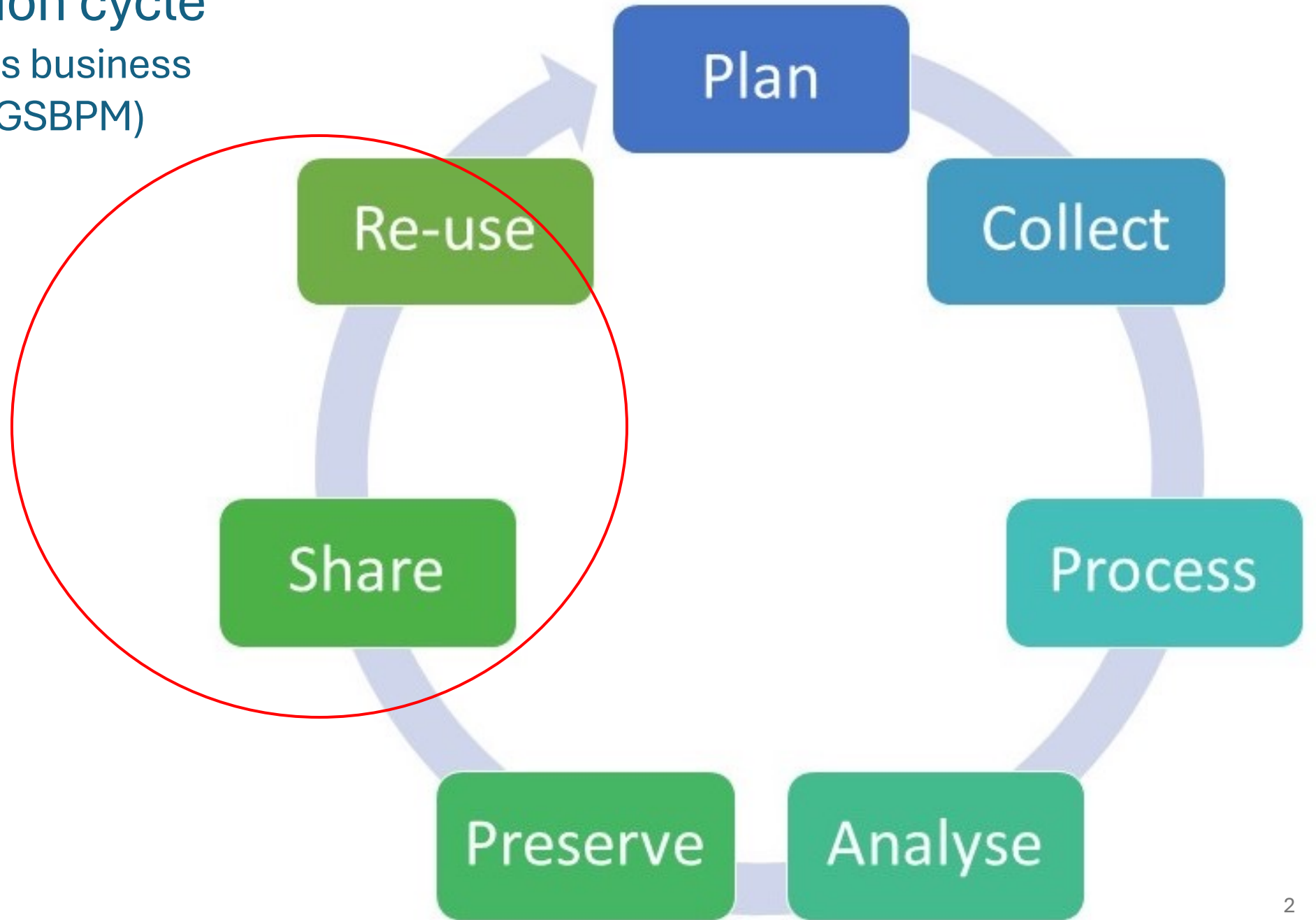# Data Governance

Understanding Data anonymization (Data protection and privacy)

Data collection cycle
Generic Statistics business process model (GSBPM)

P4 Processes

1. Data is not only an **input**; but also **output** of e-government

2. Data is used in **both front- and back-office** of e-government

3. **Some data are used; many are not,** including those generated through e-services (administrative)

4. Some data are not used **optimally**; some data are also **misused**

5. While there is a **lack of data**, there is also **data and information overload**

6. Government's quadrupole role: **producer, consumer, regulator, and platform provider (enabler)**

Reuse
Plan
Share
Colle[c]
Preserve
Process
Analyse

Why disseminate data

...to inform decisions

Elicit immediate action

Share new information or insights

Educate about recent findings or accomplishment

Justify program activities

Promote behavior change

Solicit support or participation

Document magnitude of health problem

Prepare for an upcoming intervention or program

Types of data disseminations

- Hard-copy dissemination (Analytical reports, tables)
  - Workshops and awareness
- Websites and data portals (e,g SPCs Pacific Data hub,)
- Machine-to-machine (M2M) refers to computers exchanging data via application programming interfaces (APIs)
- Multimedia dissemination include CD-ROMS, DVDs and USB keys
- Mobile apps
- GIS Portals - spatial attributes

Vanuatu legal framework

- The 2016 RTI law gives a legal basis to the right to information and requires all government agencies to release information. It also establishes a complaints mechanism if agencies do not release information.

- 2023 National Data Protection and Privacy Policy aims to create the foundations of a data protection and privacy law that will give effect to associated rights enshrined in the Constitution to ensure respect for, and the protection of, personal data and associated rights of individuals, and in particular the right to privacy.

2023 Statistics Act provides for Access, Disclosure and protection of information

- ❏ **Part 3**:Statistics Advisory Council (SAC)
- ❏ **Part 4**: Access, Disclosure and protection of information.
- ❏ **Part 5**: Collection of information for census and survey
- ❏ **Part 6**: Confidentiality of information and undertaking secrecy
- ❏ **Part 7**: Offences

General Data Protection Regulation (GDPR)

Adopted by many countries to "harmonise" data privacy laws across all of its members countries as well as providing greater protection and rights to individuals

Microdata

- Refers to data about an individual, person, household, business, community or other entity.
- Microdata may be data collected by surveys, censuses or obtained from administrative records (Health, Education, VNPF, Customs, Telcos, NDMO.
- Microdata are collected using valuable resources of a country.
- Microdata are usually of significance yet access to countries' microdata is often limited by the fear that **confidentiality** protection cannot be guaranteed.
- Microdata must be used and reused.

Data Protection:

The process of safeguarding important information from corruption, compromise, or loss. It involves implementing various measures, such as encryption, access controls, and backups, to ensure the confidentiality, integrity, and availability of data.

Data protection often involves complying with regulations and standards to ensure that sensitive information is handled securely.

Data Privacy

Focuses on the appropriate handling of personal data. It concerns the ethical and legal considerations surrounding the collection, storage, and sharing of individuals' personal information.

This includes obtaining consent for data collection, informing individuals about how their data will be used, and providing them with options to control their data.

Microdata Anonymization/Data de-identification

Data anonymization involves altering or removing personally identifiable information (PII) from datasets to prevent the identification of individuals.

The goal is to retain the utility of the data for analysis while minimizing privacy risks.

Confidentiality

- Refers to **limiting** data access and disclosure to authorized users and **preventing** access by or disclosure to unauthorized ones
- Confidentiality is also related to the broader concept of data privacy -- limiting access to individuals' personal information.

Confidentiality

UN Fundamental Principle of Official Statistics.
"Individual data collected by statistical agencies for statistical compilation, whether or not they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."

Data confidentiality

- How to make sure confidentiality
  - Legal arrangements
    - Legislation be in place before any microdata are released.
    - Legally enforceable agreement may be one of the requirements of access. It should be possible to set up an arrangement where a prior agreement needs to be signed even where access to microdata is **available online**.
  - Technical arrangements
    - The risk of confidentiality can be reduced by the use of a number of techniques.
    - The downside is that these techniques may reduce the usefulness of the underlying microdata.

Data anonymization

- Anonymization is process of removing or modifying the **identifying variables** contained in the microdata dataset.

- Identifying variables are those describe characteristic of a person

  - *Direct identifiers,* which are variables such as names, addresses, or identity card numbers.

  - *Indirect identifiers,* which are characteristics (residence, age, sex, and profession) that may be shared by several respondents, and whose combination could lead to the re-identification of one of them.

Microdata anonymization techniques

- **Data Reduction**
  - Removing variables: The first obvious application of this method is the removal of direct identifiers from the datasets (race, religion, income, etc.)
  - **Removing records**: Removing records can be adopted as an extreme measure of data protection when the unit is identifiable in spite of the application of other protection techniques.
  - **Global recoding**: The global recoding method consists in aggregating the values observed in a variable into pre-defined classes (for example, recoding the age into five-year age groups, or the number of employees in three-size classes: small, medium and large).

Why disseminate microdata

- Support and diversify research work (tabular data only answer predefined questions)
    - Maximize use and usefulness of data
    - Increased return on investments; more use → funding
- Reduced duplication of data collection
- Improved data quality, higher credibility
    - Relevance, reliability through feedback from users
    - Transparency and replicability as scientific safeguards
- Satisfy a legal requirement, and/or a contractual obligation with sponsors

Costs and risks of microdata dissemination

- Confidentiality / privacy protection are key to maintain trust from respondents; must preserve reputation and comply with the law

- Exposure to criticism

- Non-official estimates vs official estimates

- May have to respond to inquiries

- Financial cost (data preparation, anonymization, dissemination)

- Building and maintaining technical expertise

Thank yu tumas