NISR
NATIONAL INSTITUTE OF STATISTICS OF RWANDA

RISA
Rwanda Information Society Authority

NATIONAL DATA CLASSIFICATION AND ACCESS GUIDELINE

Securing Data, Enabling Access, Advancing Development

List of Abbreviations

ABAC	:	Attribute-Based Access Control
AI	:	Artificial intelligence
CASB	:	Cloud Access Security Broker
CDO	:	Chief Data Officer
CKAN	:	Comprehensive Knowledge Archive Network
DGO	:	Digital Governance Officer
DLP	:	Data Loss Prevention
DPP	:	Data Protection and Privacy
DSA	:	Data Sharing Agreement
DUA	:	Data Use Agreement
GoR	:	Government of Rwanda.
IAM	:	Identity and Access Management
ID	:	Identifier
LMS	:	Learning Management System
MFA	:	Multi-Factor Authentication,
MoU	:	Memorandum of Understanding
NGOs	:	Non-Governmental Organizations
NISR	:	National Institute of Statistics of Rwanda
PII	:	Personally Identifiable Information
RBAC	:	Role-Based Access Control
SIEM	:	Security Information and Event Management

Table of Contents

Acknowledgment

The National Data Classification and Access Guidelines were developed following the approval of the National Data Sharing Policy and the National Data Governance Framework. These guidelines establish a coherent, government-wide approach to classifying, protecting, and granting access to public-sector data. They ensure that all data is managed in accordance with national standards that safeguard confidentiality, maintain integrity, and promote secure and responsible use.

The National Institute of Statistics of Rwanda (NISR) coordinated the development of these guidelines in close collaboration with the Rwanda Information Society Authority (RISA), the Ministry of ICT and Innovation (MINICT), and technical partners, including Cenfri. Their contributions were essential to ensuring alignment with Rwanda's Digital Transformation Agenda and the national vision for a secure, interoperable, and trusted data ecosystem.
NISR extends its appreciation to sector institutions, members of the National Statistical System (NSS), and all stakeholders whose insights and feedback strengthened the relevance and practicality of these guidelines.

The publication of the National Data Classification and Access Guidelines reaffirms Rwanda's commitment to strengthening data governance, enhancing data protection, and enabling responsible access to data as a strategic national asset that supports innovation, accountability, and sustainable development.

1

Introduction

Rwanda is making significant strides in digital transformation, guided by Vision 2050, with data management emerging as a strategic national priority. Effective data classification and access are essential to enable data-driven innovation and responsible adoption of emerging technologies such as artificial intelligence (AI) while preserving citizen privacy.

These guidelines establish a standardized framework for data classification, access, and protection of government data across the Government of Rwanda (GoR)[1]. It defines clear data classification levels based on sensitivity and introduces practical measures for managing data access, both internally and externally. These measures include well-defined roles, responsibilities, approval processes, and access controls, which safeguard sensitive information while enabling secure, lawful, and responsible data sharing.

Aligned with Rwanda's digitization efforts, cybersecurity standards, AI governance frameworks, and open data policies, the guideline wants to balance data accessibility with robust safeguards. Fostering a secure and trusted environment enables evidence-informed decision-making, improves public service delivery, and national development.

To address existing challenges in data classification and access, the National Institute of Statistics of Rwanda (NISR), as the custodian of national data governance [2]and government data stewardship[3], have developed these guidelines to provide clear and actionable protocols. These protocols promote interoperability, transparency, accountability, and trust in the use of government data. Furthermore, they ensure that all data is managed in accordance with ethical principles, existing legal and regulatory frameworks, and the evolving demands of a digitally empowered society.

1.1. Purpose of data classification and access guidelines

The data classification and access guideline serves as a strategic framework for organizing, processing, and accessing national data according to its sensitivity, value, and intended purpose.

The objectives are:

(1) To ensure secure data access , facilitate data sharing, enable effective utilization to support government operations, and foster innovation, while maintaining public trust.

---

1 Government of Rwanda (GoR) in this framework means all public institutions and government-affiliated entities in Rwanda, including ministries, departments, agencies, local governments, and state-owned enterprises.

2 The National Data Sharing Policy provides NISR responsibility for providing oversight, technical guidance, and institutional support for data governance activities across the Government of Rwanda (GoR) entities, in line with national guidelines, standards, and applicable legal frameworks.

3 In this context, "data steward" NISR is legally mandated as a custodian of government data. While it is not the owner of all institutional datasets, it holds the responsibility to establish and enforce standards, safeguard data integrity and quality, promote interoperability, and oversee the responsible management and use of data across public institutions to support evidence-informed policymaking and national development.

(2) To define categories of data and outline roles, responsibilities, and access controls,
This guideline promotes:

- *National data security and privacy protection*
- *Efficient data governance and interoperability*
- *Compliance with legal and regulatory requirements*
- *Digital transformation and AI readiness*
- *Data sharing within and across institutions*
- *Foundation for secure, ethical, and informed decision-making, enabling the government to manage data as a valuable national asset.*

1.2 Scope and Coverage

These guidelines apply to GoR and any contracted entities entrusted by GoR with managing data in Rwanda, including Government-owned enterprises, where applicable, and third-party contractors, consultants, and partners operating under formal agreements with the Government of Rwanda. They ensure robust protection of data throughout its lifecycle, with access granted exclusively on a need-to-know and least privilege basis. It mandates secure handling, access controls, encryption, and continuous monitoring to prevent unauthorized access or disclosure. All stakeholders must comply with defined roles, adhere to security policies, and undergo regular training to safeguard national interests.

Similarly, these Data classification and access guidelines apply to all categories of government-managed data, including administrative records, surveys, censuses, and big data.

2

Data classification

Data classification is conducted according to data sensitivity, value, and associated risk, guided by the two principles of the CIA Triad, namely Confidentiality (Preventing unauthorized access or disclosure) and Integrity (Maintaining the accuracy and reliability of data).
This process enables institutions to implement effective data management, to protect sensitive information, and ensure compliance with legal and regulatory requirements.

2.1 Data Classification Levels

Data classification levels are categories used to organize data based on sensitivity, value, and the risk associated with unauthorized access. These levels help determine how data should be handled, stored, and protected. Typical classification levels include **Public**, **Internal**, **Confidential,** and **Restricted (Highly confidential)**. The details are clearly displayed on the table below.

Table 1: National data classification Levels

Level	Description	Examples	Access Control & Handling
Public	Data intended for unrestricted public access. Disclosure poses no risk to individuals, institutions, or national interests.	Anonymized survey microdata Aggregated survey results for public release Open administrative datasets with anonymized data	No restrictions; can be freely shared and published. Basic integrity checks recommended.
Internal	Data is meant for internal use within an institution. Unauthorized disclosure may cause minimal operational or reputational risk. Can include multiple **sub-levels**: • Low: Routine administrative statistics • Medium: Draft statistical tables or internal analysis • High: Internal datasets used for policy formulation	Draft survey tables not yet validated Internal department-level administrative statistics Preliminary census summaries or internal socio-economic indicators	Accessible only to institutional staff. Login credentials required. Basic IT protections, such as network controls. Higher internal levels may require restricted folders, role-based access, or encryption.

Level	Description	Examples	Access Control & Handling
Confidential	Sensitive data. Unauthorized access could cause harm to individuals, institutions, or partners. Disclosure may affect operations, policy decisions, or trust.	Statistical microdata with limited identifiers Internal research datasets containing sensitive administrative variables Administrative data linked across departments for analysis	Limited to specific departments or roles. Encryption is recommended for storage and transmission. Role-based access and access logging required. Periodic audit of usage.
Restricted (Highly Confidential)	Data with the highest sensitivity. Unauthorized access could result in significant harm, legal liability, or breach of privacy.	Raw survey microdata containing personal identifiers Health statistics at the individual or household level Administrative datasets with Personally Identifiable Information (PII) Sensitive economic or policy-related datasets before official release	Strict access control with pre-approved roles only. Encryption at rest and in transit. Multi-factor authentication and secure access environments. Access logs and monitoring are required. Compliance with data protection laws and regulations.

2.2 Data classification principles

These principles provide a clear and unified approach for classifying and safeguarding government data across the GoR. They are designed to ensure that GoR:

- Manages data responsibly, with appropriate security measures,
- Maintains transparency,
- Supports national development objectives,
- Enhances public trust, and
- Enables digital transformation initiatives.

Principle 1: Understand the Nature of the Data

Before classifying any dataset, institutions must assess its characteristics to determine how it should be handled, stored, protected, and shared. This assessment ensures the correct classification, appropriate safeguards, and compliance with legal, ethical, and organizational requirements.
Key considerations include:

- **Data Type and Structure:** Identify whether the data is structured or unstructured, as this affects storage, processing, and protection.

- **Ownership and Custodianship:** Determine the Data Owner, responsible for legal accountability, access decisions, proper use, and the Data Custodian, responsible for operational management and safeguarding of the data.
- **Sensitivity:** Identify personal, confidential, or sensitive information requiring special protection.
- **Source:** Consider the origin of the data (e.g., surveys, administrative systems, third-party providers), which affects reliability and compliance obligations.
- **Potential Risks:** Evaluate legal, security, and reputational risks associated with the collection, processing, or sharing of the data.

Implementation Guidelines

1. **Data Inventory[4]:** Create and maintain a comprehensive record of all datasets, including source, format, ownership, and sensitivity. Classify datasets as personal, operational, or institutional. Review and update the inventory regularly to ensure accuracy and completeness.
2. **Data Catalog[5]:** Develop a searchable and user-friendly data catalog that draws from the inventory. Include enriched metadata such as dataset description, classification, lineage, usage rules, and access controls. The catalog enables users to discover datasets, understand their context, and access them responsibly while maintaining security and compliance.
3. **Purpose and Usage Assessment:** Define why each dataset was collected, how it is used, and by whom. Specify whether it is intended for internal use or external sharing, ensuring usage aligns with the data owner's directives.
4. **Metadata Documentation:** Record key metadata for each dataset, including source, ownership, classification, and sensitivity. Assign responsibility for managing and safeguarding metadata to ensure accountability and traceability.

Principle 2: Open by Default, Protected When Needed

Data collected and managed by the GoR institutions should be made publicly accessible whenever possible to promote usability, innovation, transparency, and accountability. At the same time, access must be restricted when datasets contain sensitive, personal, or confidential information that could compromise individual privacy, organizational integrity, or national interests.

Implementation Guidelines

1. **Define Public vs Restricted Data[6]:** Clearly identify which datasets can be shared publicly and which require restricted access based on sensitivity and legal obligations.

2. **Classify and Protect Sensitive Data:** Apply national data classification levels to sensitive datasets and enforce appropriate access controls.

---

4 Data Inventory is a record of all datasets in an organization, listing their source, format, ownership, and sensitivity.

5 Data Catalogue is a centralized inventory describing datasets, their sources, owners, quality, and usage.

6 Data are classified based on four classification levels (Public, Internal, Confidential, and Restricted (Highly confidential)) as highlighted in Table 1.

3. **Regular Review of Access Restrictions:** Periodically reassess and update access restrictions to ensure they remain valid and appropriate.

4. **Ensure Transparency:** Make data-sharing agreements, access protocols, and classification rules publicly available where possible to build trust and accountability.

Principle 3: Classify Based on Risk, Sensitivity, and Lifecycle

All data must be classified as early as possible, either at creation or upon receipt, according to its sensitivity and the risks associated with misuse, unauthorized access, or disclosure. Classifications should be regularly reviewed and updated throughout the data lifecycle to ensure continued relevance and protection.

Implementation Guidelines

1. **Apply National Classification Levels**: Use the standard levels (Public, Internal, Confidential, and Restricted) to categorize datasets based on sensitivity and risk.

2. **Conservative Approach for Mixed Datasets:** For datasets containing multiple types of information or uncertain sensitivity, apply the highest applicable classification (confidential or restricted) to ensure protection.

3. **Integrate Classification into Systems:** Embed classification protocols into data management and recordkeeping systems to enforce consistent application.

4. **Periodic Reassessment:** Review and, if necessary, reclassify data at fixed intervals at least every 24 months or whenever there is a significant change in content, context, or risk.

5. **Use Automation Where Feasible:** Implement automated classification tools to support and complement human judgment, improving efficiency and reducing errors.

Principle 4: Define Roles and Ensure Accountability

Institutions must clearly assign responsibilities for data classification, access management, and usage monitoring. Clear role definitions ensure accountability, promote consistent data practices, and reduce the risk of errors, misuse, or mismanagement.

Implementation Guidelines

1. **Assign Data Roles:** Designate Data Owners to oversee classification and authorize access, Data Custodians to manage and safeguard data operationally, and Data Stewards to maintain data quality and integrity.
2. **Separate Duties:** Ensure that responsibilities are divided appropriately to prevent conflicts of interest and maintain checks and balances.
3. **Capacity Building:** Provide regular training and guidance to staff to reinforce understanding of their roles, responsibilities, and accountability in data management.

More details on roles and responsibilities are further defined in the National Data Governance Framework, which GoR institutions should adopt and adhere to.

Principle 5: Limit Access through Need-to-Know and Least Privilege

Access to data should only be granted to staff whose current assignment requires such data and only at the necessary access level. This reduces the risk of misuse or accidental exposure. In the context of data sharing, this principle should be applied together with the Principle of Data Minimisation, meaning that only the minimum amount of data required for the specific purpose should be shared or accessed.

Implementation Guidelines:

1. Implement role-based access controls across all systems.
2. Enforce the "need-to-know" and "least privilege" principles.
3. Regularly audit access permissions and remove outdated accounts.
4. Require a highly secured access environment (on-premises or controlled), approvals, and logs for accessing restricted or confidential data.
5. Apply multi-factor authentication for all sensitive data access.

Principle 6: Ensure Consistency Across Institutions

GoR must follow the same classification rules, using consistent labels and standards. This ensures secure and reliable data sharing between entities.

Implementation Guidelines:

1. Adopt national classification levels.
2. Integrate classification guidelines into policies and data and information management systems.
3. Train and strengthen all staff who are involved in data classification and access management.
4. Establish a central registry of classification standards and updates
5. Conduct regular cross-institutional compliance assessments

Principle 7: Document and Monitor All Classification Actions

All classification decisions and data access must be fully traceable. Institutions should document who classified data, the rationale for the classification level, the date of classification, and monitor the usage of classified data.

Implementation Guidelines:

1. Maintain **classification records** in metadata repositories or data registries
2. Log all access, sharing, or modification of classified data.
3. Conduct regular audits to ensure compliance.
4. Utilize dashboards, reports, and automated monitoring tools to identify unusual access patterns.

Principle 8: Protect Privacy and Uphold Ethical Use

Data that identifies or affects individuals must be handled fairly, and with respect for individual rights, in full compliance with Rwanda's Data Protection and Privacy Law (No. 058/2021). Data classification should never be used to avoid accountability or to enable discrimination.

Implementation Guidelines:

1. Conduct privacy impact assessments for personal data.
2. Apply anonymization or pseudonymization where possible.
3. Ensure ethical data utilization, particularly in AI, statistics, and service delivery.

2.3 Classification Decision Matrix

This matrix helps institutions decide the appropriate classification level by assessing key risk and sensitivity criteria. It aligns with the four-level framework: Public, Internal, Confidential, and Restricted.

Table 2: GoR Classification Decision Matrix

Criteria	Public	Internal	Confidential	Restricted
Impact of Disclosure	No harm.	Minor inconvenience or internal disruption.	Legal, financial, privacy, or reputational harm to individuals or institutions.	Severe damage to national security, public safety, or core institutional integrity.
Legal Requirements	Legally open; no restrictions.	Access controlled by institutional discretion or internal policy.	Subject to national data protection and statistical laws.	Protected under statutory provisions related to national security, classified information, or critical infrastructure.
Stakeholder Impact	Enhance transparency and public trust.	Neutral or limited impact outside the institution.	It could negatively affect individuals' or institutional credibility.	Catastrophic to individuals, government, or diplomatic relations.
National Security	No impact.	No impact.	Moderate to high potential impact.	Critical national security relevance.
Privacy Concerns	None (fully anonymized or aggregated data).	Minimal (non-sensitive identifiers).	Directly identify individuals (PII, health, or financial data) or sensitive business information.	Extremely sensitive personal, operational, or protected identity information, including intelligence or defense-related data.

3 Data Access Control

3.1 Access rights by classification level for different users

Access rights are determined according to the sensitivity of the data and its designated classification level. This approach ensures that information is protected proportionately to the potential risks of unauthorized disclosure, while enabling appropriate sharing to support operational efficiency and transparency. Each classification level, Public, Internal, Confidential, and Restricted, has clearly defined access rules, authorized user groups, and security controls to safeguard the integrity, privacy, and lawful use of the data.

Table 3: Access rights by data classification level for different users

Classification	Internal Staff	External Institutional Users (Gov/Public Entities)	Independent External Users (Researchers, NGOs, etc.)
Public	Full access	Full access via official platforms	Full access via portals or request
Internal	Relevant staff only	Access by formal request + MoU/Approval	Access only if anonymized/ pseudonymised and formally approved
Confidential	Strict role-based access	Access only under legal agreement or national mandate	Access only if legally mandated, and data are anonymized/pseudonymised
Restricted (Highly Confidential)	Clearance only (specific authorization)	No access unless explicitly defined by law	Strictly prohibited

3.2 Conditions for External Data Access

Before sharing data externally, data must be classified and anonymised or pseudonymised where applicable.

A. External Government/Public Institutions:

The national data sharing policy provides the modality for data sharing between GoR institutions

1. Submission of a Data Sharing Request Form.
2. Provision of a justification and intended use for the data.
3. Execution of a Data Sharing Agreement (DSA) or Memorandum of Understanding (MoU).
4. Access is granted exclusively through security systems or portals.
5. Further redistribution of data is prohibited without authorization.

B. Independent External Users (Researchers, NGOs, Partners, etc.):

1. Public or open data are freely accessible via official websites or the NISR microdata portal
2. For non-public data, users are required to submit a data access request form to NISR, clearly stating the purpose and intended use.
3. Users must provide a clear purpose for research or a public interest justification
4. A Data Use Agreement (DUA) between the data user and data provider must be signed, specifying the scope, duration, confidentiality obligations, and publication rules. Only anonymized, aggregated data may be shared unless special official approval is given.
5. NISR will facilitate access on behalf of GoR through a secure environment via a remote access solution that allows users to analyze the data in collaboration with the data owners.
6. Requested data must be used only for approved purposes

3.3. Risks and Challenges

Effective data sharing and access require proactive risk management. The following are key risks and their implications:

- Privacy violations occur when personal data is disclosed without proper authorization or utilized beyond its originally intended purpose, thereby potentially compromising individuals' fundamental privacy rights and undermining their reasonable expectations of confidentiality.
- Data Breaches are security weaknesses, including vulnerabilities in third-party systems, that can lead to unauthorized disclosure of sensitive or personal data.
- Loss of Control: Once data is shared externally, GoR loses oversight of how and where it's used. External sharing doesn't always involve physical data transfer between organizations, but when it does, clear protocols and audits should be in place.
- Compliance Risk, Failure to adhere to relevant laws, regulations, or contractual terms regarding data handling may result in legal penalties or sanctions.

Reputational damage is the misuse or improper management of data by internal or external parties that can damage public trust and harm the institution's credibility

3.4. Implementation of Data Access and the Responsible Role

Each defined role plays a vital part in managing data access. The Data Owner is responsible for approving who can access specific data and the justification for all access requests. The Data Custodian implements technical access control to enforce data owner authorization decisions. The Data Steward ensures that access aligns with institutional policies and actively monitors usage.
The Data User is responsible for proper handling of authorized data and compliance with all applicable data sharing agreements. The Chief Data Officer (CDO) or similar role provides oversight, ensuring policies are followed and access is appropriately audited. Finally, the Data Auditor reviews overall compliance and suggests necessary improvements to strengthen access control practices.

4 Roles and responsibilities in data classification and access

The national data governance framework provides clear guidance on roles and responsibilities as a foundation for effective data classification and access management. This ensures ownership, accountability, and decision-making across all data functions. While institutional arrangements may differ depending on mandate, size, and capacity, GoR institutions should integrate these roles into their operational structures and document them accordingly. Functions may be combined if needed, provided they remain well-defined and effectively managed.

Each institution will establish a data team responsible for providing strategic oversight of data governance activities. The team will be chaired by the Chief Data Officer (CDO) and will include the Officer, Data Owners, Data Stewards/Data Governance and Data Custodian.

In the context of this guideline, the data team will:

- Approves organisation-wide data classification policy and access standards
- Resolves escalated access disputes
- Ensures alignment with national laws, regulations, and institutional requirements
- Provides strategic oversight and accountability

The table below presents a summary of the key roles and their responsibilities as recommended in the national framework.

Table 4: Roles and responsibilities in data classification and access management

Role	Key Responsibilities
Chief Data Officer (CDO)	• Provides overall leadership for data governance across the institution • Ensures classification and access align with strategic objectives • Escalates unresolved issues to the DGC • Reports regularly on compliance, risks, and improvements
Data Owner(s)	• Determines classification level of datasets (Public, Internal, Confidential, Restricted) • Approves or denies access requests based on business/ statistical need • Ensuring access rights are role-appropriate • Reviews access rights periodically

Role	Key Responsibilities
Data Steward(s)/ Data Governance Officer (DGO)	• Overseas day-to-day implementation of the classification and access framework • Ensures integration with risk, compliance, and IT security programs • Coordinates between Data Owners, Custodians, and Stewards • Provides technical/procedural guidance to staff • Tracks compliance and reports to the CDO • Applies classification labels as defined by Data Owners • Maintains classification metadata in registers/catalogues • Monitors compliance with access rules • Ensures correct handling of data according to classification • Requests access through approved channels when required
Data Custodian(s) / IT Admin	• Implements/enforces technical access controls (role-based permissions, encryption, logging) • Ensures secure storage, backup, and transmission aligned with classification • Defines security controls per classification level • Audits access logs and monitors for breaches • Reports unauthorized access or misuse immediately
Data User(s)	• Accesses and uses data only for approved purposes • Completes with classification, access, and handling rules • Reports on incidents or suspected misuse • Protect data according to training and institutional policies

5 Compliance and Monitoring in data classification and accessibility.

Compliance and monitoring are essential elements of a strong data governance program. They ensure that sensitive information is classified correctly, protected from unauthorized access, and used in line with institutional legal and regulatory requirements. Without clear compliance and monitoring practices, organizations risk data breaches, privacy violations, and loss of trust.

5.1. Compliance Activities in Classification and Access

Compliance activities focus on setting clear rules, implementing effective controls, and maintaining accountability in how data is classified and accessed. These activities establish the foundation for protecting sensitive data while meeting regulatory obligations.

1. Policy Definition: Define data classification policies based on sensitivity, business value, and compliance needs. Establish access control policies using Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), or rule-based models.
2. Data Discovery and Labeling: Deploy automated tools such as Data Loss Prevention (DLP) systems and cloud-native data discovery tools to detect, classify, and label data according to content and context.
3. Access Control Implementation: Enforce access through Identity and Access Management (IAM) systems, including Role-Based Access Control, Multi-Factor Authentication (MFA), and time-bound or context-aware permissions.
4. User Access Reviews: Conduct regular reviews of access rights. Remove unnecessary or inactive accounts to reduce compliance risks.
5. Audit Logging: Maintain logs of data access events, permission changes, and system configuration updates. Ensure logs comply with legal, regulatory, and biometric data requirements to support effective monitoring and compliance.

5.2. Monitoring Mechanisms for Compliance

Monitoring mechanisms ensures that compliance activities remain effective over time. They provide ongoing oversight, detect potential security risks, and validate that policies and procedures are being applied consistently. By combining real-time monitoring with regular audits and automated tools, institutions can maintain a proactive approach to compliance.

1. Real-Time Monitoring: Use Security Information and Event Management (SIEM) tools to track unauthorized access, detect anomalies in data usage, and generate security alerts.
2. Periodic Audits: Conduct internal and external audits to confirm data classification accuracy, proper enforcement of access control policies, and resolution of logged incidents.
3. Automated Tools: Apply Data Loss Prevention systems to prevent unauthorized data sharing and Cloud Access Security Broker (CASB) solutions to monitor and enforce security policies for cloud-

based data access.

5.3 Best Practices

Data Classification

Effective data classification ensures that information is properly categorized and managed according to sensitivity and business value. Best practices include the use of automated classification tools enhanced by artificial intelligence (AI) and machine learning (ML), conducting periodic reviews and reclassification of data, and ensuring consistent and visible labeling through headers, footers, or metadata tags.

Access Control

Access to data must follow the principles of least privilege and just-in-time authorization. Implementing zero-trust architecture ensures verification is required before granting access, including for internal requests. A centralized Identity and Access Management system should be maintained to enforce consistent access policies across the organization.

Monitoring

Monitoring mechanisms should provide visibility across all systems and be integrated with incident response workflows. Logs from multiple systems should be correlated to detect anomalies, and regular compliance assessments must be conducted to update controls based on findings.

5.4. Documentation and Evidence for Audits

GoR, preparing for compliance audits, must maintain comprehensive records. Key documents include:

1. A metadata documenting all data assets and their classifications.
2. An access control matrix specifying who can access what and the justification for each permission.
3. Access logs and audit trails capturing all relevant data activity.
4. Risk assessments and mitigation plans highlighting potential vulnerabilities and planned responses.
5. Training records showing staff education on data handling policies.
6. Policy documents and enforcement evidence demonstrating consistent application of governance rules.

6 Data Linkage and Integration

Data linkage and integration refer to the process of combining datasets from one or more sources, either within a single government entity or across multiple institutions, to create enriched datasets that support evidence-based policymaking, monitoring, and research. While the benefits of this process are significant, including improved insights and efficiency, the activity involves handling sensitive and often confidential information. As a result, it requires strict governance to safeguard privacy, ensure security, and maintain data integrity

6.1. Governance Principles

For statistical purposes, Data linkage and integration projects are coordinated by the National Institute of Statistics of Rwanda (NISR) on behalf of the Government. This approach helps ensure that linked datasets are accurate, consistent, and reliable for producing official statistics, while also safeguarding confidentiality and complying with data protection standards. By coordinating these activities, NISR promotes standardized methodologies, avoids duplication, and ensures that statistical outputs are coherent and useful for evidence-based policymaking and public reporting. Institutions or researchers wishing to undertake such projects are invited to submit a standardized Data Linkage and Integration Request Form, detailing the project objectives, motivation for the analysis, datasets involved, compliance with legal requirements, and the security measures to be applied, such as anonymization or pseudonymization. Requests are reviewed in line with national statistical policies and best practices.

6.2 Execution of Linkage and Integration

All approved data linkage operations will take place in secure environments managed or supervised by NISR. Only authorized and trained personnel with the appropriate clearance may perform the linkage. Where necessary, remote secure access solutions may be used, ensuring that linked and integrated datasets do not leave NISR's controlled environment.

6.3. Register of Linkage Projects

NISR will maintain a national registry of all approved data linkage and integration projects. This registry will document the purpose and description of each project, the datasets and institutions involved, the approving authority and date of approval, the custodian of the linked dataset, and the conditions for access. The registry will promote transparency, strengthen accountability, and prevent duplication of effort across institutions

6.4. Ownership and Stewardship

Ownership of the original datasets remains with the contributing institutions. Linked datasets, however, are considered derivative datasets and are governed jointly by the contributing government entities, with NISR serving as custodian. Access to these linked datasets will strictly follow the classification and access control rules outlined in this guideline.

6.5. Safeguards and Security Measures

Before release or use, all linked datasets must undergo a formal risk assessment to evaluate potential threats such as re-identification, sensitivity, or confidentiality breaches. Linked data may only be shared outside NISR's secure environment in anonymised or pseudonymised form, and always under formal agreements that define conditions of use. Any misuse, unauthorized disclosure, or security breach must be reported immediately and addressed in accordance with the national data governance framework.

7

Annexes

Glossary of terms

1. **Authorized Individuals** means organizational personnel, contractors, or third parties with authorized access to the information system in which the organization has the authority to impose rules of behavior regarding system access.
2. **CDO** means the Chief Digital Officer, or a person appointed to fulfil that role, responsible for ensuring that information security guidelines and procedures (including this Data Classification Guideline) are implemented across the Government of Rwanda (GoR) institutions, public institutions, and partner institutions.
3. **Data Classification:** The process of categorizing data according to its sensitivity, value, and potential impact to the institution if compromise occurs.
4. **Public Data**: Information approved for open access by the public (e.g.: marketing materials, press releases). No restrictions on dissemination.
5. **Internal Data** Non-sensitive data designated for internal use only (e.g.: policies, internal memos). Unauthorized disclosure may cause minimal risk.
6. **Confidential Data** Sensitive data requiring access restrictions (e.g.: employee records, internal reports). Unauthorized disclosure may cause harm to individuals or the institution.
7. **Restricted Data:** Unauthorized access could result in legal penalties or severe reputational/financial harm.
8. **Personal Data and Sensitive Data** as legally defined under Article 3 of Rwanda's Data Protection and Privacy (DPP) Law (Law N° 058/2021) of 13/10/2021. Unauthorized access, disclosure, or misuse of such data may result to legal penalties as stipulated by this law, as well as severe reputational and financial harm.
9. **PII (Personally Identifiable Information)** Any data that may directly or indirectly identify a natural person. Classification aligns with Personal Data.
10. **Data Owner** The individual or department responsible to define data classification levels, establish access controls, and protection measures for a specific dataset.
11. **Data Custodian** IT or security personnel responsible for implementing and maintaining security controls (e.g: encryption, access permissions) as defined by the Data Owner.
12. **Encryption** The process of converting data into a secure format to prevent unauthorized access during storage or transmission.
13. **Data Breach** Unauthorized access, disclosure, or loss of classified data, whether accidental or malicious.
14. **Access Control**: Policies, procedures, and technologies used to ensure that only authorized individuals or systems can access specific data or systems. This includes authentication, authorization, and enforcement mechanisms.
15. **Access Monitoring:** The continuous process of observing and tracking data access events in real-

time to identify unauthorized or abnormal activities. This includes reviewing access logs, tracking login attempts, and identifying potential security breaches.

16. **Access Request**: A formalized procedure whereby a user submits a request to gain access for specific data or systems. This request typically includes access level specification, justification for access and duration of access.
17. **Audit Trail**: A record of all actions and events related to data access, including who accessed the data, when, and what actions were taken. Audit trails are used for monitoring, accountability, and security auditing.
18. **Authentication**: The process of verifying the identity of a user or system attempting to access data or resources. Methods of authentication include passwords, biometric verification, and multi-factor authentication (MFA).
19. **Authorization**: The process of determining whether a user or system has the right to access a specific resource or perform a certain action after their identity has been authenticated.
20. **Compliance**: Adhering to legal, regulatory, and institutional standards or requirements for data security and privacy.
21. Comprehensive Knowledge Archive Network (CKAN): It is an open-source data management system designed for open data portal, data storage and distribution of open data.
22. **Data Breach:** An incident where unauthorized individuals gain access to sensitive or confidential data, potentially leading to data loss, exposure, or misuse.
23. Data Classification: The process of categorizing data into different levels based on its sensitivity, value, and importance.
24. Data Disposal: The process of securely deleting or destroying data when it is no longer needed or when retention periods have expired. This is crucial for preventing unauthorized access to obsolete or sensitive information.
25. Data Encryption: A security process that converts data into an unreadable format using algorithms, ensuring that unauthorized individuals cannot access or understand the data. It applies to data in transit (during transmission) and data at rest (stored data).
26. Data Owner: The individual or entity responsible for the management, classification, and security of a specific set of data.
27. Data Retention: The practice of retaining data for a specific period, based on legal, regulatory, or business needs, after which it should be securely archived or destroyed.
28. Faceted search: It is a method of searching through data by using facets, i.e. attributes in the data, to gradually filter down a large selection of data to a smaller one that we are looking for.
29. Federated identity: It is a method of linking a user's identity across multiple separate identity management systems. It allows users to quickly move between systems while maintaining security.
30. **Multi-factor Authentication (MFA):** A security mechanism requiring multiple verification factors to access a system, such as something they know (password), something they have (security token), or something they are (fingerprint or face recognition).
31. **Role-based Access Control (RBAC):** A method of regulating access to resources based on the roles of individual users within an institution. Each role has specific permissions associated with it, and users are assigned to roles according to their job functions.
32. **Administrative data:** Records related to institutional operations and internal management.
33. **Statistical data:** Conducted and compiled by the National Institute of Statistics of Rwanda and

sector-specific statistical units.

34. **Operational data:** Supporting service delivery, mandates, and institutional functions.
35. **Financial and procurement data:** Including budgets, payments, contracts, tenders, and financial statements.
36. **Official communications:** Internal and external communication such as reports, emails, meeting records, and correspondence.