

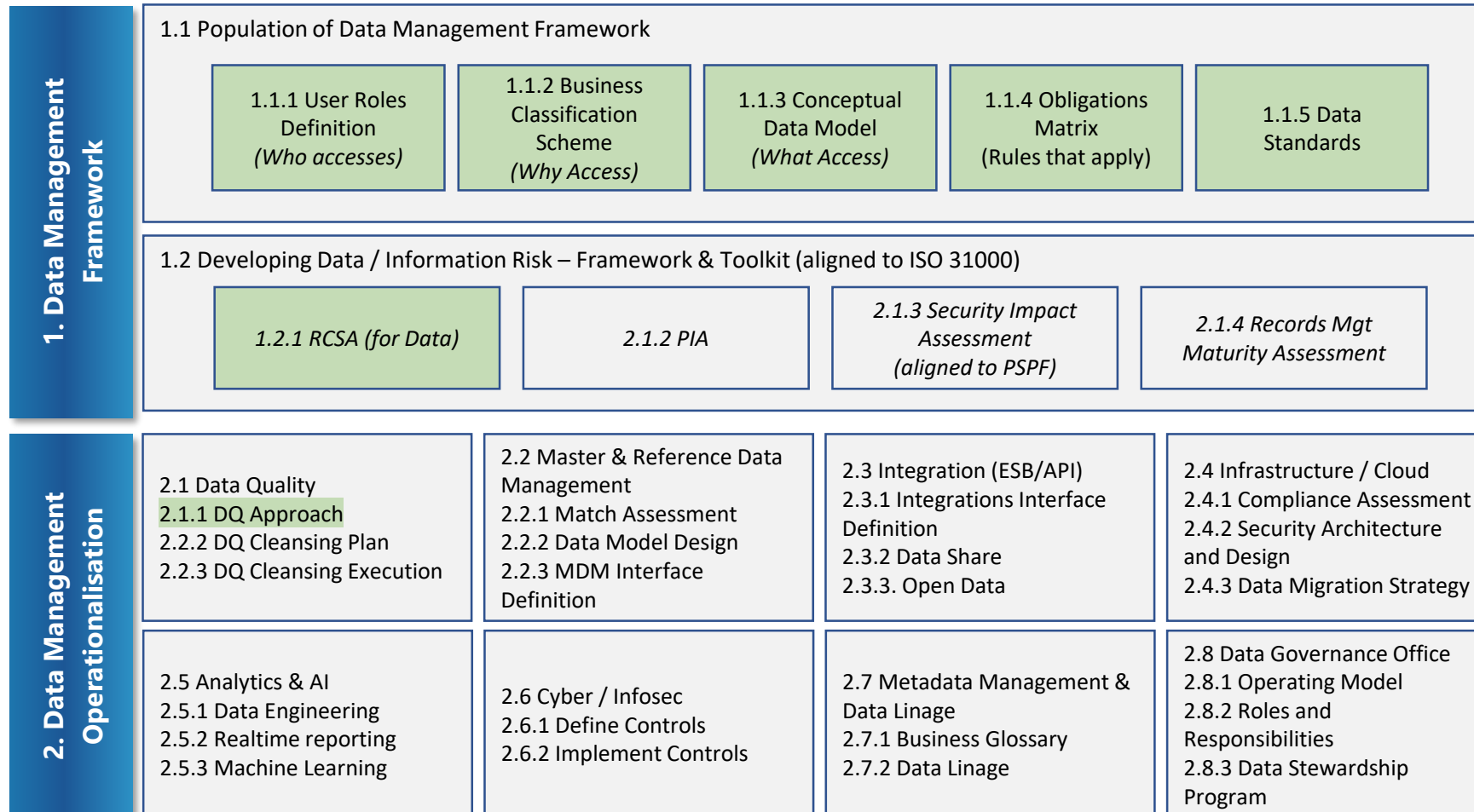
Adapting Information Management for the Open Data Economy


Cognitivo Consulting

March 2019

Data Management Services

Cognitivo has a comprehensive catalogue of services and experience establishing and operationalising data management functions within organisations



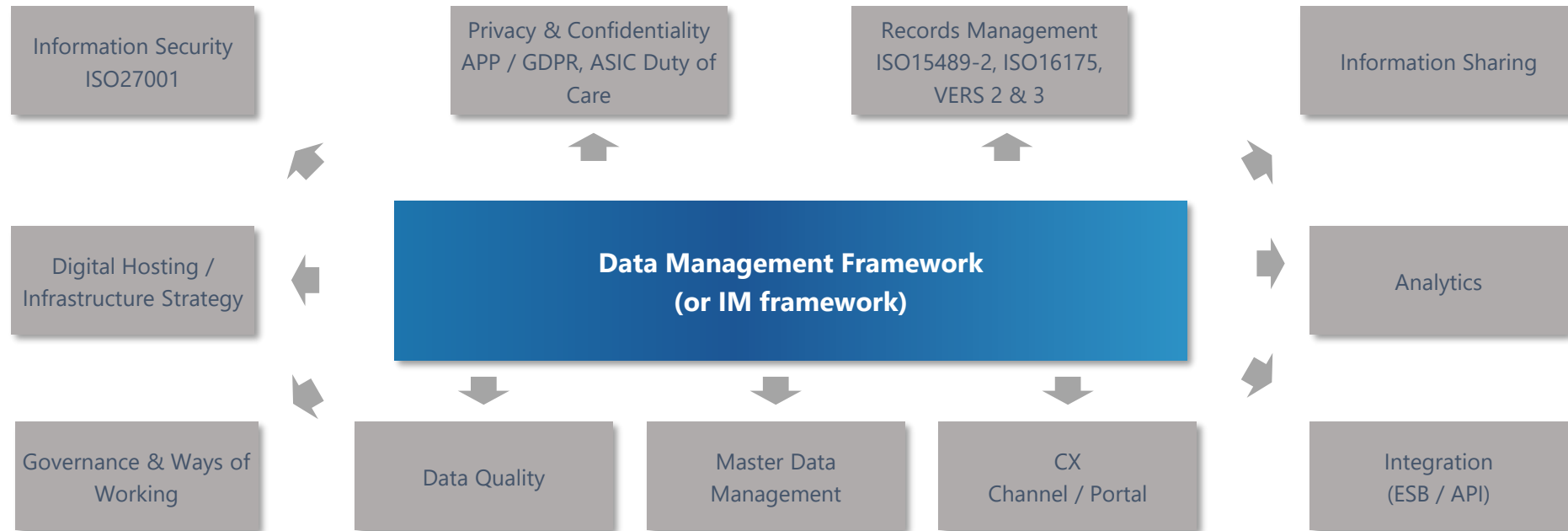
 Recommended scope of initial engagement for organisations looking to establish a Data Management Capability

Our Approach

- Two phase approach:
 - Phase 1: Building foundational/core data management capabilities
 - Phase 2: Operationalising data management processes
- The phase 1 ensures population of data management framework with the organisation specific roles definition, data access requirements, the rules/obligations around data and the information risk management toolkit.
- The phase 2 helps organisations to operationalise their data management activities, and maximise value from their data.

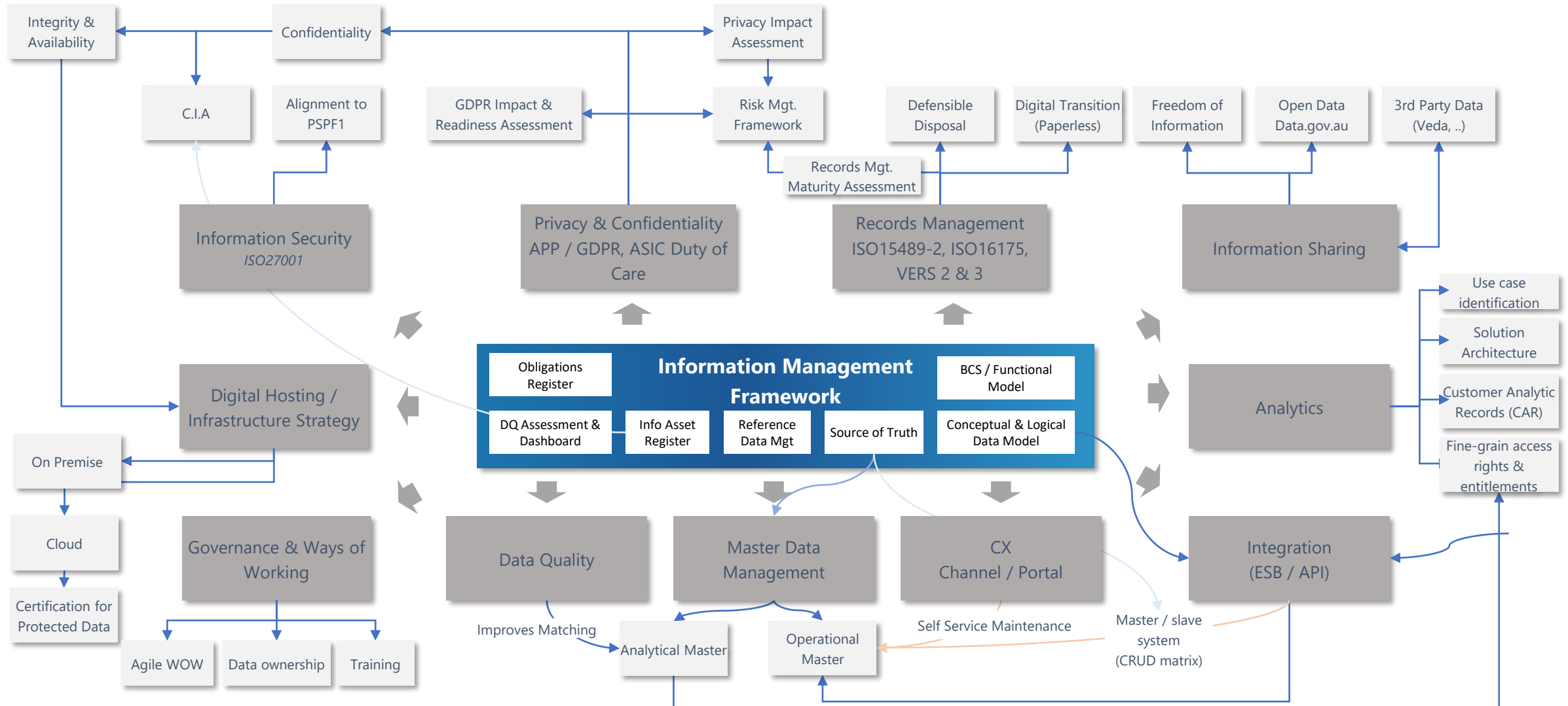
Data Management is a crucial enabler to digital transformation

Data Management is the connective tissue between multiple business and technology objectives & disciplines



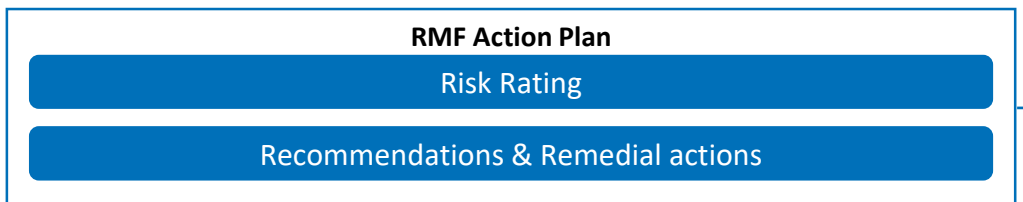
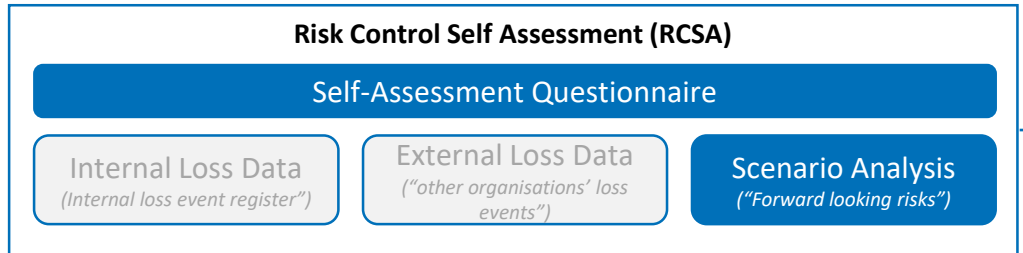
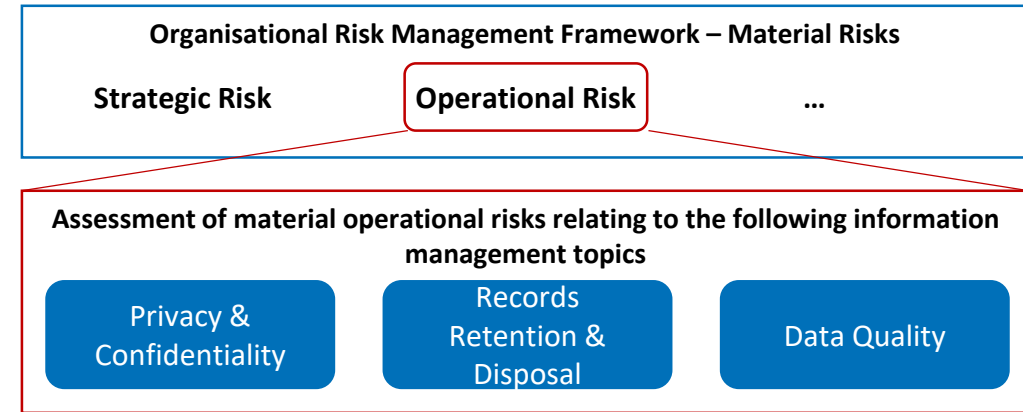
However a broad range of tools and practices must be aligned

Cognitivo has experience in connecting the relevant tools and practices to enable implementation of Information Management Framework



Our framework is aligned to operational risk methodology

We have a systematic risk control self assessment(RCSA) tool and approach for effective operational risk methodology implementation



Risk Management Framework – w.r.t Operational Risk Management

- Operational risks are linked to the Business Plan objectives and take into consideration risks which will prevent organisations from delivering their business plans and ongoing services.
- Risk assessment is undertaken in accordance with this Framework to determine the risks in meeting its obligations and objectives.
- Scope of this workstream includes the assessment of information / data risk which we deem to be a type of operational risk.

RCSA provides a systematic means of identifying control gaps that threaten the achievement of defined business or process objectives and monitoring what management is doing to close these gaps.

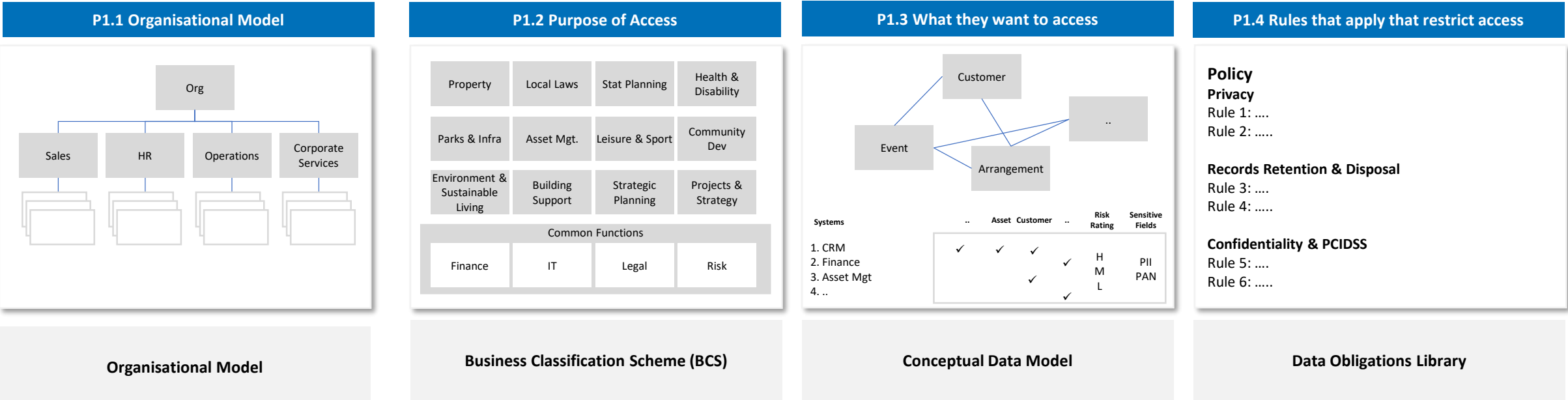
A typical RCSA incorporates:

- Identification of business objectives
- Identification of risks that could threaten the achievement of those objectives
- Identify controls to prevent risks events from occurring
- Determine responsibility and ownership of those controls
- An assessment of the effectiveness of controls in operation and level of residual risk (qualitative assessment within this phase)

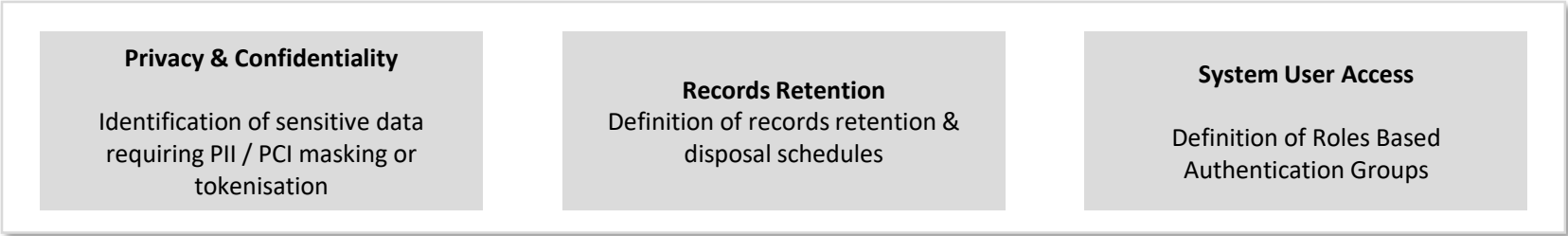
- Organisations operational risk profile to be maintained.
- Periodic review of Organisations operational risk registers are undertaken and reported.
- Management oversight of remediation progress.

Unified Data Classification Approach

Cognitivo’s unified data classification approach can drive control requirements for Privacy, Records & User Access

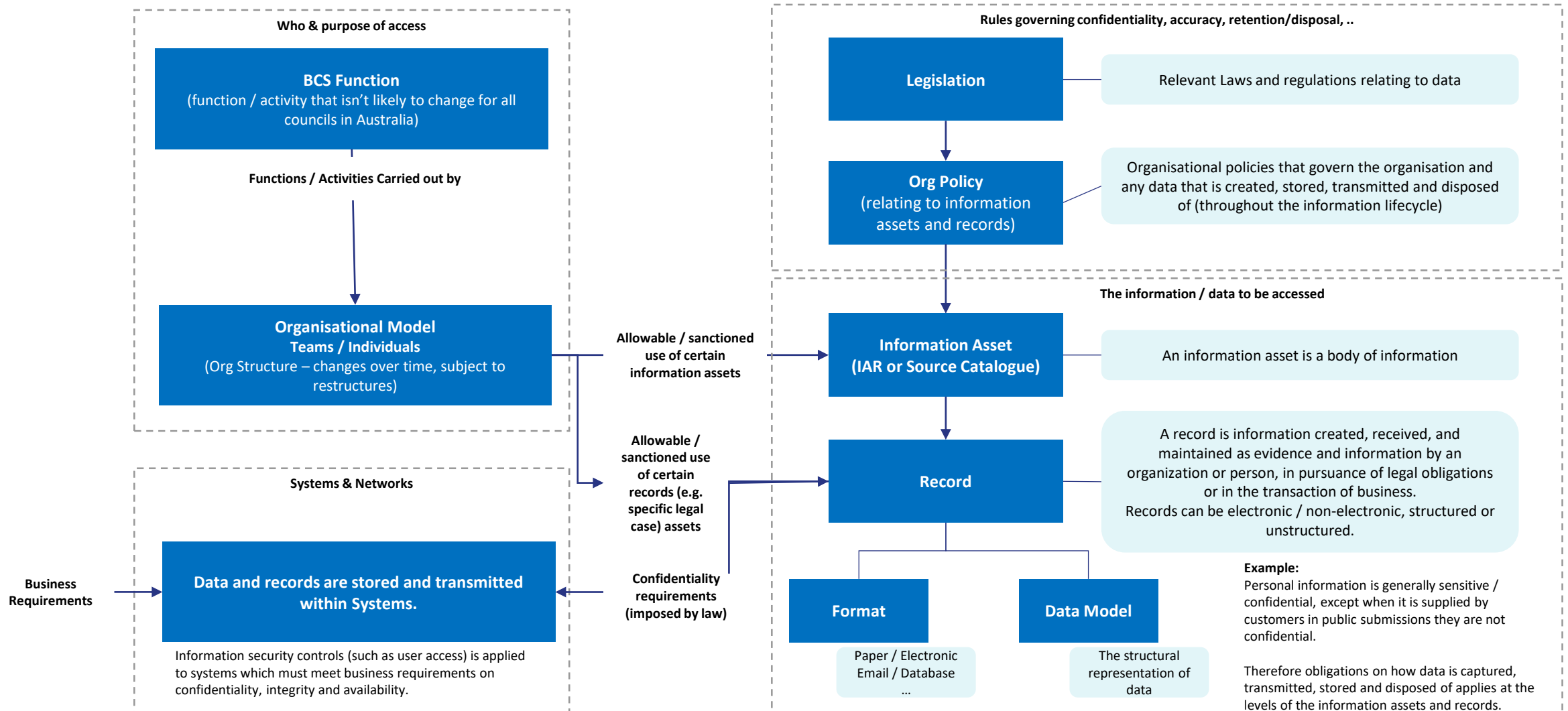


Drives Information Lifecycle Management & Data Risk Controls



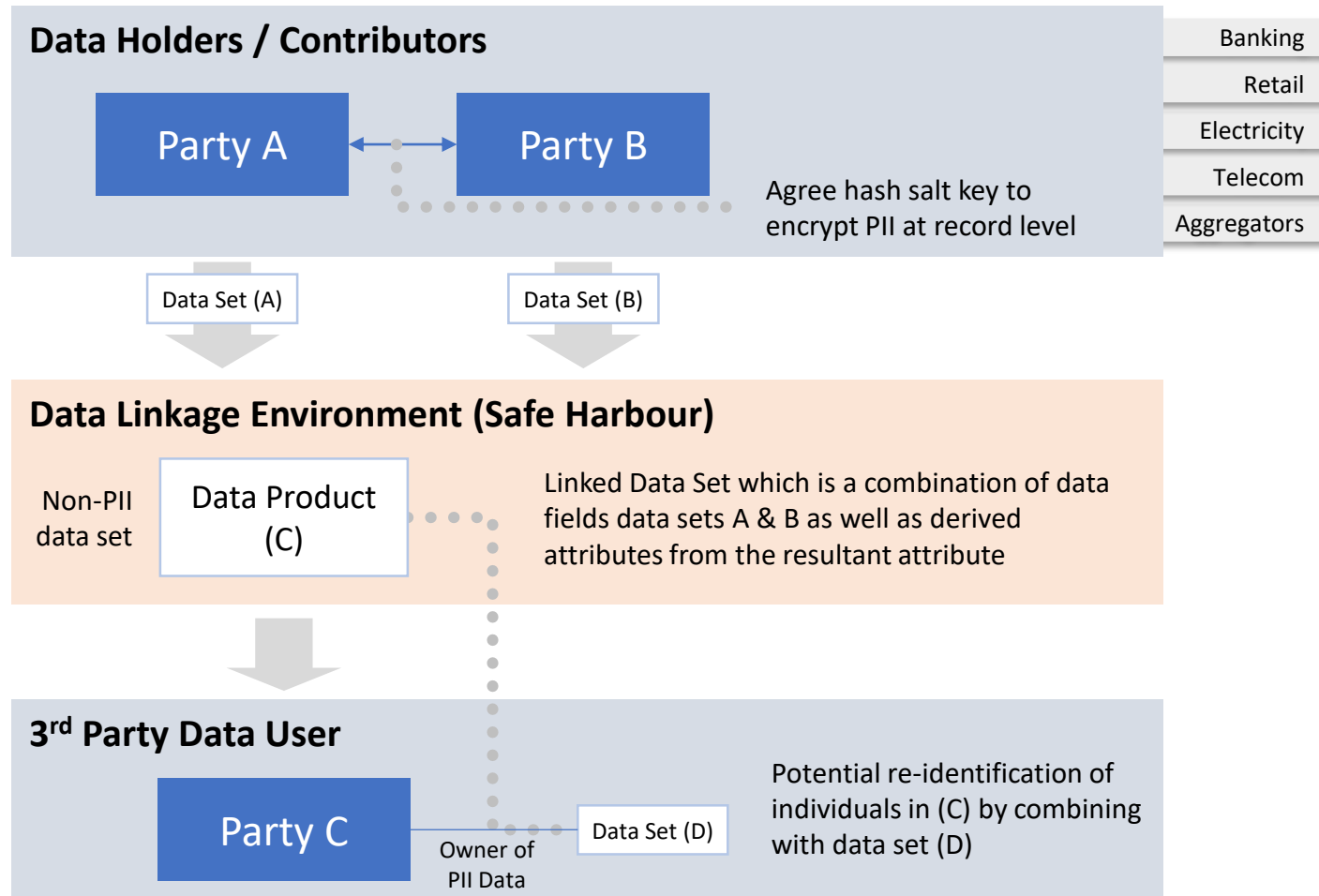
This alignment links Data Mgt to existing assets & controls

Our framework provides a holistic view on how the business and compliance requirements impacts the organisation information assets, and the controls required to be in placed



New risks in the open data / data sharing economy

Data linkage will introduce new data risks associated with re-identification, in the data sharing world, user access controls will not help, new statistical disclosure risk measurement and avoidance controls will need to be implemented



- The purpose of the Statistical Disclosure Avoidance (SDA) framework is support the protection of individual privacy by testing and measuring the level of risk of re-identification of previously de-identified data
- Current methods use rules-based controls on restricted (PII/PAN/HIPA) data fields and minimum cohort sizes
- E.g. between 2-3pm there were only 5 pensioners tapping their opal card, a bank may have enough demographic data to re-identify these individuals
- Furthermore, testing of outbound data sets that vary in cohort size for personal information risk is rarely (if ever) undertaken

No method currently exists to quantify re-identification risk

Organisations need to begin with qualitative methods to size data sharing risks that can be linked to existing tools in the Information / Data management tool kit

High Level Approach

1

Framework Definition

- Data Classification (document sensitive data fields)
- Define SDA methodology & guidelines (aligned to organisation’s Risk Management Framework for risk levels)
- Development of minimum criteria such as minimum identifiable cohort sizes for different data concepts (e.g. consumer, businesses etc.)

2

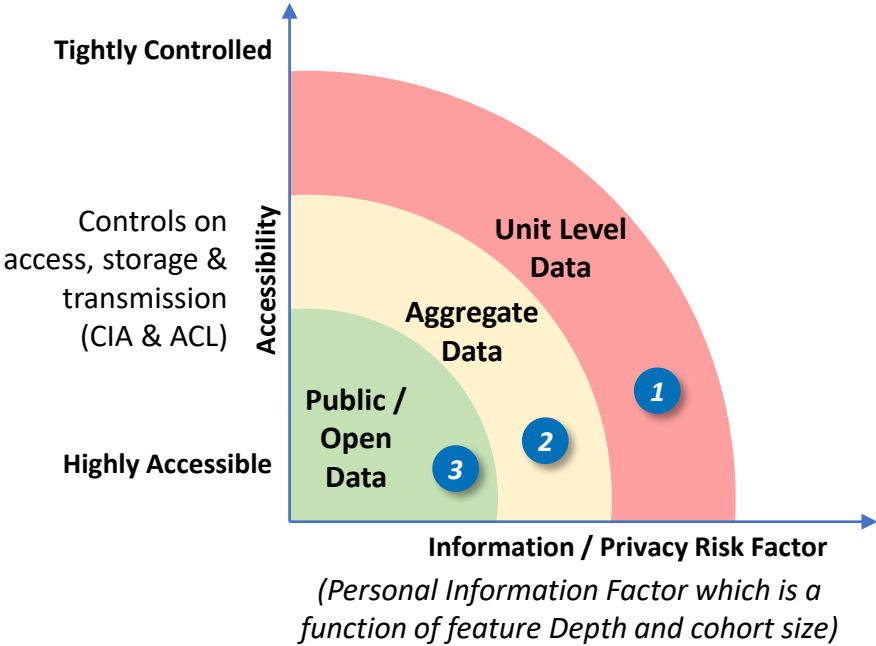
Define Data Sets / Products

- Define outbound datasets (data sets to be shared externally) and record these data sets within the organisation’s information asset register

3

Testing & Monitoring

- Static testing of data fields based on specifications (looking at sensitive data fields and indicative cohort sizes)
- Deploy machine-based and online monitoring capability attached to outbound API’s.
- Online data risk monitoring is capable of aggregating data across streaming data sets to determine privacy risk



Sample Information Asset Register

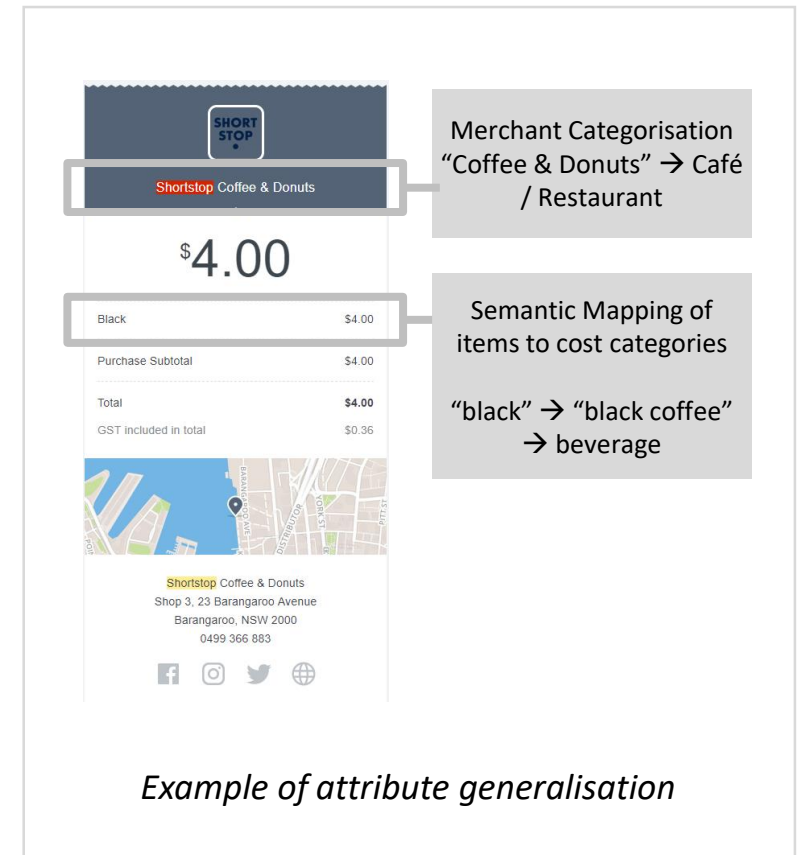
#	Information Asset (Data Set)	Description	Permitted Users & Purpose	Custodian	Confidentiality	PII	Health	PAN/PCI	Retention
1	Monthly Credit Card Spend in NSW	Transactional Level spend data for NSW	Business Insights	Head of Unsecured Loans	Highly Confidential	No	No	Masked PAN	7 years
2	New mortgages by state	Regulatory Report	APRA, RBA, ABS for economic planning	CRO	Confidential	No	No	No	20 years
3	Product Reference Data	List of all products, features, fees etc.	Public consumption	CMO	Public	No	No	No	10 years

Privacy Preservation Techniques

A number of privacy preservation techniques are available, what to use in which scenario must be driven by the Data Catalogue

A selection of Privacy preservation techniques:

1. **Minimum cohort sizes** – The most basic, but ensure data sets are sufficiently deep (row counts) so that there are no unique outliers
2. **Tokenisation & Anonymisation** – Substituting a sensitive element with a non-sensitive equivalent (a token), this is a technique widely adopted to protect payment card numbers (PCI DSS) however given token's are unique, this technique offers little protection against 3rd party reidentification
3. **Redaction** – Omission of data fields / attributes is effective but reduced the utility of data sets
4. **Generalisation** – By provided a summary or rolled-up version of the data element, for example, instead of address use the ABS's statistical area definitions, or a product category instead of a purchase item
5. **Differential Privacy** – A process of actively measuring and modifying queries in databases in order to maximise the statistical integrity of a query while protecting privacy of individual records.
6. **Bring the model to the data** – Secure data vaults or safe harbours where data science models can be executed securely. However this increases consequences associated with cyber attacks and insider data breaches.



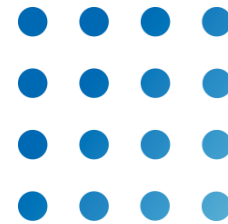
Contact Information



contact@cognitivo.com.au



www.cognitivo.com.au



cognitivo
CONSULTING