identifiers could be Digital Object Identifier (DOI), Archival Resource Key (ARK), or a persistent URL (PURL).
* Even as identifier standards may change over time, datasets can evolve over time as well.

**From the above, how should personal information be taken care of?**