PLAYBOOK

How to Design and Implement a Governance, Risk, and Compliance Framework for Enterprise Data Analytics

RICHARD GEERING
VP EMEA, Immuta

2019

IMMUTA

Contents

Purpose, Scope, and Audience

Purpose

The data privacy and protection risks posed by enterprise data analytic projects can be wide ranging. Technology plays a central role both in giving rise to and in managing those risks, but is only part of the picture. Many of the most significant risk issues inherent in enterprise data analytic projects result from the familiar challenges of organisational design and culture, governance, resourcing, and, critically, the context within which an enterprise operates.

This document has been written to address those challenges; it describes the organisational factors which must be considered in designing and implementing data policies and access control (AC) for analytic initiatives within an enterprise.

Scope

AC is the process of determining and enforcing who can view specific data and, in some cases, how the data appears to users.[1] While the risks discussed are those specifically relating to data protection and privacy resulting from authorised access, managing these risks will also reduce data security risks related to unauthorised access.

Audience

The paper is aimed at a broad audience spanning business, data protection, compliance, legal, and IT professionals within firms undertaking data-centric initiatives, and consultants and systems integrators who may support such activities.

1 See Appendix A for a description of different AC models

Document Overview

AC is a core component of the internal control environment of a firm. An internal control is any action taken by management, the board, and other parties to manage risk and increase the likelihood that established objectives and goals will be achieved. In the case of a data analytic project, internal controls increase the probability that it will be executed successfully and in compliance with internal policy and regulations.[2]

Businesses must negotiate the challenge of allocating finite resources between production– and protection–focused activities. Controls, while essential, are typically established in response to business needs. Consequently, this document is structured around the lifecycle of a data analytic project, from establishing the context and idea creation, through analysis and results, to monitoring and review.

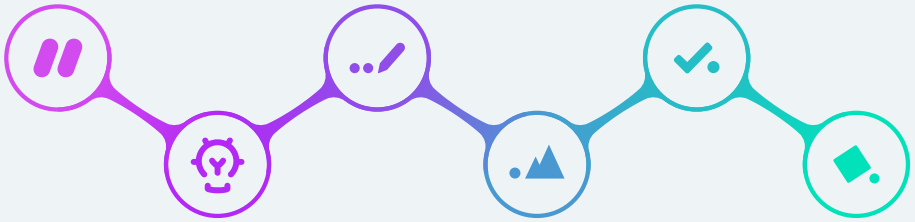
The document is arranged in two parts. The chapters in the first section describe, at a high level, the components and activities necessary to safely manage the risks related to data privacy and protection in a data analytic project; this is the "what". The second part addresses key elements in detail, and describes specific control implementation methods; this is the "why," "who," "when,'' and "how."

The degree of formality an enterprise will apply to data analytic project governance will depend on internal and external factors; a risk–based approach should be adopted. Firms operating in regulated industries such as finance are obligated to apply strict project management (PM) standards to material projects. Smaller firms in non–regulated sectors may employ a much lighter touch PM framework.

2 See Appendix C

Data Project Lifecycle

The lifecycle of a data analytics project has been broken down into the following six steps:



01 Context
What are the relevant external and internal factors?

02 Conception & Initiation
What is the idea and objective of the project?

03 Definition & Design
How will this be accomplished?

04 Risk Assessment & Management
What are the risks; is the control environment adequate?

05 Execution
Implement the analysis and risk management steps.

06 Monitor & Review
Are the risks within tolerance; what needs remediation?

Figure 1: Data project lifecycle stages

01 Context

Establishing the external and internal factors which are relevant to the design and execution of the data project is a core element of the risk-based approach concept.

External factors will include the regulatory environment and sector-specific expectations. Internal factors will include organisational culture and capabilities.

02 Conception & Initiation

The genesis of a data analytic project will likely be in response to a business challenge, need, or opportunity.

Emerging technologies or data may allow longstanding questions to be addressed, or new legislation may require additional regulatory reporting. In any event, before questions related to access control can be addressed, some fundamental aspects must be defined and documented, including

- **Project drivers** – Why this project, and why now?
- **Stakeholders** – Which business and support functions are involved? Who is the sponsor? Who owns the project outcome?[3] Specific to access control, two areas are critical: organisational design and decision authority – see Appendix B for an in-depth analysis of the stakeholder topic.
- **Type of project** – What is the expected project outcome? Is this aimed at research and development, a proof of concept or prototype, or a productionised data project?
- **Scope** – What is included and excluded from the project?
- **Timeline** – What are the proposed start and end dates and key milestones? Is this a "now" initiative? Will it take days, months, or years? How will we know if it's on track?
- **Preliminary capability assessment** – Can it be done; should it be done?

Defining the scope, drivers, and expected deliverables allows business and control functions to understand the potential demands on their environments and highlight any potential concerns or gaps at the outset.[4]

Preliminary Capability, Feasibility Assessment

It may be that further analysis is required before a go/no go decision can be made. If the firm lacks the expertise or resources to answer the questions above, that is a good indicator that gaps exist and must either be closed or the project shelved. Sometimes "no" or "not now" may be the right business answer.

KEY CHECKPOINT

The agreement or sign-off should be achieved and documented before progressing – both from the perspective of risk management and to avoid wasting resources.

3 The Association for Project Management identifies stakeholder engagement and management as the most important ingredients for successful project delivery. – https://www.apm.org.uk/resources/find-a-resource/stakeholder-engagement/key-principles/

4 Second Line of Defense groups, typically Risk, Compliance, Finance – see Appendix B

03 Definition and Design: Data Project

The Conception and Initiation phase deals predominantly with business-related questions, while the Definition and Design phase of the project develops some of the answers to these questions and addresses specifics related to operational, control, and IT processes.

In addition to fleshing out the aspects covered within the first phase, now is the time to identify specific data assets and the constraints with which the project must conform. The critical elements are

- **Objective** – As more context and facts are available it will be possible to refine the definition of the project objectives and make them more specific.
- **Data** – Within this context, data encompasses a large range of information needed to execute the project. Beyond the underlying measurements and observations which will feed insights, it will include metadata, intended usage, expected outcomes, and desired use cases. Metadata of the data assets necessary for the project, including owners, project members, constraints (see below), schema, and field descriptors, should be documented.
- **Constraints** – Constraints describe the environmental (e.g., regulatory, internal policy) and data centric (e.g., protected category, confidential) limits to which the data project must conform.
- **Controls** – What are the people, processes, and systems assets available to ensure the data is used compliantly? See Appendix C for a discussion of internal controls, including preventive, detective, directive, and corrective types.
- **Processing Activities** – How is the data going to be used? What are the people, processes, and systems that will touch the data?
- **Stakeholders** – During the definition phase of the project, additional information and insight can be obtained from stakeholders and subject matter experts (SME). Also, commitment and buy-in can be secured.

Definition and Design: Data Internal Control Environment

A key consideration is ensuring that the organisational design (including decision authorities), data protection policies, and the internal control environment are fit for the purpose. These topics are discussed in detail in Appendices B and C and should be revisited during the formal risk assessment stage.

04 Risk Assessment and Management

The risks assessed in this section are those directly associated with access control and data protection only. Other risks (e.g., project delivery risks), are outside the scope of this paper. Appendix D provides greater detail regarding risk assessment and specific methodologies.

The interaction between the project objectives, data, processing activities, constraints, and the control environment determines the feasibility and risk of the data project.

Risk based approach

KEY CHECKPOINT

Depending on the nature of the environment (e.g., regulated, confidential) and the materiality of the project, a risk-based approach to the formal risk assessment should be employed. The decision regarding the risk management approach should be documented.

If it is decided to undertake a formal risk assessment, then there are a variety of methods which may be employed if the firm lacks its own internal framework. The International Organisation for Standardisation (ISO) has published a standard on risk management and assessment techniques which is appropriate for this application.[5]

The elements comprising ISO's risk management methodology are

1. Communication and consultation	2. Establishing the context	3. Risk Assessment Comprising of: risk identification, risk analysis and risk evaluation	4. Risk Treatment	5. Monitoring & Review

The first two steps have been accomplished in Step 01 above and are the pre-work for the risk assessment proper. The last step, monitoring and review, is dealt with in Step 06.

Figure 2: ISO risk management methodology

5 www.iso.org

Risk Assessment: Identification

The constraints as identified in the Definition and Design section are a good starting point for identifying the risks, hazards, and potential adverse outcomes associated with the project.[6]

Additionally, reviews of historic data, both internal and external, and SME input are also useful sources of risk identification.

Risk Assessment: Analysis

A key advantage of the ISO framework is its flexibility, which can be applied to any risk management challenge. However, for this component of the overall risk assessment process, it may be beneficial to augment the ISO standard with frameworks such as those published by French data protection regulator, CNiL, or the National Institute of Standards and Technology, which are tailored to data protection risk analysis.[7] The CNiL framework in particular provides a quasi-quantitative method for risk assessment. Appendix D provides an in-depth analysis of how this part of the process can be implemented.

Identified risks should be analysed in relation to the existing controls, and a record of both the risk and its associated control(s) recorded.

Risk Assessment: Evaluation

The last step within the risk assessment stage is to compare the level and types of residual risk against the firm's tolerance.[8]

Each firm will have its own tolerance, or appetite, for different dimensions of risk.[9] For example, a social media firm may have a higher tolerance for reputational risk than that of a public sector body.

KEY CHECKPOINT

The level of risk versus a firm's risk appetite, and resulting treatment decisions, should be adequately documented.

6 Recital 75 of GDPR lists in excess of two dozen potential adverse outcomes as well as hazards which may result from the inappropriate processing of personal data; associated risks can be inferred from this list.

7 https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf, https://www.nist.gov/sites/default/files/documents/2019/02/27/outline_privacy_framework_2.27.19.pdf

8 Residual risk is the remaining risk after taking into account respective controls; inherent risk is the risk pre-controls.

9 The amount of risk that an organisation is willing to seek or accept in the pursuit of its long-term objectives – The Institute of Risk Management (IRM). See Appendix E

Risk Assessment: Treatment

The commonly accepted risk management options are:

Terminate (avoid)	If the mismatch between the estimated data project risk and an institution's risk appetite is too great, and/or the nature of the risk is such that no leeway is acceptable, deciding to terminate the project and avoid the risk entirely may be the best option. Or it may be that mitigation is possible, but proves to be uneconomic or compromises the project to such an extent that avoidance is the only practical option.
Tolerate (accept)	Conversely, if the risk falls within the firm's risk appetite, or strategic factors mean that the risk reward equation is favourable, then the risk may be accepted. The rationale for such an action should be adequately documented.
Transfer (to others)	Given the nature of data projects, it is less likely that the option of risk transference is available to an institution. This is especially true in the case of reputational risk. Insurance may be available to ameliorate risk in some other cases.
Treat (mitigate)	Designing effective risk mitigation mechanisms in relation to data analytics initiatives is an area in which firms can differentiate themselves and gain a competitive advantage. Mitigation strategies may include modifying aspects of the project to reduce risk, although care should be taken to avoid fatally compromising the initiative.

Enhancing the internal control environment is a further option.

Preventive Controls Access Control	The most opportune point to control the risks associated with data protection and privacy in relation to data analytic projects is at the point of data access, usually with preventive controls.
Directive Controls People and Process	While access controls are critical in managing data privacy risk in analytic projects, they are not the only potential control point. Data protection regulations such as HIPAA and GDPR are prescriptive in how data should be processed within data analytic projects.[10] Ensuring compliance systematically (via preventive controls) within a project once data has been accessed and is actually being processed is challenging, though technical solutions can be of help in minimising the burden. More appropriate controls are afforded through directive controls. Ensuring data analysts are aware of risks and obligations inherent in data processing, that they are well supervised by competent managers, and that processes are well documented and understood are key controls.
Detective Controls Monitor Data Access and Usage	Monitoring how data is being accessed can be an effective control to ensure data is being used compliantly — and to limit the impact in the case it is not. In addition to traditional internal control methods, such as segregation of duties and system enforced privileges, other techniques such as pseudonymisation, masking, and differential privacy are also available in relation to data protection and privacy.[11]
Corrective Controls Identify and Remediate Issues	Corrective controls serve two purposes; firstly, they identify issues and ensure restoration to an acceptable or normal situation. Secondly, and more importantly, corrective controls draw attention to systems and process failures and ensure they are remediated upstream to prevent future occurrences.

10 https://www.hhs.gov/hipaa/index.html; https://eugdpr.org/

11 Masking entails altering the attributes of a table so that individual records can't be easily traced back to the original individual. Pseudonymisation is a de-identification procedure which replaces identifiable values within a data record with artificial identifiers, or pseudonyms. Differential privacy is a constraint on response to an aggregate query or algorithm executed over a database, limiting an observer's ability to assess any single individual's impact on the query with arbitrary confidence.

05 Execution

Augmentation of the internal control environment (e.g., access control) and related monitoring and reporting activities may be required in addition to actual data analysis.

If a firm is operating in a greenfield site, lacking foundational aspects of AC or data governance such as internal policy or risk appetite statement, these resources should be developed in tandem with the data project. Sufficient resources and priority should be given to governance-related workstreams to ensure these do not fall behind business-related topics.

It is acknowledged that finite resources must be allocated between production (business) and protection (control) activities, and often production wins. Typically the balance is only reset following a catastrophe or near-death experience. Evaluating how much to invest in protection-related activities and securing resources to achieve an effective and efficient control environment are key business skills, which over the long term differentiate successful enterprises from merely lucky ones.[12]

Controls: Design vs Effectiveness

In the prior Risk Management phase, the control environment was assessed relative to the risks it must help manage. Additional controls may have been designed to ensure the residual risks of the project were commensurate with the firm's risk appetite.

Care should be taken to ensure both the existing and new controls are effective. This may require unit- or sample-based testing. Near-misses, where controls are ineffective but no actual compromise of data protection has occurred should be treated seriously, are "free lessons".[13]

Execution: Data Access Controls

Additional data access controls specific to this data analytic project may need to be configured or implemented via an access control plane or other systems. Care should be taken to ensure that the appropriate roles and responsibilities are in effect to ensure potential conflicts of interest are managed. See Appendix B.

12 https://www.sciencedirect.com/science/article/abs/pii/S0024630108000137, https://www.gelaw.com/wp-content/uploads/2015/02/Does_Corp_Gov_Matter.pdf

13 https://www.bbts.org.uk/downloads/ac14/presentations/0900_fri_qs1_alison_watt.pdf/

06 Monitor and Review

The last phase of the data analytic project comprises two activities: monitoring and review; ISO groups them together as the last step of its risk management framework. However, care should be taken to ensure specific items do not distract from broader themes and vice versa.

KEY CHECKPOINT

Monitor

Specific risks and controls identified during the risk assessment phase should be tracked to ensure that they remain within expected parameters during the execution and post-execution phase.

Throughout the data analytic lifecycle there should be sufficient management reporting to allow stakeholders to understand how the project is tracking and to be aware of any risks or issues.

The execution phase, when data is accessed, is the point at which any latent defects in the project or control environment are likely to come to light, and extra scrutiny should be focused on controls and audit logs at this point.

> The aggregate risk vs tolerance should be monitored to ensure the firm complies with its risk appetite.

Review

Not forgetting the risk-based approach axiom, which stresses the importance of managing checkpoints at each project phase transition where key business and control function stakeholders can review progress and raise any issues, is best practice.

If the project is designed to deliver a productionised data analytic project, operational reporting should be part of the delivery objectives. As important as designing and producing reports is, ensuring they are being read, understood, and acted upon by a designated person or team. Escalation criteria, processes, and paths also must be defined.

Defining an endpoint and, with some level of formality, declaring that the project has finished are critical, but sometimes overlooked items. In the case of a data analytic project whose goal was to achieve a working prototype, care must be taken to avoid scope-creep and erroneously allowing a product which was not designed to cope with the demands of such an environment to be deployed into production.

In all cases, and especially in the case of operational products, it is essential that any outstanding issues are assigned, tracked, and resolved in a timely manner. Furthermore, stakeholders must understand any gaps, either between the planned functionality or in the control environment.

Conclusion

Risk management, whether aimed at the risk associated with data privacy and protection risks within a data analytic project or any other risk, is a question of context, suitability, and discipline.

Context determines the kinds of hazards that may be encountered and the potential resulting adverse outcomes that should be anticipated.

Firms should maintain perspective and ensure that their internal control environments are commensurate with, and suitable for, inflight and planned data analytic projects. Care must be taken to ensure a healthy balance between production- and protection-focused activities. This theme of **suitability** is dealt with in detail in the last appendix.

The best plans and risk assessments will come to naught without the **discipline** of good corporate hygiene, which involves documentation, sign off, follow up, issue tracking, and ongoing monitoring.

Finally, perhaps the most important aspect of managing a risk is to make sure it is clear who owns it — if a risk is assigned to nobody, that is exactly who will manage it.

Appendices

The appendices form the second part of the document; they expound on areas and concepts covered in the first section of this document.

The first three Appendices — Data Access Control, Organisational Design, and Internal Controls — serve as foundational elements for the last chapter: Risk Management.

APPENDIX A

Data Access Control

IN THIS SECTION

Data Access Policy

Institutions should establish and maintain policies which govern under what circumstances data can be accessed for analytic purposes.

Policies should be sufficiently detailed to avoid misinterpretation or ambiguity. In the most general case, policies may be designed to determine access permissions based on attributes related to

- **Subject (user)** – e.g., role, location
- **Object (data)** – e.g., sensitivity, intended use, purpose
- **Operation (purpose)** – e.g., intended outcome of the analytic project
- **Environment (context)** – e.g., time of day, business urgency, legislation

Policies are effected via access control models, typically system-based preventive controls; in the case of complex policies requiring a degree of human interpretation, directive and detective controls may be employed.

Access Control Models

Role-Based Access Control (RBAC)

RBAC has long been employed as a standard implementation of data access control policy. It is a type of group-based access control which grants certain data access privileges to users based on their role(s). For example, employees within the Finance team can access the ClientBalances table. Bob works in Finance; therefore, Bob can view ClientBalances.

The benefits of RBAC are that if members of the Finance team need access to many data assets in the course of their role, data access administrators can grant a new member of the team all the access they need by simply adding them to the Finance group. This benefit is predicated on the assumption that the number of groups (roles) is limited relative to data assets and employees.

In practice, however, role proliferation driven by custom access requests often means that this model can become unscalable for large enterprises.[14] A specific challenge is for employees to identify of which particular established groups (roles) they need membership so that they can do their jobs, without needing to ask for new groups (roles) to be created for them.

14 See Data Governance Anti-Patterns: Stop Conflating WHO, WHY, and WHAT

Attribute-Based Access Control (ABAC)

ABAC exploits the four dimensions described above to avoid the need to create custom roles to manage specific data access situations. Instead it leverages user (subject) and data (object) attributes to achieve policy objectives. (In fact, RBAC is a specific, and limited, implementation of ABAC.)

The benefit of ABAC is that it is infinitely configurable and can accommodate any policy demands. The challenge is that for an institution to benefit from the flexibility ABAC affords, a significant upfront investment in policy, metadata, and organisational design may be required.

In certain situations (e.g., GDPR), RBAC is not sufficiently featured to satisfy policy demands, and people and process mitigants must be used to augment RBAC policies. GDPR-compliant policies can be fully implemented using ABAC and suitable infrastructure (e.g., tagging, access control planes).

Purpose-Based Access Control (PBAC)

PBAC is a special case of the generic ABAC-type model implemented via leveraging object and operation attributes related to the intended and actual use of data.

For example, if data is collected related to natural subjects (i.e., persons) with their consent for a particular purpose, it should be subsequently used for that specific purpose only. This constraint would be articulated via a purpose-based access control policy.

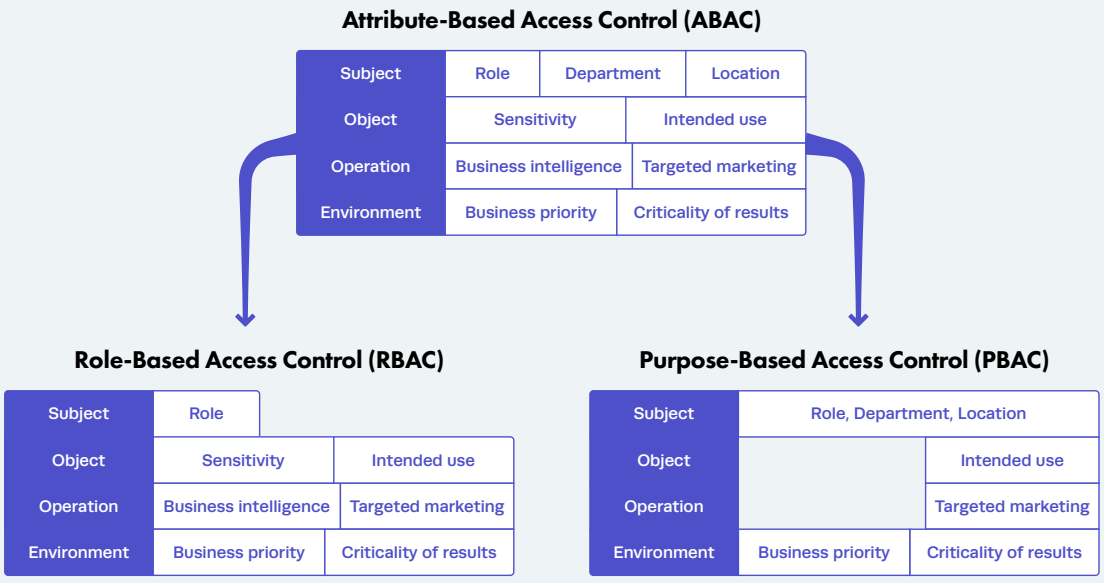


Figure 3: RBAC and PBAC are special case implementations of ABAC.

Tagging

Tags are a type of metadata which can be associated with data at the element, row, or column level. Tags can be leveraged as subject or object attributes via ABAC–type policies to implement data access control.

The tagging process may be driven top–down as part of a controlled vocabulary or bottom–up by data owners and users. In each case, adequate controls need to be implemented to ensure that the process to derive the taxonomy and apply it to data is complete and accurate.

Access Control Planes

Whatever access control model is adopted, there remains the question of how and where to implement the data access policy.

Enterprises typically employ identity and access management (IAM) frameworks and technologies to manage user profiles, including data access privileges. However, even with the advent of enterprise data storage technologies, many data analytic projects require access to data assets which span different locations, technologies, and business areas. This creates an IAM challenge which has led to the creation of access control plane solutions, where data access policy is enforced at the point of demand rather than on each data storage asset.

Using access control planes, data access policies (whether simple RBAC or complex ABAC policies) can be enforced in one layer, irrespective of the underlying storage or downstream technologies and use cases.

APPENDIX B

Organisational Design

This section describes the roles and responsibilities relevant to internal control design and implementation specific to data access.

IN THIS SECTION

Objectives and Roles

All businesses must negotiate the challenge of allocating finite resources between production (business focused) and protection (control focused) activities. The relative split between each part will depend on various factors, including the maturity and nature of the business and its industry sector and risk appetite.

In highly regulated and mature industries, authorities may demand that roles focusing on production are separated from those dedicated to protection. Financial services is such a sector, and it is perhaps the most evolved in terms of organisational design principles to manage conflicts of interest and risk, via the Three Lines of Defense framework.

Defenses in Depth: The Three Lines

The Bank for International Settlements overhauled its guidance for banks post the 2008 financial crisis and defines the responsibilities of each of the lines of defense as follows:[15]

- The business line – the first line of defense – has "ownership" of risk, whereby it acknowledges and manages the risk that it incurs in conducting its activities.
- The risk management function is responsible for further identifying, measuring, monitoring, and reporting risk on an enterprise-wide basis as part of the second line of defense, independently from the first line of defense. The compliance function is also deemed part of the second line of defense.
- The internal audit function is charged with the third line of defense, conducting risk-based and general audits and reviews to provide assurance to the board that the overall governance framework, including the risk governance framework, is effective and that policies and processes are in place and consistently applied.

While the BIS guidance is aimed at the financial services sector, firms operating in other industries which share commonalities in terms of data protection and privacy risk may consider formalising, or functionalising, roles along these lines.

15 https://www.bis.org/bcbs/publ/d328.pdf

Scale will also determine whether a firm has sufficient resources to separate roles into different lines; employees within smaller firms will necessarily wear several hats in relation to data access and protection.

However, in all cases, firms should be aware of actual or potential conflicts of interest and ensure they are documented and managed via a risk-based approach using an agreed framework (e.g., terminate, tolerate, transfer, treat). Segregation of duties can be used as a key control, as described in the Internal Controls appendix.

Control Functions

Groups within the second line of defense are collectively known as the "control functions". These will include Group Risk Management, Compliance, Finance, and, potentially, Legal.

Roles - Data Access Project and Access Controls

The field of data governance is still evolving and developing in maturity. For example, in contrast to, say, the role of a Chief Risk Officer, there is no accepted best practice definition of the responsibilities of a Chief Data Officer.

Consequently, it is challenging to define a one-size-fits-all target operating model for data access control. However, the following roles are likely to be present, or required, in some form in firms involved in material data analytic projects.

First Line of Defense

Project Business Sponsor
Responsible for defining the project's objectives and ensuring they are met.

Data Governance
Responsible for designing and implementing business focussed rules to use data safely and effectively. Data Governance may act as an internal control function, based in the first line, but independent from day-to-day operations. Depending on the maturity of the organisation, Data Systems Admin may be incorporated into this team.

Data Analyst
Responsible for delivering the data project and results; identifying and sourcing data.

Data Systems Admin
Responsible for configuring access control and other data-focused applications.

Data Owner
Responsible for the business area which produces the data processed in the data analytic project. In some cases data ownership may be transferred to a specialist business function supporting enterprise-wide data analytic initiatives.

IT
Responsible for business or enterprise-wide information technology and data storage applications.

2 Second Line of Defense

Chief Data Officer

Responsible for the firm's enterprise-wide data and information strategy, governance, control, policy development, and effective exploitation.[16]

Data Protection/Compliance/ Risk/Security Officer

In relation to data protection and privacy, these roles are focussed on monitoring and oversight to ensure that the organisation processes data in compliance with applicable data protection rules.

3 Third Line of Defense

Internal Auditor

The role of internal audit is to provide independent assurance that an organisation's risk management, governance, and internal control processes are operating effectively. Internal audit perform periodic and trigger-based reviews on business areas, control functions, or other targeted areas.

Responsibility Assignment Matrix

A responsibility assignment matrix is an effective way of describing the respective roles and their responsibilities throughout the data analytic project.

Below is an example using the Responsible/Accountable/Consulted/Informed (RACI) model.[17]

The process steps of the project are recorded as rows and the stakeholders by columns in a table. The characters at each of the intersections of the rows and columns describe the specific function, if any, for that process step.

Definitions of each of the R / A / C / I letters are as follows:

- **Responsible:** those who do the work to complete the task. There is at least one role with a participation type of responsible, although others can be delegated to assist in the work required.
- **Accountable:** the one ultimately answerable for the correct and thorough completion of the deliverable or task; the one who ensures the prerequisites of the task are met and who delegates the work to those responsible. There must be only one accountable specified for each task.
- **Consulted:** those whose opinions are sought, typically subject matter experts, and with whom there is two-way communication.
- **Informed:** those who are kept up-to-date on progress, often only on completion of the task or deliverable, and with whom there is just one-way communication.

16 https://www.gartner.com/smarterwithgartner/understanding-the-chief-data-officer-role/
17 https://en.wikipedia.org/wiki/Responsibility_assignment_matrix

An example of a RACI specific to data analytic projects is shown below.

		1ST LINE OF DEFENSE						2ND LINE OF DEFENSE		3RD LINE OF DEFENSE
PROJECT STAGE		Project Business Sponsor	Data Governance	Data Analyst	Data Systems Admin	Data Owner	IT	Chief Data Officer	Data Protection/ Compliance/Risk/ Security Officer	Auditor
Conception and Initiation		A / R	C			C	I	C	C	
Definition & Design	Data Project	A / R	C	R		C	R	C	C	
	Data Access Permissions	A / R	C	I	I	C	I	C	C	I
Risk Assessment and Management		A / R	C	R		C	C	C	R	I
Execution	Data Project	A / R	C	R	R		R	I	I	I
	Data Access Permissions	A	R	I	R			I	I	
Monitor & Review		A / R	R	R				I	R	I

Table 1: Data analytic project RACI

In this example, accountability for the risks remains with the business owners within the first line of defense.

Throughout the design and execution phases, input from other stakeholders within the first and second lines of defense is sought to ensure both production and protection aspects of the project are considered.

Responsibility for certain tasks (e.g., configuration of access control systems) may be delegated to other areas.

Each institution should tailor the values within the cells to suit its particular environment and circumstances.

However, the power of the RACI approach is based on ensuring that material process steps and stakeholders are identified and participate in the RACI model development and agree on the particular descriptions of their responsibilities and contributions to the project.

Institutions should ensure they design their RACI to be appropriate to the risks that they seek to manage, in particular with respect to conflicts of interest and segregation of duties. Secondly, they should act in congruence with their agreed RACI to be effective.

Data Policy Design

In addition to defining roles and responsibilities with respect to the implementation of data policy via access controls, firms must define the equivalent process for creating and maintaining data access policy.

There is a significant overlap between the stakeholder constituents of each process, although the policy creation activity may also involve the legal function, which is likely absent from policy implementation activities.

	1ST LINE OF DEFENSE			2ND LINE OF DEFENSE			3RD LINE OF DEFENSE
TASK	**Senior Mngt/ Project Sponsor**	**Data Governance/ Admin**	**Data Owner**	**Chief Data Officer**	**Compliance / Data Protection**	**Legal**	**Auditor**
1. Create Policy Framework	C		C	A /R	C	C	I
2. Monitor for new legistation	I			I	A /R	R	
3. Create local standards / procedures	A / R		R	C	C	C	I
4. Implement / operationalise procedures	A	R	C	I	I		
5. Compliance Monitoring & Reporting	A / R	i	I	C	R	I	I
6. Periodic policy review	R /C		C	A / R	C	C	I

Table 2: Data policy RACI

Here, the office of the CDO owns the data policy framework. It seeks input from other stakeholders, including the business, but ultimately it has final authority on how the framework is designed and managed.

Monitoring for new legislation or regulations which may impact data policy is the remit of the Compliance and/or DPO function. They may refer to the Legal department for matters related to the interpretation of legislation.

Once enterprise-wide policies are determined, it is the job of each business line to create and operationalise local standards. This means business areas will need to develop policies appropriate to their context and needs, but subordinate to and compliant with parent policies. These policies must be developed in consultation with the second line functions, including the CDO.

Ensuring compliance with data policies is anchored in the first instance with the business line. Again, the first line of defense owns the risk. Compliance is also responsible for providing oversight and monitoring.

Finally, the office of the CDO will execute periodic reviews of the global policies and mandate that business lines also review their respective policies.

APPENDIX C

Internal Controls

IN THIS SECTION

The Institute of Internal Auditors defines a control as:

> *Any action taken by management, the board and other parties to manage risk and increase the likelihood that established objectives and goals will be achieved. Management plans, organises and directs the performance of sufficient actions to provide reasonable assurance that objectives and goals will be achieved.*

Critically, controls are not created in and of themselves; their purpose is to improve business outcomes through risk management. In this regard, controls go hand-in-hand with both good business practice, and the discipline of risk management. Business objectives must be a foundational element of control frameworks; COSO's excellent methodology recommends as much.[18]

Control Types

The Chartered Institute of Internal Auditors list four types of control:[19]

- **Preventive** – avoid an adverse outcome materialising (e.g., access control and segregation of duties).
- **Detective** – discover in a timely manner that an adverse outcome has occurred (e.g., access logs, and exception reporting).
- **Directive** – procedures and guidance to reduce risk (e.g., documentation, training, and supervision).
- **Corrective** – systemic controls to restore normal state (e.g., error handling and operational incident reviews).

18 https://www.coso.org/Pages/ic.aspx
19 https://www.iia.org.uk/resources/control?downloadPdf=true

People, Process, Systems

The familiar organisational dimensions of people, process, and systems can be considered in relation to the four control types.

A key control design principle is to ensure that the right resources are being used in an appropriate way to effect a control, and the holistic control environment should be understood. For example, system-enforced access controls are very effective in preventing unauthorised access to data. However, systematic controls, at some level, rely on people, and a people-based control, such as segregation of duties, is likely also necessary.

Furthermore, each of the elements of people-, process-, and systems-based controls may be relevant to each of the control types; the overreliance on a single pairing (e.g., implementing preventive controls exclusively systematically) should be avoided.

Control Objectives and Ownership

Key to an effective control environment is ensuring that each control is defined in terms of its control objective: what specific risk is it seeking to address?

To what extent will it reduce the identified risk? Equally important is identifying who, or which group, is responsible for implementing and operationalising the control.

APPENDIX D

Risk Management

This section describes how risk management techniques based on industry standard methodologies such as failure modes and effects analysis (FMEA), fault tree analysis (FTA), and the ISO standard can be used to identify, assess, and manage data protection and privacy risk within data analytic projects.

IN THIS SECTION

Risk-Based Approach

Throughout this document, the concept of a risk-based approach has been surfaced; context is everything. In the case of institutions dealing with very sensitive data, when the consequences of unauthorised access may be unacceptable, then a detailed risk assessment is essential. In other cases the investment may not be justified.

Risk Management Terminology

Before diving into the risk analysis, it is necessary to review some terminology; even among professionals, basic terms are sometimes misused:[20]

- **Adverse Outcome or Harm** – materialisation of one or more of the following affecting either a data subject and/or the institution: loss of privacy, loss of confidentiality, financial loss, strategic/operational impairment, personal injury/death (including moral harm), reputational damage.
- **Hazard** – a potential source of harm or adverse effect on a person or organisation.
- **Risk or Likelihood** – the probability that a person may be harmed or suffer adverse effects if exposed to a hazard; the chance of a risk materialising in a given timeframe.
- **Impact** – the severity of an adverse effect or outcome.
- **Detectability** – the probability of identifying a defect and correcting it before it materialises into an adverse outcome.
- **Failure Mode** – the mechanism or process through which a hazard materialises into an adverse outcome.
- **Inherent and Residual risk** – residual risk is that which remains after taking into account the ameliorative effects of respective controls on inherent risk.

20 Unless otherwise stated, for the sake of readability, throughout this document the term "risk" has been employed to refer to both likelihoods and hazards.

Risk Management Methodology

ISO's five-step framework is as a good starting point:

- communication and consultation
- establishing the context
- risk assessment (comprising risk identification, risk analysis and risk evaluation)
- risk treatment
- monitoring and review

Whichever risk framework is adopted, the process of risk management is to identify the risks, establish how significant they are, decide on a plan, execute, and review.

Communication and Consultation

This phase is designed to ensure relevant stakeholders are aware of the risk management programme and objectives, and to secure their buy-in. It is a two-way process, those responsible for the risk management activities must also seek to understand stakeholders' objectives and priorities and to ensure the programme integrates with other programmes (e.g., change management, project management).

There are parallels here to the notion of privacy by design. As noted in the UK's Information Commissioner's Office guidance, and also within the CNiL framework, ensuring risk management is front of mind from the outset of a data analytic project is more likely to result in the most effective and efficient outcome.[21]

Context

Context is the key determinant in judgements related to "risk-based approach"; it is the basis on which decisions of this kind are made. Key dimensions to consider are:

Environmental

- Regulatory landscape
- Sector-specific expectations

Organisational

- Risk appetite
- Culture
- Technology infrastructure
- Internal control environment
- Staff training, awareness and competence

The type of hazards present are also a key aspect of the context. In the field of data privacy and protection the principal hazard is the confidentiality, sensitivity, or secrecy of data. Large datasets, data processing at scale and automated decision making also pose intrinsic dangers and should be considered hazards in their own right.

21 https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/ See notes pg 10 section 2 of CNIL's framework: https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf

Risk Assessment

The risk assessment process comprises three steps: ***identification, analysis and evaluation.***

This section works through two different, and complementary, methods of accomplishing the identification and analysis steps:[22]

- Failure modes and effects analysis, FMEA
- Fault tree analysis, FTA

In each case, establishing the scope of the risk study is critical; risk studies should be congruent with the context and scope of the overall data project and sufficiently comprehensive to evaluate the risks of adverse outcomes materialising.

Failure Modes and Effects Analysis

Failure modes and effects analysis (FMEA) is an industry standard mechanism for estimating the materiality of "risk" in a system. Using the FMEA method, the end-to-end data processing system is reviewed and potential failures identified and scored for their respective impact, likelihood, and detectability.

Typically, FMEA is implemented using a scale of 1 to n (with n being the most adverse score) for each of the dimensions. In the case of impact and likelihood, a high number indicates a more significant risk. Conversely, a low score for detectability indicates a better chance of avoiding an adverse outcome. This means that the product of impact, likelihood, and detectability (also called the risk priority number, or RPN) will result in a score between 1 and n^3. The RPN allows firms to identify the highest risk items and tackle them first.

The FMEA mechanism consists of four steps:

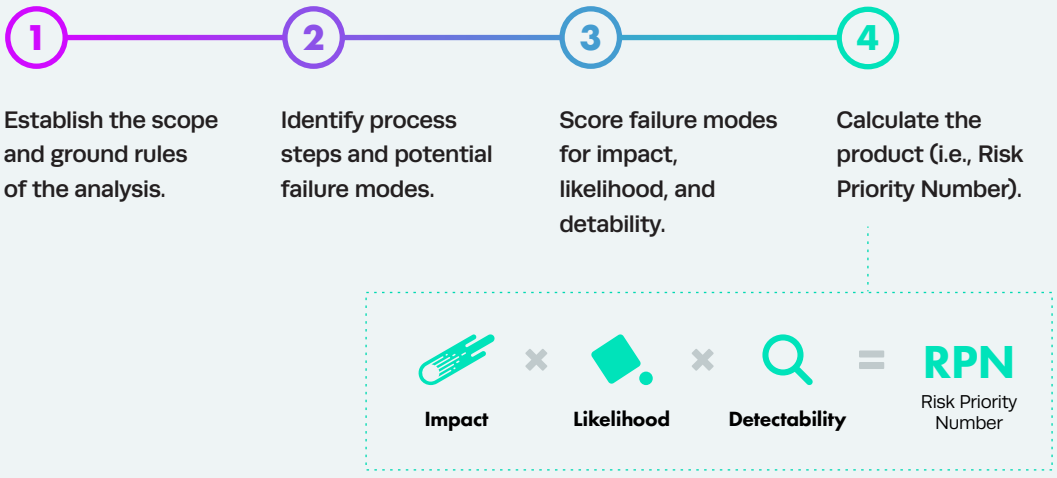


Figure 4: FMEA process diagram

22 See below for a comparison of each method.

Fault Tree Analysis

Fault tree analysis (FTA) is a technique for identifying and analysing factors than can contribute to a specified undesired and critical event (the "top event"). Causal factors are inductively identified and organised in a tree diagram which depicts failure modes and their logical relationship to the top event.

The FTA process consists of six steps:

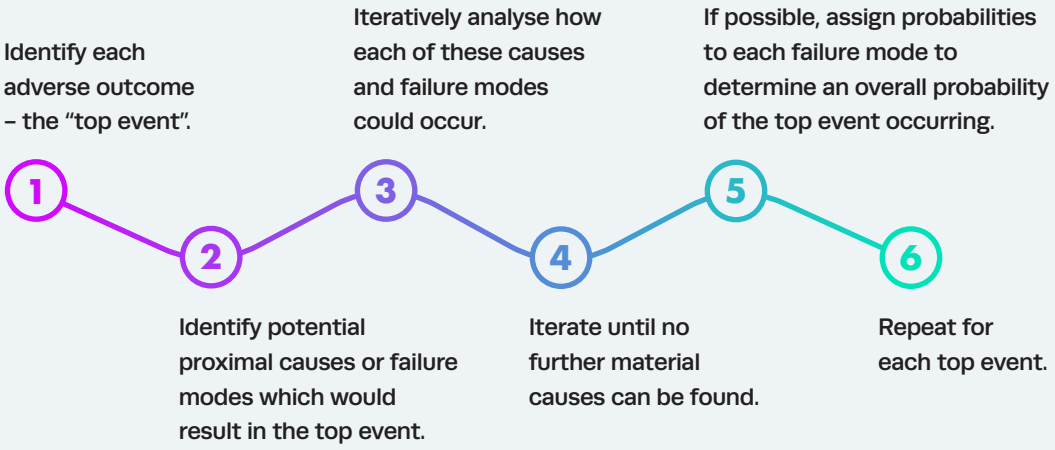


Figure 5: Fault tree analysis process

FMEA vs FTA

Each of the methods is described within the ISO standard, with associated strengths and weaknesses documented. (Notwithstanding the ISO definition asserts FTA is a deductive method, it is more readily argued to be inductive.)

In the field of data privacy and protection, if the processes are well known and documented, then the FMEA method may be preferred, since it is probably more intuitive and well known. Risk managers can work forward on the basis of the documented processes.

However, in the case where the business processes supporting data privacy and protection are less mature, or a firm is developing its capabilities in the field of data analytics, it may be that the top events (adverse outcomes) are articulated more concretely than the processes.

In such cases, employing the FTA methodology and working backwards from the adverse outcomes may be easier and will enable a business to evaluate and develop its control environment.

See Table 3 for an example of how FTA and FMEA compare in practice

Risk Identification - Failure Modes

Whether using FMEA or FTA, the failure modes (i.e., the mechanisms through which hazards materialise into adverse outcomes) must be identified.

As stated above, in the case of FMEA, a defined process is evaluated and each step reviewed for how it might fail and cause or contribute to an adverse outcome. If using FTA, the failures which cause each adverse outcome, either the top event or a sub-process failure lower in the tree, must be found. In the abstract (i.e., this document), it is challenging to evaluate particular failure modes; however, whether using FMEA or FTA, some generalities can be found by leveraging the ABAC concept, which will be discussed on page 33, following the FMEA- and FTA-specific sections below.

Risk Analysis

Risk analysis is the process of determining the materiality of the identified risks. At this point of the risk assessment process, the FMEA and FTA paths diverge and are dealt with separately below.

Risk Analysis: FMEA

Under FMEA, each specific failure mode, and in a complex system there may be dozens, must be scored according to its respective impact, likelihood, and detectability dimensions, with each RPN as the resulting product.

The failure modes are specific to, and derived from, the data analytic project process. This must be sufficiently defined and documented to successfully execute an FMEA study.

Each institution, according to its context, should develop rules or guidance related to impact, likelihood, and detectability. It is acknowledged that given the paucity of data, this may be a challenging exercise, but experience has shown that having some guidance, even if subjective to a degree, is better than none.

FMEA: Impact

The French data protection regulator, CNiL, has developed a framework based on the EU data security standard EBIOS, which can help in respect to estimating impact within FMEA studies.[23] It combines the potential harm (prejudicial effects) with the specificity of the data (level of identification) to give an overall impact score – see pages 12–14 of its guidance for implementation details.[24]

23 https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/rm-ra-tools/t_ebios.html

24 https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf

FMEA: Likelihood

Likelihood of each failure mode is assessed given the existing internal control environment.

Experience from the field of operational risk shows that in addition to the degree of manual intervention vs systematisation, key factors driving the probability of failure are the level of complexity, scale, and change as compared with the capabilities of the internal control environment.

Any material gap between the operational risk challenges and the internal control environment will most likely result in adverse outcomes materialising.

For example, if an institution finds managing its users' IAM profiles challenging, and these profiles are leveraged within its data access control environment, then the risk of a privacy breach is all too obvious.

The complexity of an institution's data landscape, particularly if it is spread across multiple jurisdictions (given that different data protection legislation may exist), is also a key driver of the likelihood of a privacy failure.

Organisational change causes operational risk incidents to spike, as the demands on staff increase and control infrastructures struggle to keep pace with new activities.

FMEA: Detectability

The extent to which a defect can be trapped before it results in an adverse event depends on the controls and reporting throughout the data analytic project's lifecycle.

For example, if a particular failure mode is associated with incorrect tags being applied to data, yet a sufficiently robust control to detect erroneous tags exists, then detectability will be high, resulting in a low score. (Recall that a low score is desired, since it reduces the overall product of impact, likelihood and detectability.)

Detective and corrective controls also improve detectability, limiting the timeframe or extent of any unauthorised access.

Risk Analysis: FTA

Using the FTA method, the first step is to identify the top events. Within the context of data protection and privacy, these will be the material adverse outcomes as noted above (page 26) – loss of privacy, loss of confidentiality, financial loss, strategic/operational impairment, personal injury/death (including moral harm), and reputational damage.

For each discrete top event, the immediate prior unwanted event/s (prior condition/s) on which the top event is dependent and the mechanism/s through which the prior

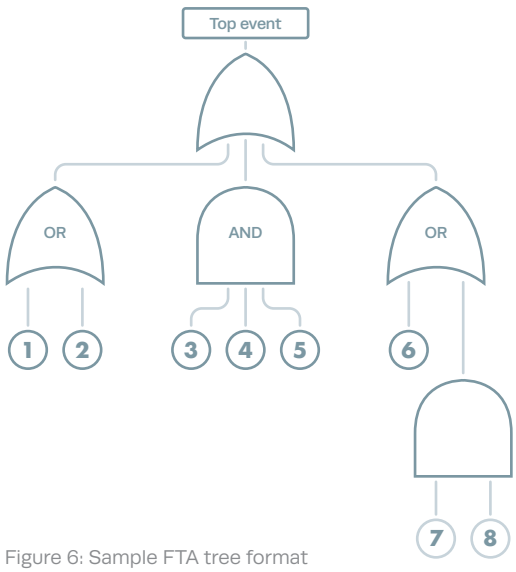


Figure 6: Sample FTA tree format

event/s are propagated into the top event are identified. These data are represented via logic gates in a tree format.

This process is continued to build the lower levels of the tree until the end of the process or no further material or plausible conditions are identified.

During the analysis process, relevant internal controls should be taken into account.

Probabilites can be assigned to the sub-event failures, and an overall probability of the top event occurring can be calculated. Previous experience of failures and/or expert input can be used to estimate respective values.

Securing reliable data in this field can be challenging; however, even in the absence of quantitative data, FTA studies are valuable sources of qualitative information in the way risks can propagate through the data analytic project.

Risk Evaluation

Whichever risk methodology is used, whether FMEA or FTA or another technique, the key steps once the risk analysis has been concluded are to review the results and determine and execute an action plan.

The outputs of the risk studies will comprise both quantitative data and qualitative insights. Both these dimensions should be reviewed in light of the context and risk appetite that has been established at the outset of the risk review phase. Decisions should be documented and include

- Go/no-go on the data project;
- Any necessary project modification to manage risk;
- Statement of appropriateness of control environment, including people, process, systems dimensions;
- Statement of risks by type (e.g. financial, reputational) relative to risk appetite; and
- Risk treatment plan.

In regard to the risk treatment plan, risks can be divided into three categories:

- **Unacceptable** – effective treatment essential
- **Grey area** – decide based on cost/benefit approach
- **Non-material** – risks are negligible relative to risk appetite and can be left unmanaged; however, these risks should not be allowed to drop off the radar

Risk Treatment

Treatment options have been discussed in Step 04 of the Data Project section above and include terminate, tolerate, transfer, and treat.

Leveraging ABAC Concepts to Augment FMEA and FTA Analyses

The FMEA risk analysis method relies on a documented data lifecycle process. Using the FTA model, prior events are identified on which subsequent events depend.

In either case, ensuring that all failure modes or prior dependencies are identified and analysed can be challenging. However, the attribute-based access control model can be leveraged to assist in this task.

Recall that ABAC-type policies use attributes related to the dimensions of subject, object, operation, and environment to define access authorisation. Consequently, unauthorised access to data can result from a failure or deficiency in any of these factors (attributes).

Even if an institution does not explicitly use ABAC-type policies and relies instead on, say, RBAC, these risks still exist within their firm. The risks are just not visible in the policies; they're latent within their organisation.

The dimensions of ABAC provide a framework for identifying potential failure modes or events on which top events may be dependent. An example, including the policy itself as a fifth dimension, is shown below.

ABAC Dimension	ABAC Policy Example	FMEA Failure Mode	FTA Condition
Policy	Users acting under the purpose of fraud prevention are allowed to access row-level data.	Data Governance department configured policy wrongly in ABAC system.	Wrong policy in ABAC system.
Object (Data)	Data access driven by sensitivity of data.	Data tagging process failure.	Sensitive data is erroneously tagged as non-sensitive.
Environment (Context)	Data (erroneously) asserted to not be subject to GDPR.	Failed to execute business process review.	Business process review not executed.
Operation (Process)	Cross-jurisdiction access permitted for fraud prevention purpose.	Project purpose wrongly cloned from anti-fraud project.	Data project wrongly defined as fraud prevention initiative.
Subject (User)	Data access driven by users' location.	User location tag wrongly updated.	User location attribute wrong.

Table 3: ABAC dimensions and FMEA vs FTA comparison

Furthermore, well designed and effective controls are necessary for each of the dimensions.

At each process step (FMEA) or prior failure condition (FTA), the ABAC dimensions should be considered to ensure that all potential failure paths have been considered.

For example, considering the last FTA condition in the above table – "user location attribute wrong" – how could this happen? One mechanism/condition may be that those managing the location attributes are not aware that this data element is being leveraged downstream in a business critical process and must be updated immediately when a user moves from one location to another. This is potentially a failure in the "environment/ context" dimension.

APPENDIX E

Risk Appetite

As defined by the Institute of Risk Management, a firm's risk appetite is the amount of risk that an organisation is willing to seek or accept in the pursuit of its long-term objectives.[25]

IN THIS SECTION

Following from this definition, an organisation's objectives are a key factor which must be considered in defining its risk appetite, and its risk appetite must remain congruent with its objectives as they inevitably change over time. There is a strong parallel with the Context topic as discussed as part of the risk management process within Step 04 and Appendix D.

Risk appetite provides the benchmark against which risks are evaluated, as described within the risk assessment Appendix D. Without a well-articulated risk appetite statement, the risk management process will terminate in a cul de sac.

Risk Appetite Ownership

All risks are owned within the business, and, ultimately, by the Board; this also applies to a statement of risk appetite. However, to be effectively managed, risk appetite must be articulated in ways to be relevant throughout the organisation at a strategic, tactical, and operational level.

In practice this will mean that if at a strategic level an institution has a low tolerance to unauthorised access to sensitive data, then this must be adequately articulated at both a tactical and operational level.

At a tactical level, this would be reflected in the type of data analytics projects the firm would be prepared to undertake and would be considered within its contextual risk management.

At an operational level, the risk appetite statement may take the form of the number of breaches or near misses that is tolerable and may, in this case, be zero.

Clearly, the different tiers of risk appetite and management are owned by different parts of the organisation.

25 Specific to the situation at hand, only those risks which are related to data privacy and protection are considered. Also, see https://www.theirm.org/media/3779216/64355_Riskapp_A4_web.pdf

Risk Capability

IRM's paper introduces the concept of "risk capability" based on an institution's "risk capacity" and "organisational maturity".[26] Risk capacity is the amount of risk an organisation can absorb without materially reducing its confidence in achieving its strategic objectives. In the context of this paper, organisational maturity is the overall soundness of a firm's organisational design, infrastructure, and internal control environment in respect to data privacy and protection.

The IRM asserts that an institution's risk appetite should be set with reference to its risk capability, and this paper adopts that view.

Risk Capacity

Dimensions of Risk

As defined in Appendix D, the dimensions of adverse outcome or harm with respect to data protection and privacy include

- loss of privacy
- loss of confidentiality
- financial loss
- strategic/operational impairment
- personal injury/death (including moral harm)
- reputational damage

Although, depending on the institutional context, there may be other risks, too.

Capacity vs Dimensions

Organisations must determine their risk capacity in respect to these dimensions. Clearly, some of these dimensions will be more contentious than others or may exhibit strongly asymmetric properties. For example, while adverse outcomes of a financial nature can be remedied with like, in the case of moral harm, there may be no effective remedies available.

26 https://www.theirm.org/media/3779216/64355_Riskapp_A4_web.pdf

Organisational Maturity

Clearly there is a strong overlap between those institutional features which were considered during the contextual analysis within the risk management section and those which determine organisational maturity. Building on context, the following dimensions should be considered:

- Culture
- Organisational design
- Control environment
- Risk management maturity
- Staff training and awareness

The table to the right shows the relationship between risk capacity and organisational capability.

From the perspective of the firm managing its risks, the optimal situation is that it is able to absorb significant risks (again, dependent on the dimensions of risk) and also have the organisational skills, experience, and resources to manage the risk adequately.

Having a high capacity for risk but lacking the organisational capability to manage it may result in a firm finding itself outside of its risk appetite without warning.

Firms who have invested to develop their organisational capabilities but lack the ability to absorb risk may need to revisit their business model and consider partnering with institutions who can complement their deficiencies.

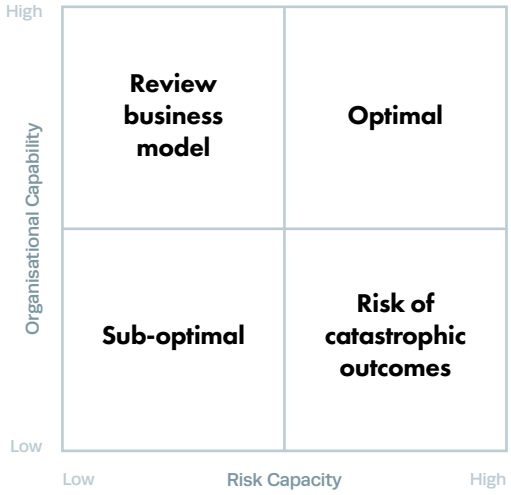


Figure 7: Risk capacity vs organisational capability.

APPENDIX F

Suitability - safety by design

Managing the data protection and privacy risks of a data analytic project is fundamentally a matter of how well-suited an organisation is to accomplish the project, its tolerance for the inevitable issues and incidents that will occur, and its skills in designing mitigation strategies and managing incidents.

These relationships are illustrated in the diagram below.

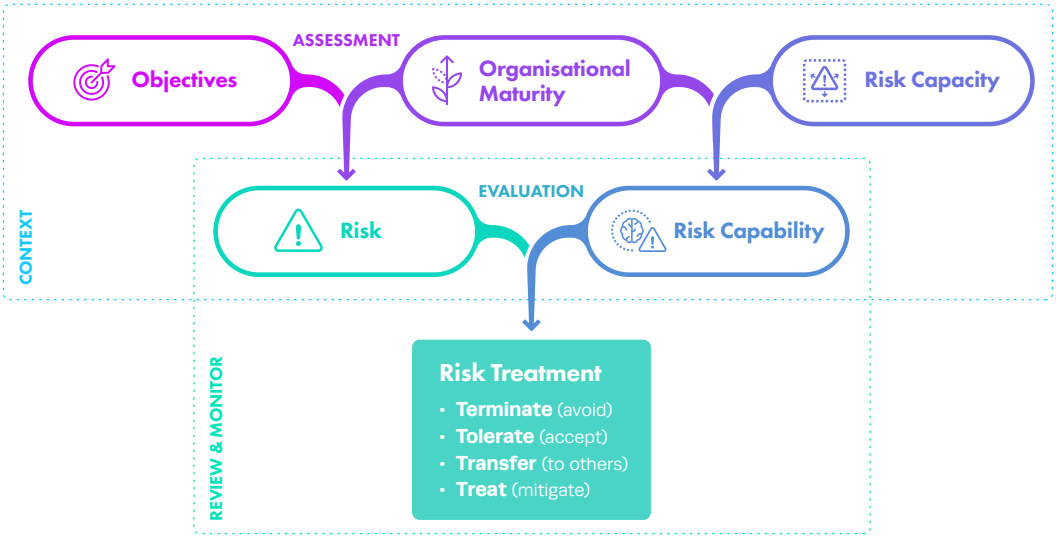


Figure 8: Risk assessment process

Given the objectives are documented and known, the context has been established, and the types of hazards and adverse outcomes that should be anticipated are clear, then the risk, represented by the red disc, is largely determined by the firm's ability to meet the demands of the data analytic project.

A high degree of suitability between the demands of the data analytic project and the firm's ability to rise to them will likely result in minimal residual risk relative to risk appetite, minimising the need for additional risk treatment.

This is the principle of safety by design: it is preferable to ensure suitability between the objectives, resources, and capabilities from the outset, rather than layer on *ex post facto* controls.

About Immuta

The Immuta Automated Data Governance Platform creates trust across security, legal, compliance, and business teams so they can work together to ensure timely access to critical business data with minimal risks. Its automated, scalable, no code approach makes it easy for users across an organisation to access the data they need on demand, while protecting privacy and enforcing regulatory policies on all data.

IMMUTA

7878 Diamondback Dr, Suite C, College Park, MD 20740 | immuta.com | (800) 655-0982