



APRA

Prudential Practice Guide

CPG 235 – Managing Data Risk

September 2013

Disclaimer and copyright

This prudential practice guide is not legal advice and users are encouraged to obtain professional advice about the application of any legislation or prudential standard relevant to their particular circumstances and to exercise their own skill and care in relation to any material contained in this guide.

APRA disclaims any liability for any loss or damage arising out of any use of this prudential practice guide.

© Australian Prudential Regulation Authority (APRA)

This work is licensed under the Creative Commons Attribution 3.0 Australia Licence (CCBY 3.0).

 This licence allows you to copy, distribute and adapt this work, provided you attribute the work and do not suggest that APRA endorses you or your work. To view a full copy of the terms of this licence, visit www.creativecommons.org/licenses/by/3.0/au/.

About this guide

Prudential practice guides (PPGs) provide guidance on APRA's view of sound practice in particular areas. PPGs frequently discuss legal requirements from legislation, regulations or APRA's prudential standards, but do not themselves create enforceable requirements.

This PPG aims to assist regulated entities in managing data risk. It is designed to provide guidance to senior management, risk management and technical specialists (both management and operational).

The PPG targets areas where APRA continues to identify weaknesses as part of its ongoing supervisory activities. The PPG does not seek to provide an all-encompassing framework, or to replace or endorse existing industry standards and guidelines.

Subject to meeting APRA's prudential requirements, a regulated entity has the flexibility to manage data risk in a manner that is best suited to achieving its business objectives. Not all of the practices outlined in this PPG will be relevant for every regulated entity and some aspects may vary depending upon the size, complexity and risk profile of the entity.

Contents

| | |
|--|-----------|
| Introduction | 6 |
| Data and data risk | 7 |
| Definition | 7 |
| Data risk management | 7 |
| Data quality | 8 |
| Classification by criticality and sensitivity | 8 |
| Industry baselines | 8 |
| A systematic and formalised approach | 8 |
| Overarching framework | 8 |
| Principles-based approach | 9 |
| Roles and responsibilities | 9 |
| Ongoing compliance | 10 |
| Ongoing assessment of effectiveness | 10 |
| Data architecture | 10 |
| Staff awareness | 11 |
| Training and awareness programs | 11 |
| Staff education areas | 11 |
| Data life-cycle management | 11 |
| Data risk considered at all stages | 11 |
| Capture | 11 |
| Processing | 11 |
| Retention | 12 |
| Publication | 12 |
| Disposal | 12 |
| Other control considerations | 13 |
| Auditability | 13 |
| Desensitisation | 13 |
| End-user computing | 13 |
| Outsourcing/offshoring of data management responsibilities | 13 |

| | |
|---|-----------|
| Data validation | 14 |
| Assessment of fitness for use | 14 |
| Data cleansing | 15 |
| Monitoring and management of data issues | 15 |
| Monitoring processes | 15 |
| Data issue management | 15 |
| Data quality metrics | 16 |
| Data risk management assurance | 16 |
| Assurance program | 16 |
| Frequency of assurance | 16 |

Introduction

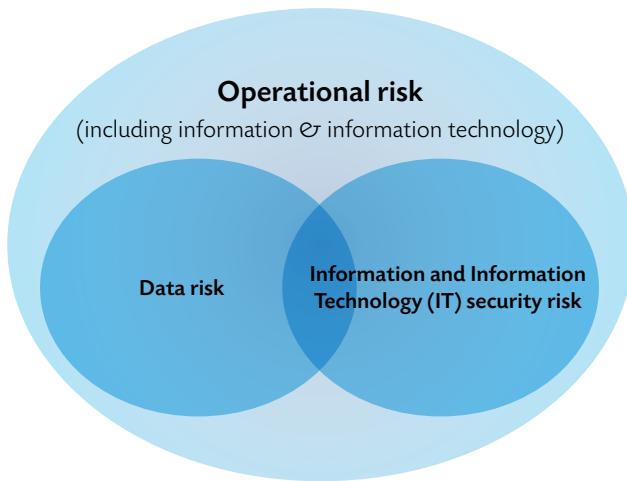
1. The management of data and associated risks is important for a broad range of business objectives including meeting financial and other obligations to stakeholders, effective management and proper governance. This prudential practice guide (PPG) provides guidance on data risk management where weaknesses continue to be identified as part of APRA's ongoing supervision activities.
2. While this PPG provides guidance for managing data and complying with APRA's prudential requirements, it does not seek to be an all-encompassing framework. APRA expects that a regulated entity using a risk-based approach will implement controls around data, including in areas not addressed in this PPG, appropriate for the size, nature and complexity of its operations.
3. Data is essential for a regulated entity to achieve its business objectives. Furthermore, reliance on data has increased as a result of process automation and greater reliance on analytics and business intelligence to support decision-making. Consequently, stakeholders including the Board of directors (Board), senior management, shareholders, customers and regulators have heightened expectations regarding the effective management of data. This trend has enhanced the importance of treating data as an asset¹ in its own right.
4. This PPG aims to provide guidance to senior management, risk management, business and technical specialists. The multiple audiences reflect the pervasive nature of data, and the need for sound risk management disciplines and a solid business understanding to effectively manage a regulated entity's data risk profile. Additionally, effective data risk management can facilitate business initiatives and assist compliance with other regulatory and legal requirements.
5. As with any process, governance is vital to ensure that data risk management and related business processes are properly designed and operating effectively to meet the needs of the regulated entity. In APRA's view, effective governance of data risk management would be aligned to the broader corporate governance frameworks and involve the clear articulation of Board and senior management responsibilities and expectations, formally delegated powers of authority and regular oversight.
6. Subject to the requirements of APRA's prudential standards, an APRA-regulated entity has the flexibility to manage data risk in the way most suited to achieving its business objectives.
7. A regulated entity would typically use discretion in adopting whichever industry standards and guidelines it sees fit-for-purpose in specific control areas. This PPG does not seek to replace or endorse any existing industry standards or guidelines.
8. The relevance of the content of this PPG will differ for each regulated entity, depending upon factors such as the nature, size, complexity, risk profile and risk appetite of the entity. The nature and specific usage of the data (current or potential) will also have an impact on the application of this PPG. APRA envisages that an entity's approach to managing data risk would also take into consideration the resources the entity has at its disposal, including whether the business is supported by an in-house technology function or an external service provider. Such factors will assist an entity in determining the relevance and extent to which it adopts the practices in this PPG.
9. This PPG also provides examples to illustrate a range of controls that could be deployed to address a stated principle. These examples are not intended to be exhaustive compliance checklists.

¹ 'Asset' is used here to represent anything deemed to be of value (either financial or otherwise) by an entity.

Data and data risk

Definition

10. Data² refers to the representation of facts, figures and ideas. It is often viewed as the lowest level of abstraction from which information and knowledge are derived.
11. Data risk encompasses the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events impacting on data. Consideration of data risk is relevant regardless of whether the data is in hard copy or soft copy form. Examples include:
 - (a) fraud due to theft of data;
 - (b) business disruption due to data corruption or unavailability;
 - (c) execution delivery failure due to inaccurate data; and
 - (d) breach of legal or compliance obligations resulting from disclosure of sensitive data.
12. For the purposes of this PPG, data risk is considered to be a subset of operational risk, which includes information and information technology risk. In addition, information and information technology security risk overlaps with data risk (refer to the diagram below).³



- 2 For the purposes of this PPG, data encompasses a broad range of categories including data that is entered, calculated, derived and structured or unstructured.
- 3 For further details, refer to *Prudential Practice Guide PPG 234 – Management of security risk in information and information technology* (PPG 234), which incorporates data (both hard and soft copy) as a subset of information and information technology assets.

Data risk can adversely affect a regulated entity and could result in a failure to meet business objectives (including regulatory and legal requirements). Consequently, it is important that business functions understand and manage the risks associated with the data required for the successful execution of their processes. Additionally, an understanding of data risk is beneficial when managing other types of risk.

Data risk management

13. A regulated entity would typically manage data risk in alignment with the operational risk framework and, where relevant, in conjunction with other risk management frameworks (e.g. credit, market and insurance risk management frameworks), depending on the nature of the data involved.
14. A goal of data risk management is to ensure that the overall business objectives of a regulated entity continue to be met. Therefore, it is important that an individual business unit's objectives are not considered in isolation, but rather in the context of the objectives of the entity as a whole. Consequently, the design of controls for a particular data set would typically take into account all usage of that data.
15. The adequacy of data controls in ensuring that a regulated entity operates within its risk appetite would normally be assessed as part of introducing new business processes and then on a regular basis thereafter (or following material change to either the process, usage of data, internal controls or external environments). The assessment would typically take into account the end-to-end use of the data and related control environment (including compensating controls). Changes to the control environment would typically follow normal business case practices, taking into account the likelihood and impact of an event against the cost of the control.

Data quality

16. In APRA's view, a useful technique for managing data risk is through the assessment and management of data quality. Data quality can be assessed using a range of dimensions. The relevance of each of these dimensions will vary depending upon the nature of the data. Dimensions typically considered in the assessment of data quality include:
- (a) accuracy: the degree to which data is error free and aligns with what it represents;
 - (b) completeness: the extent to which data is not missing and is of sufficient breadth and depth for the intended purpose;
 - (c) consistency: the degree to which related data is in alignment with respect to dimensions such as definition, value, range, type and format, as applicable;
 - (d) timeliness: the degree to which data is up-to-date;
 - (e) availability: accessibility and usability of data when required; and
 - (f) fitness for use: the degree to which data is relevant, appropriate for the intended purpose and meets business specifications.
17. Other dimensions that could also be relevant, depending on the nature and use of specific data, include:
- (a) confidentiality: restriction of data access to authorised users, software and hardware;
 - (b) accountability: the ability to attribute the responsibility for an action;
 - (c) authenticity: the condition of being genuine; and
 - (d) non-repudiation: the concept that an event cannot later be denied.

Classification by criticality and sensitivity

18. For the purposes of managing data risk, a regulated entity would typically classify data based on business criticality and sensitivity. The assessment would typically take into account the end-to-end use of the data. A regulated entity could seek to leverage the existing business impact analysis process to achieve this. The entity's data classification method and granularity would normally be determined by the requirements of the business.

Industry baselines

19. A regulated entity could find it useful to regularly assess the completeness of its data risk management processes by comparison to peers and established control frameworks and standards.

A systematic and formalised approach

Overarching framework

20. In order to ensure that data risk management is not conducted in an *ad hoc* and fragmented manner, a regulated entity would typically adopt a systematic and formalised approach that ensures data risk is taken into consideration as part of its change management and business-as-usual processes. This could be encapsulated in a formally approved data risk management framework outlining the entity's approach to managing data risk that:
- (a) includes a hierarchy of policies, standards, guidelines, procedures and other documentation supporting business processes;
 - (b) aligns with other enterprise frameworks such as operational risk, security, project management, system development, business continuity management, outsourcing/offshoring management and risk management;
 - (c) includes the expectations of the Board and senior management;

- (d) assigns a designated owner or owners;
- (e) outlines the roles and responsibilities of staff to ensure effective data risk management outcomes;
- (f) enables the design and implementation of data controls. The strength of controls would normally be commensurate with the criticality and sensitivity of the data involved; and
- (g) is reviewed on a regular basis, with periodic assessment for completeness against current practices and industry standards.

A data management framework could be defined at an enterprise-wide level, a business unit level, or as a component of other enterprise frameworks, as appropriate.

21. The establishment and ongoing development of the data risk management framework would normally be:

- (a) directed by a data risk management strategy and supporting program of work with a clearly defined budget, resource requirements, timeframes and milestones; and
- (b) an integral part of a regulated entity's change management and business-as-usual processes.

A data risk management strategy would typically be aligned with the regulated entity's business, information technology, and security strategies as appropriate.

Principles-based approach

22. APRA envisages that a regulated entity would adopt a set of high-level principles in order to establish a sound foundation for data risk management. Data risk management principles could include:

- (a) access to data is only granted where required to conduct business processes;
- (b) data validation, correction and cleansing occur as close to the point of capture as possible;
- (c) automation (where viable) is used as an alternative to manual processes;

- (d) timely detection and reporting of data issues to minimise the time in which an issue can impact on the entity;
- (e) assessment of data quality to ensure it is acceptable for the intended purpose; and
- (f) design of the control environment is based on the assumption that staff do not know what the data risk management policies and procedures are.

In addition, a number of specific security management principles are also relevant (refer to *Prudential Practice Guide 234 Management of security risk in information and information technology* for further details).

Roles and responsibilities

23. A key element in effective data risk management is the allocation of formal roles and responsibilities (pertaining to data) to appropriately skilled staff. This would typically articulate the data risk management responsibilities of staff, customers, service providers and other third parties. Common areas of consideration when formalising data management roles and responsibilities include:

- (a) data roles and responsibilities for general staff and data users;
- (b) data-specific roles and responsibilities, as applicable (e.g. data officers⁴, data custodians⁵, data owners/stewards⁶, designated business sponsor). These could form part of an individual's broader roles and responsibilities;
- (c) governance functions and reporting mechanisms to assess the ongoing effectiveness of the data risk management framework and ensure a continued focus on data risk and the escalation of data issues;
- (d) risk management, assurance and compliance roles;

⁴ A data officer is responsible for data processing and usage.

⁵ A data custodian is responsible for the safe custody, transport and storage of data.

⁶ A data owner/steward is responsible for authorising access to data and its quality.

- (e) data risk management framework roles (if applicable) including maintenance, ongoing review, compliance monitoring, training and awareness; and
- (f) responsibilities for data monitoring and management.

Ongoing compliance

24. APRA expects that a regulated entity would implement processes that ensure compliance with regulatory and legal requirements and data risk management requirements. This would typically include ongoing checks by the compliance function (or equivalent), supported by reporting mechanisms (e.g. metrics, exceptions) and management reviews.
25. A regulated entity would be expected to implement an exemption policy for handling instances of non-compliance with the data risk management framework (if relevant), including management of the exemption register, authority for granting exemptions, expiry of exemptions and the review of exemptions granted. Where exemptions are granted, APRA envisages that an entity would review and assess the adequacy of compensating controls initially and on an ongoing basis. Compensating controls would normally reduce the residual risk in line with the entity's risk appetite.

Ongoing assessment of effectiveness

26. APRA envisages that a regulated entity would regularly assess data quality and evaluate the effectiveness of data risk management, and make any necessary adjustments to ensure identified control gaps are treated in a timely and systematic manner. This could involve establishing a data improvement program that specifies target metrics, timeframes for resolution and associated action plans for closing any gaps identified. Typically, action plans would be prioritised and tracked.

Data architecture

27. In order to ensure that data risk management is effective, it is important that a regulated entity:
- (a) understands the nature and characteristics of the data used for business purposes;
 - (b) is able to assess the quality of the data;
 - (c) understands the flow of data and processing undertaken (i.e. data lineage); and
 - (d) understands the data risks and associated controls.
28. Data risk management could be supported by the use of data architecture practices. These practices assist in understanding how data is captured, processed, retained, published and disposed of. The sophistication of the data architecture⁷ would normally be commensurate with data risk. A data architecture could include:
- (a) a data strategy as a component of the broader business and information technology strategies, as relevant;
 - (b) information on the characteristics of the data, commonly referred to as metadata.⁸ This could include definitions, descriptions, sources, usages, update mechanisms, owners, authorised users, criticality, sensitivity and quality requirements;
 - (c) diagrams and detailed technical information that describe the underlying data structure⁹, the flow of data, key systems and data repositories and interfaces;
 - (d) description of the controls necessary across the various stages of the data life-cycle¹⁰; and
 - (e) standards and guidelines to facilitate the development of systems, data repositories, interfaces (including exchange of data with external parties) and data controls. This would normally include approved technologies (e.g. applications, data base management systems and data integration tools).

⁷ This can range from system documentation provided by vendors to an enterprise-wide data architecture.

⁸ Metadata is often embodied in a data dictionary.

⁹ Data structure is often embodied in data models.

¹⁰ Refers to the end-to-end life-cycle of data from the initial point of capture through to disposal. This differs from the system development life-cycle.

29. APRA envisages that the data architecture would normally align with a regulated entity's established policies, standards and guidelines. An entity would normally maintain the data architecture as part of its change management, project management and system life-cycle processes. This includes controls to ensure alignment to the standards and guidelines embodied in the data architecture.

Staff awareness

Training and awareness programs

30. A regulated entity would be likely to benefit from developing an initial and ongoing training and awareness program. For staff who do not have specific data risk management responsibilities, this would typically be incorporated as part of ongoing business process-specific or broader risk management training, as applicable.
31. A regulated entity could also consider incorporating data risk management responsibilities as a component of staff performance plans, as appropriate.

Staff education areas

32. In APRA's view, a regulated entity would regularly educate users as to their responsibilities in maintaining data quality. Common areas covered could include:
- ensuring the quality of data entered;
 - verifying the level of data quality prior to its use;
 - mechanisms for reporting data quality issues and concerns; and
 - adherence to the regulated entity's data-related policies and standards.

Data life-cycle management

Data risk considered at all stages

33. APRA envisages that a regulated entity would ensure that data risk is considered at each stage of its life-cycle and that appropriate controls are implemented to ensure that data requirements are met. Data-related life-cycle stages typically include data capture, processing, retention, publication and disposal.

Capture

34. Data capture controls, including manual entry of data as well as automated data feeds from internal business units and external sources, would typically be designed to ensure that newly introduced data meets data quality requirements. Controls in this area could include:
- user interfaces that conduct appropriate validation before data is accepted;
 - mechanisms to detect if automated data feeds are functioning as expected and to prevent erroneous data from progressing beyond the capture stage and prevent downstream processing from proceeding; and
 - specification of data quality requirements and the mechanisms for handling data quality issues included in agreements with internal and external parties.

Processing

35. A regulated entity would typically implement controls to ensure that data processing (the application of business rules to data, including regulatory and legal requirements) and the output generated continue to meet data quality requirements. This would usually include controls over:
- data integration (combining data from different sources) to manage the extraction, transformation and loading mechanisms;

- (b) acquisition and implementation via approved development, change and project management methodologies to ensure that data quality is not compromised by changes to the production environment;
- (c) exception handling to identify and respond to data quality issues in a timely manner; and
- (d) error-handling to ensure data is able be restored or corrected to a known level of data quality. This is commonly achieved through a variety of mechanisms including database management system checkpoint and rollback capabilities, data backup and recovery, and the design of automated processes so they can be re-run if required.

Retention

- 36. Data retention controls would typically be in place to ensure that data requirements are not compromised as a result of risks associated with the storage of data. This includes data hosting that is outsourced and/or located offshore. APRA's prudential standards and prudential practice guides on security, business continuity management and outsourcing provide specific requirements and guidance in this area.
- 37. A regulated entity could find it beneficial to develop a formal retention strategy that addresses the risks associated with data accessibility, and takes into account archiving and recovery requirements. Common issues in this area include accidental deletion, data corruption, changes in technology and poor asset management. The retention strategy would normally include mechanisms to ensure that data retention complies with business requirements, including regulatory and legal requirements.
- 38. As part of data retention, a regulated entity would normally implement robust protocols for data correction including approval and review of data changes, and maintenance of audit trails for tracking data changes. These controls would typically include appropriate segregation of duties, to reduce the potential for the actions of an individual to compromise data quality.

Publication

- 39. Data publication refers to the production of information for internal and external stakeholders (e.g. operational information, management information, customer information, media releases, regulatory reporting). Controls would typically be in place to ensure that published data meets the understood content and quality requirements of users. Examples include:
 - (a) acquisition and implementation controls as part of the introduction of new publication mechanisms (e.g. management review and approval, change management, project management and system development life-cycle);
 - (b) validation and monitoring controls to ensure published data continues to meet the specified requirements of users; and
 - (c) processes to manage data issues raised by users.
- 40. In APRA's view, it is important that data quality requirements are clearly specified and that confidentiality is not compromised through the publication of data.
- 41. Additionally, depending on the nature of usage, there could be benefit in a regulated entity including metrics with the data to provide users with an indication of the level of data quality (e.g. the level of completeness and accuracy).

Disposal

- 42. Disposal controls would typically be in place to ensure:
 - (a) data is disposed of in compliance with the retention strategy; and
 - (b) business requirements with respect to confidentiality are not compromised as hardware, software or data reach the end of their useful life or the hardware/software is recommissioned for another use.

Examples include the deletion of sensitive information prior to the disposal or recommissioning of hardware, archiving data prior to decommissioning systems and the removal of data following disaster recovery testing, if appropriate.

Other control considerations

Auditability

43. Auditability (the ability to confirm the origin of data and provide transparency of all alterations) is a key element to verifying data quality. It involves the examination of data and associated audit trails¹¹, data architecture and other supporting material. APRA envisages that a regulated entity would ensure that data is sufficiently auditible in order to satisfy the entity's business requirements (including regulatory and legal), facilitate independent audit, assist in dispute resolution (including non-repudiation) and assist in the provision of forensic evidence if required.

Desensitisation

44. Desensitisation (the process of reducing a data set's sensitivity to a level which complies with the authorised access of the end-user) is a useful approach for maintaining the confidentiality of data while extending its usage. Common approaches include the use of cryptographic¹² or de-identification¹³ techniques when transferring data to a less trusted environment (including the public domain). The strength of desensitisation should take into account the ability to reconstruct the original data set using other data available or brute force techniques.¹⁴

End-user computing

45. Current technologies allow for end-users to develop/configure software for the purpose of automating day-to-day business processes, facilitating decision-making and storing data. In addition, software is increasingly designed to enable extraction of data by users. This creates a risk that data life-cycle controls may be inadequate given that end-user developed/configured software is not typically subject to the controls that a technology function would apply.¹⁵
46. A regulated entity would normally introduce processes to identify the existence of end-user developed/configured software and assess its risk exposure. In APRA's view, any software that is used for the processing and retention of critical or sensitive data would comply with the relevant life-cycle controls of the entity.

Outsourcing/offshoring of data management responsibilities

47. Continued industry developments allow a regulated entity to more easily move data management responsibilities to service providers or other entities within a group (both on- and offshore). This increases the risk that data life-cycle controls may be inadequate, with problems potentially magnified when offshoring is involved. The possible causes of this increased risk include control framework variations, lack of proximity, reduced corporate allegiance, geopolitical risks and jurisdictional-specific requirements.
48. APRA expects a regulated entity to apply a cautious and measured approach when considering retaining data outside the jurisdiction it pertains to. It is important that a regulated entity is fully aware of the risks involved and makes a conscious and informed decision as to whether the additional risks are within its risk appetite.

¹¹ Evidence (e.g. log files, paperwork) of the sequence of activities that have affected data through a specific operation, procedure, or event.

¹² Methods used to transform data/information using an algorithm to make it unreadable to anyone except those possessing special knowledge, usually referred to as a key.

¹³ The removal of identifying information (e.g. name, date of birth)

¹⁴ A brute force technique is a method of defeating a desensitisation process by systematically trying a large number of possibilities.

¹⁵ For further details, refer to PPG 234.

49. When outsourcing/offshoring data management responsibilities, APRA expects that a regulated entity would be able to demonstrate the following:
- (a) ability to continue operations and meet core obligations following a loss of services;
 - (b) maintenance of the quality of critical or sensitive data;
 - (c) compliance with legislative and prudential requirements; and
 - (d) a lack of impediments (from jurisdictional hurdles or technical complications) to APRA being able to fulfil its duties as prudential regulator (including timely access to data in a usable form).
50. In APRA's view, the following would normally be applied to the assessment and ongoing management of outsourced/offshored data management responsibilities:
- (a) enterprise frameworks such as security, project management, system development, business continuity management, outsourcing/offshoring management, risk management and delegation limits;
 - (b) detailed risk assessments of the specific arrangements underlying the services offered. This would normally include assessments of the service provider, the location from which the services are to be provided and the criticality and sensitivity of the data involved;
 - (c) a detailed understanding of the extent and nature of the business processes¹⁶ and the sensitivity/criticality of the data impacted by the arrangement;
 - (d) alignment with the data architecture supporting the broader information technology and business strategies;
 - (e) a business case justifying the additional risk exposures;

- (f) Board/senior management's understanding, acceptance and approval of the resulting risk profile; and
- (g) periodic reassessment of risks in line with the entity's risk management framework.

Data validation

Assessment of fitness for use

51. Data validation is the assessment of the data against business rules to determine its fitness for use prior to further processing. It constitutes a key set of controls for ensuring that data meets quality requirements.
52. Regulated entities typically implement data validation controls (whether via manual or automated mechanisms) at the point of capture and at various points throughout the data's life-cycle. APRA envisages that the strength of the validation controls would be commensurate with the nature of the data and its classification.
53. Considerations when validating data include the level of trustworthiness (e.g. is the data from a provider with a known control environment and track record) and the extent to which data quality degrades over time. In APRA's view, regulated entities would design business processes to revalidate data on a periodic basis to minimise the degree of data quality degradation. The comprehensiveness of revalidation would normally be commensurate with the criticality of the data and the risk of degradation.
54. Common forms of data validation include verification of format, type, value range, currency, presence, consistency and completeness. Data validation can also be usefully conducted at a data-set level such as the use of:
- (a) control totalling: aggregation techniques including hash totalling¹⁷, amount totalling and record counts;
 - (b) reconciliation: comparing two sets of data and explaining variances;

¹⁶ Including those pertaining to decision-making and support.

¹⁷ The application of an algorithm to summarise a dataset in numeric terms.

- (c) benchmarking: comparing two sets of data that would normally exhibit similar characteristics, in order to highlight material variations;
 - (d) data profiling: examination of a data set and the gathering of statistics and other relevant information for the purposes of analysis to highlight any data anomalies (e.g. missing data, outliers, unexpected variances); and
 - (e) a review of data for reasonableness using expert judgement.
55. In APRA's view, where other validation controls cannot be easily implemented, a review of data for reasonableness using expert judgement would be beneficial as a minimum.
56. A regulated entity would normally document data validation processes, including their nature, frequency and level of granularity, and provide clear allocation of accountabilities for the detection, investigation, reporting and escalation of data anomalies. In APRA's view, data validation processes can be a key consideration when designing data quality metrics.

Data cleansing

57. Data cleansing is the act of detecting and correcting¹⁸ erroneous data. Erroneous data is anything that does not meet the quality objectives of the regulated entity. Entities would be expected to periodically cleanse data (e.g. as part of key business events such as member rollovers, claims, policy renewal) to ensure data quality remains at or above the required level. Data cleansing could also be required where the quality level requirements change over time (e.g. as a result of new usages or changes to existing processes) or when undergoing material change such as a system migration.

Monitoring and management of data issues

Monitoring processes

58. APRA expects that a regulated entity would have monitoring processes to identify potential data issues. The strength of monitoring controls would typically be commensurate with the criticality and sensitivity of the data. APRA envisages that alerts would be investigated in a timely manner with an appropriate response to address anomalies.
59. Clear allocation of responsibility for regular monitoring of data, with appropriate processes and tools in place to manage the volume of monitoring required, would assist in reducing the risk of a data issue not being detected.

Data issue management

60. APRA envisages that a regulated entity would develop appropriate processes to manage all stages of a data issue including detection, identification, containment, investigation, evidence gathering, resolution, return to business-as-usual and the adjustment of controls to reduce the risk of similar issues in the future. Common data issues include:
- (a) processing errors impacting on the accuracy and completeness of balances and transactions;
 - (b) lack of timeliness in updating data intended to reflect recent market conditions or assessments;
 - (c) inadequate data availability, accuracy or consistency resulting in pricing and valuation errors;
 - (d) data leakage leading to a breach of confidentiality;
 - (e) failure to accurately execute instructions in a timely manner;
 - (f) failure to maintain data quality when migrating data to another system; and
 - (g) data that is not fit-for-use, resulting in poor business decisions.

¹⁸ Correction can include data removal or addition of missing data.

61. Subject to the nature of the data, a regulated entity would:
- (a) have clear accountability and communication strategies to limit the impact of data issues. This would typically include defined mechanisms and thresholds for escalation and reporting to the Board and senior management, and customer communication where appropriate. Issue management strategies would also typically assist in compliance with regulatory and legal requirements; and
 - (b) conduct root cause analysis of the data issue, where the underlying cause of the issue is identified and analysed, with controls adjusted to reduce the likelihood of a future occurrence.

Due to resource constraints, regulated entities would normally prioritise remediation of data issues. A combination of tactical and strategic solutions may be required, depending on the root cause, including containment of identified issues.

62. In APRA's view, it could be beneficial for data users to provide feedback to data providers, both within the regulated entity as well as external parties, whenever data quality falls below the quality required.

Data quality metrics

63. Data quality metrics are a useful mechanism for assessing data quality and the success of data risk management. Typically, the use of metrics would be targeted to areas:
- (a) where there are regulatory, legal and specific industry requirements; and
 - (b) that have the greatest sensitivity/criticality, as determined by the risk assessment process.

64. Each dimension of data quality could be measured by at least one metric to enable the monitoring of progress towards set targets and the identification of issues and trends. Effective metrics would be specific, measurable, business oriented, controllable and reportable, and preferably involve the inspection of data to determine if a control is effective in maintaining data quality. Examples

of data quality metrics could include error rates, timeliness measures, materiality thresholds and reconciliation exceptions over a specified period.

65. APRA envisages that data quality gaps would be addressed over time in a systematic way. This may involve the formulation of a data quality plan that specifies target data quality metrics, timeframes for resolution and associated action plans for closing any gaps. Action plans would typically be prioritised and tracked.

Data risk management assurance

Assurance program

66. APRA expects that a regulated entity would seek regular assurance that data quality is appropriate and data risk management is effective. This would normally be implemented through the broader assurance program and result in a systematic assessment of data risk and the control environment over time. Assurance responsibilities would typically be conducted by internal audit or another independent function.

Frequency of assurance

67. A regulated entity would benefit from a multi-year schedule of testing that incorporates both adequacy and compliance-type reviews, with the program of work determined on a risk basis. Additional assurance work may be triggered by changes to vulnerabilities/threats or material changes to the business/information technology environment. Such reviews may encompass:
- (a) inspection of data;
 - (b) data risk management;
 - (c) general information technology controls;
 - (d) data architecture;
 - (e) data governance; and
 - (f) data metrics and data quality plans.
68. The schedule of testing would typically ensure that all aspects of the data control environment are assessed over time, commensurate with the sensitivity and criticality of the data.

Telephone
1300 55 88 49

Email
info@apra.gov.au

Website
www.apra.gov.au

Mail
GPO Box 9836
in all capital cities
(except Hobart and Darwin)