

Data science applications

Notes



July 19, 2024

Cherrie liang

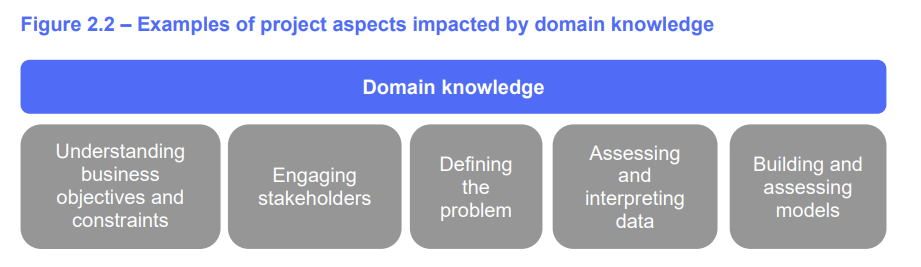
# Chapter 2: Domain Knowledge

**1. Assess the impact of complex business environments on all stages of a data science project**

**1.1. Evaluate the impact domain knowledge has on the success of a data science project**

**Domain Knowledge** includes an understanding of a domain’s: customers and other stakeholders, processes, policies, constraints, terminology (i.e. jargon) and abbreviations and trends.

* + 1. **Assess the influence of domain knowledge on all stages of a data science project**

****

**Business Objectives and Constraints:** knowledge of objectives will help knowing relevant parties, solve problems, understand the data generation process and allow for constraints.

**Engaging Stakeholders:** understand which metric is important to your stakeholders to ensure delivery of best outcome. Best communication to audience: these factors include your audience’s size, personalities, education, status, detail required, cultural background and customs, perspectives and prejudices, and context.

**Defining the Problem:** usually spend more time defining the problem than search for solution. Sometimes you might be searching solution for the wrong problem. Ask 5 whys, consider what you know on the business context, ask clarifying questions, consider restating the problem.

**Assessing and Interpreting Data:** knowledge of the data creation process and any changes (who create data, human or machine, is field optional), are there other aspects (such as below) that will affect the data interpretation.



**Building and assessing models:** using domain knowledge one can assess whether the existing model variables will achieve the best outcome, whether new data needs to be collected to build the model, if there will be information leakage.

* + 1. **Apply appropriate techniques to acquire domain knowledge**

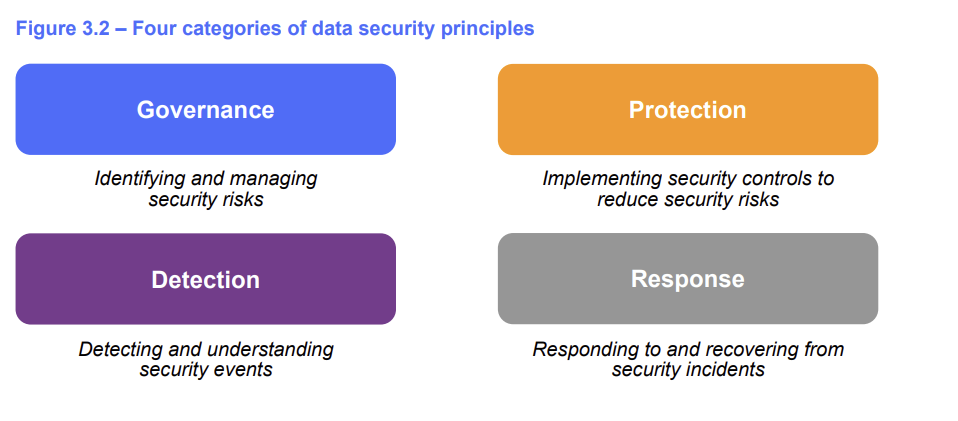
**Techniques include:**

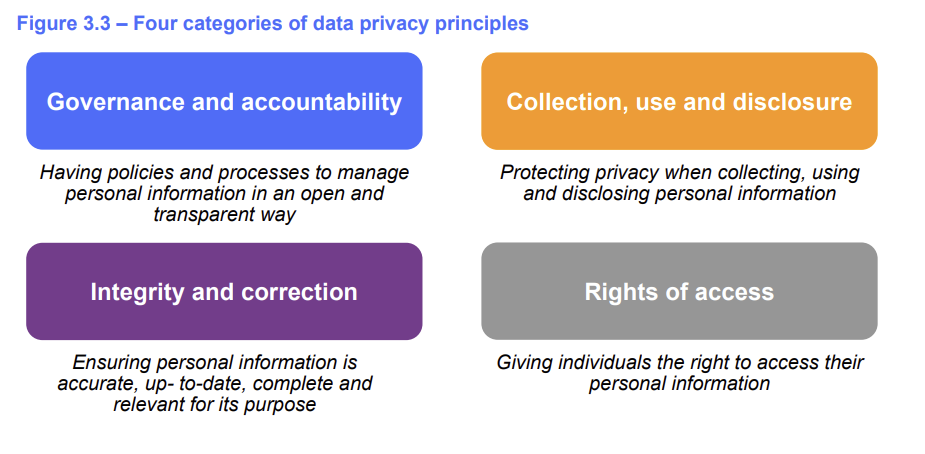
* **Plan,** keep a notebook on the topic to record the question answers, focus on what you need to know
* **Conduct initial research:** read relevant books, videos podcasts, news and case studies etc to acquire basic knowledge
* **Ask questions:** ask experts on areas you do not understand, choose the who to talk to wisely, such as people who manage/run key processes, data team if there is one, people who manage relationship with stakeholders.
* **Build a knowledge repository:** save source PDF etc to keep a record for future reference
* **Join a knowledge community**
* **Enrol in a course**
* **Walk in the shoes of users**
* **Ensure you have the right mix of people in your team.** So, each area can have an expert including business and technical processes, policies, constraints, risks and key stakeholders.

# Chapter 3: Security, privacy and ethics

**1.2. Assess the impact of security and privacy considerations on data science work**

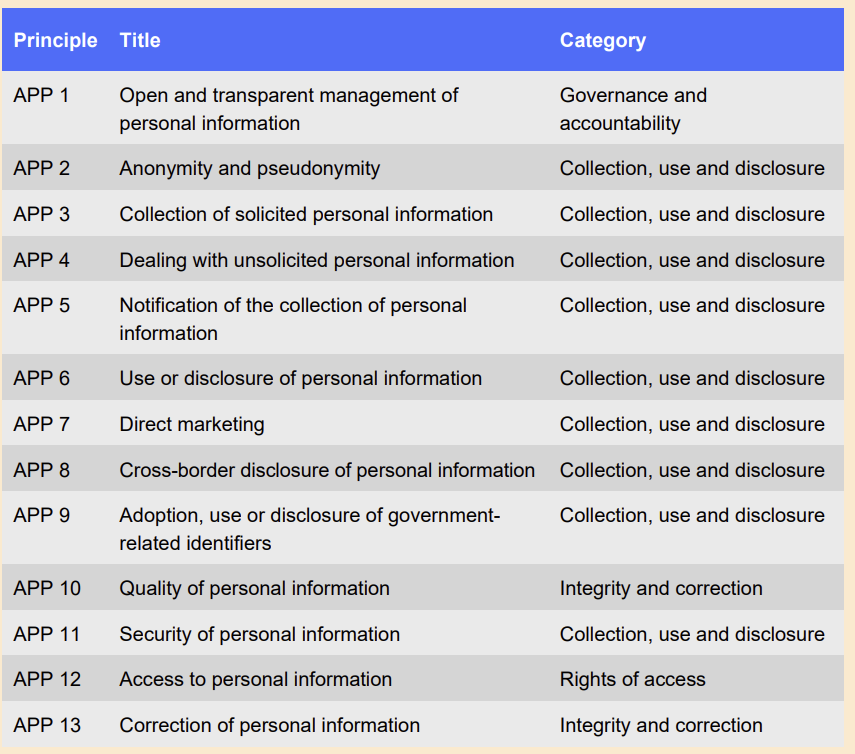
**1.2.1. Explain the principles that underpin an organisation’s data security and privacy processes**

****

****

**1.2.2. Outline the Australian security and privacy laws and regulations that apply to data science work**

**- Australian Privacy Pronciples (APP)** – apply to gov, rev > $3m, small business in healthcare or sell personal information – 13 rules



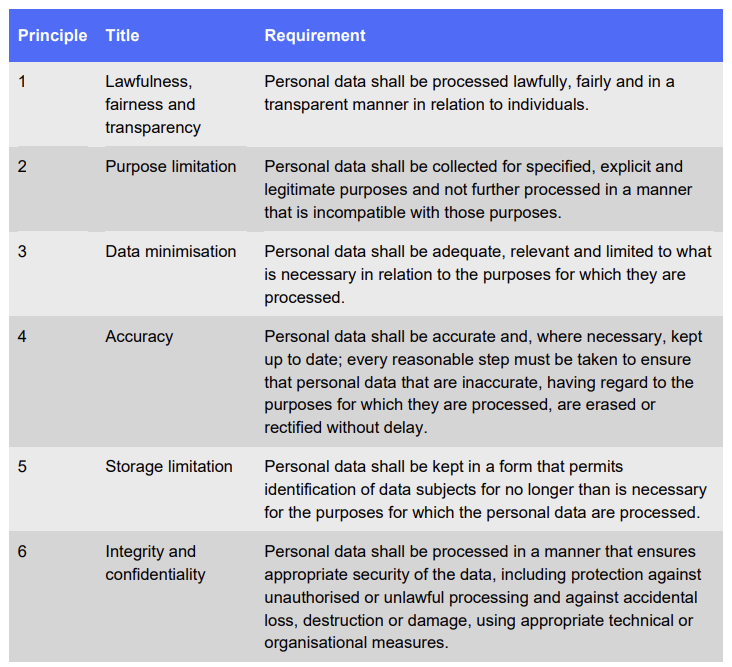
- CPS234 – information security

- CPG235 – managing data risk

- Consumer Data Right (CDR) rules - allow individual and business to efficiently access data about them kept by other businesses and the government.

**1.2.3. Contrast the security and privacy rules that relate to data science work in Australia with those that apply in the European Union and the United States of America**

**- EU** – 6 general data protection regulation principles, on eon accuracy, remaining on collection, use and disclosure of data.

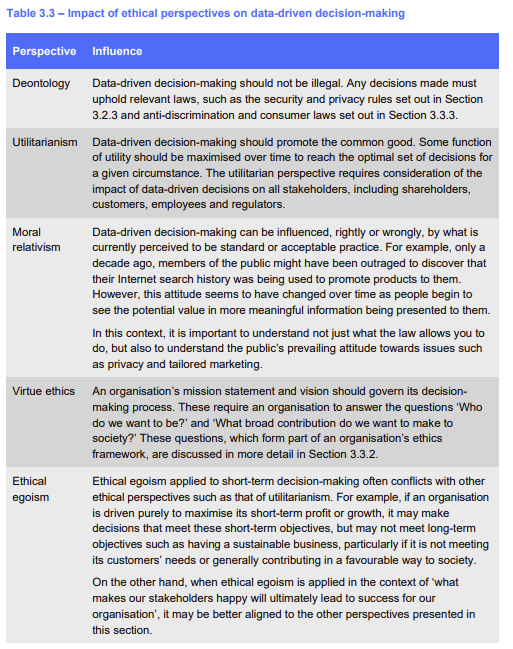


- **USA** – California Consumer Privacy Act – only applicable to California but is used as a benchmark for other states.

- Payment card industry data security standards.

**1.3. Apply ethical principles to data science work**

**1.3.1. Evaluate various ethical perspectives that influence data-driven decision-making**

****

**1.3.2. Examine an ethical framework that drives a strong data ethics culture**

Ethics framework is essentially asking 3 questions: who should we be, what outcome is best, what should we do. These translates into 3 items: purpose, values and principles.

Purpose is usually identified in organisations mission statement. Values are objectives and constraints. Principles give guidance to employees on decision making to achieve the purpose and values. Principles can include:

* Ought before can
* Non-instrumentalism
* Self-determination
* Accessibility
* Fit for purpose
* Net benefit
* Fairness
* Responsibility and
* Accountability

**1.3.3. Outline the ethical guidelines that apply to data science work in different jurisdictions**

**- Australian rules:**

- Australian Privacy Act 1988

- anti-discrimination law

- consumer laws (Corporations Act 2001)

- Human right agreements

- AI ethics principles

# Chapter 4: Data Structure

# Chapter 5: Classification

# Chapter 6: Unsupervised learning

# Chapter 7: Natural language processing

# Chapter 8: Business Optimisation

# Chapter 9: Implementation change