









University of
South Australia

Big Data Quality and Governance

1

The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume, variety and velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					

ICONS: ALEXANDRO/ADORE STOCK

©2018 TECHTARGET. ALL RIGHTS RESERVED. 

2

Data quality dimensions



Accuracy



Completeness



Consistency



Timeliness



Validity



Uniqueness

How good is your data?

2019 Enterprise Data Quality survey by *Precisely*

Only 8% of respondents reported **excellent** data quality

Less than 50% of take advantage of a data **profiling tool**
or a central data catalogue

Less than 60% of data **well understood** by stakeholders



Challenges with ensuring high quality data

2019 Enterprise Data Quality survey by *Precisely*

Many sources of data **(70%)**

Applying governance processes to manage and measure data quality **(50%)**

Volume of data **(48%)**



Consequences of poor quality data

2019 Enterprise Data Quality survey by *Precisely*

Wasted time **(92%)**

Ineffective business decisions **(72%)**

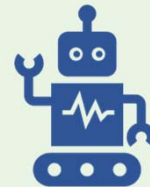
Customer dissatisfaction **(67%)**



Impact on business decisions

2019 Enterprise Data Quality survey by *Precisely*

Only 16% are confident they aren't feeding bad data into **AI and ML applications**



7

Data Quality Tools

- Gartner report, ID G00363493, published 27 March 2019



8

Data Governance – definition

- Data governance is the specification of **decision rights** and an **accountability framework** to ensure the appropriate behaviour in the valuation, creation, consumption and control of data and analytics.



Big Data Governance – benefits

Better data security and compliance

Improved data quality

Accelerated automation

Data Governance – roles



Chief Data Officer



Data owners



Data stewards

Data Governance – framework



Data standards

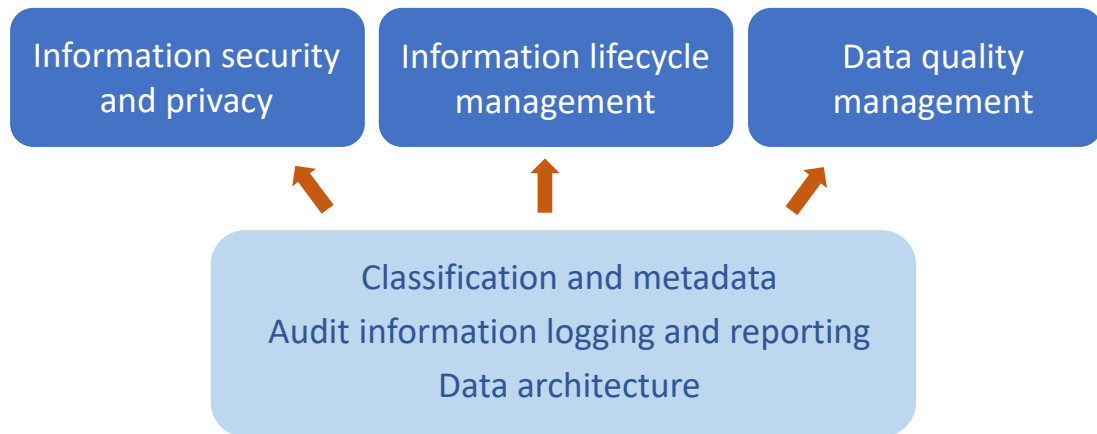


Data processes
Organisational
structure



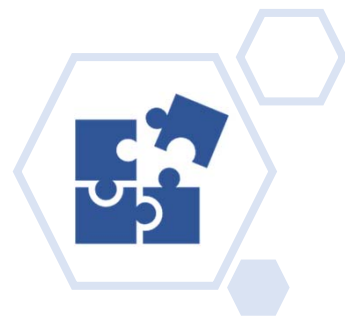
Technology

Big Data Governance – core disciplines



Big Data Governance – challenges

- Unstructured data – messy and moves fast
- Dynamic, multi-cloud environments
- Curating what's in a data lake
- Data ROT – need to clean up
- Supporting people with technology



Apache Atlas

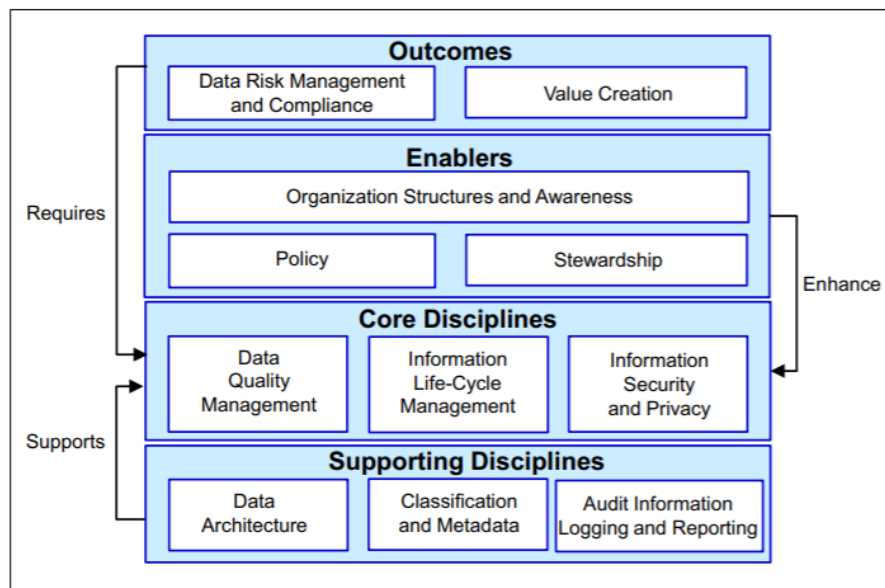


Figure 2-1 IBM Information Governance Capability Maturity Model

WARNING

This material has been reproduced and communicated to you by or on behalf of the **University of South Australia** in accordance with section 113P of the *Copyright Act 1968 (Act)*.

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice