Data Ecosystem Governance

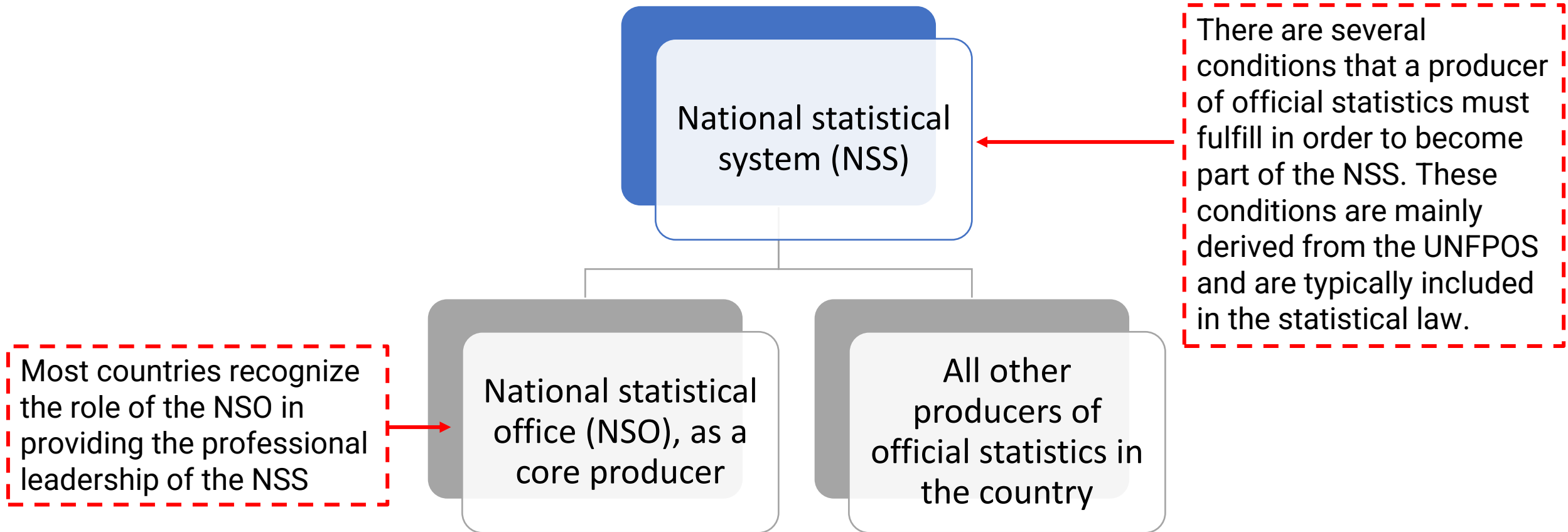# Data Ecosystem

Data Ecosystem

is the entire network of:
- **data collectors,**
- **data producers,**
- **data analysts and**
- **other main data users**

that directly or indirectly collect, process, disseminate, analyze and/or otherwise consume data and associated services within a specified country or region.

NSS
- National Statistical Office (Main producer of official statistics)
- Other producers of official statistics

- Administrative data and registers
- Other data community (Big data, Geospatial, NGOs...)
- Users

National Statistical System

```
                    National statistical
                    system (NSS)
```

```
National statistical        All other
office (NSO), as a          producers of
core producer               official statistics in
                            the country
```

There are several conditions that a producer of official statistics must fulfill in order to become part of the NSS. These conditions are mainly derived from the UNFPOS and are typically included in the statistical law.

Most countries recognize the role of the NSO in providing the professional leadership of the NSS

Data sources

Surveys and censuses are the traditional data source and are still very much used.

Geospatial data means data that have a geographic component. The most prolific source of such data is satellite imagery.

Survey

Administrative data

Geospatial data

Big data

Data collected by a government ministry, department or agency primarily for administrative purposes.

Big Data refers in general to data generated by business transactions, social media, phone logs, communication devices, web scraping, sensors etc.

UN environment programme

Census

Pros of a CENSUS

- provides a true measure of the population (no sampling error)

- benchmark data may be obtained for future studies

- detailed information about small sub-groups within the population is more likely to be available

Cons of a CENSUS

- may be difficult to enumerate all units of the population within the available time

- higher costs, both in staff and monetary terms, than for a sample

- generally, takes longer to collect, process, and release data than from a sample

UN
environment
programme

Surveys

Pros of a SAMPLE

- costs would generally be lower than for a census

- results may be available in less time

- if good sampling techniques are used, the results can be very representative of the actual population

Cons of a SAMPLE

- data may not be representative of the total population, particularly where the sample size is small

- often not suitable for producing benchmark data

- as data are collected from a subset of units and inferences made about the whole population, the data are subject to 'sampling' error

- decreased number of units will reduce the detailed information available about sub-groups within a population
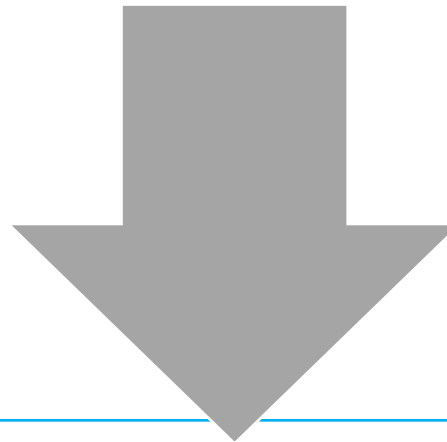
UN
environment
programme

Administrative data

Administrative data
refers to data collected by a government ministry, department or agency primarily for administrative (not research or statistical) purposes
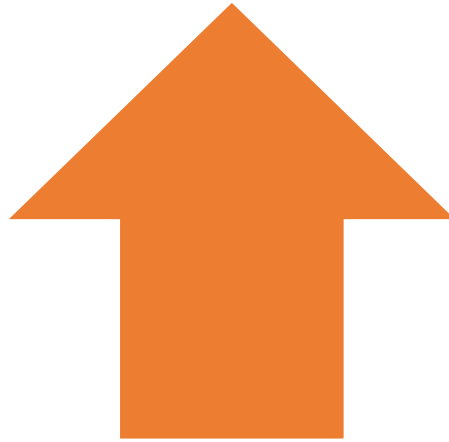
Advantages:

- Cost-effectiveness

- Reduced response burden

- Timeline and frequency

- Coverage and completeness

- Relevance

Disadvantages:

- Differences in units, concepts and definitions of variables

- Differences in classifications

- Inadequate data quality

- Lack of statistical know-how

- Need for additional data checking

UN
environment
programme

Other data sources

❖ Geospatial data refers to data that has a geographic component.

- A geographic information system (GIS) is a system designed to capture, store, manipulate, analyse, manage, and present geospatial data.

❖ Big Data refers in general to data generated by business transactions, social media, phone logs, communication devices, web scraping, sensors etc.

- Big Data is widely used in the commercial sector for business analytics, but there is less evidence of its use thus far in the world of official statistics.

UN
environment
programme

Data, statistics, indicators

Indicators

Statistics

Data

UN environment programme
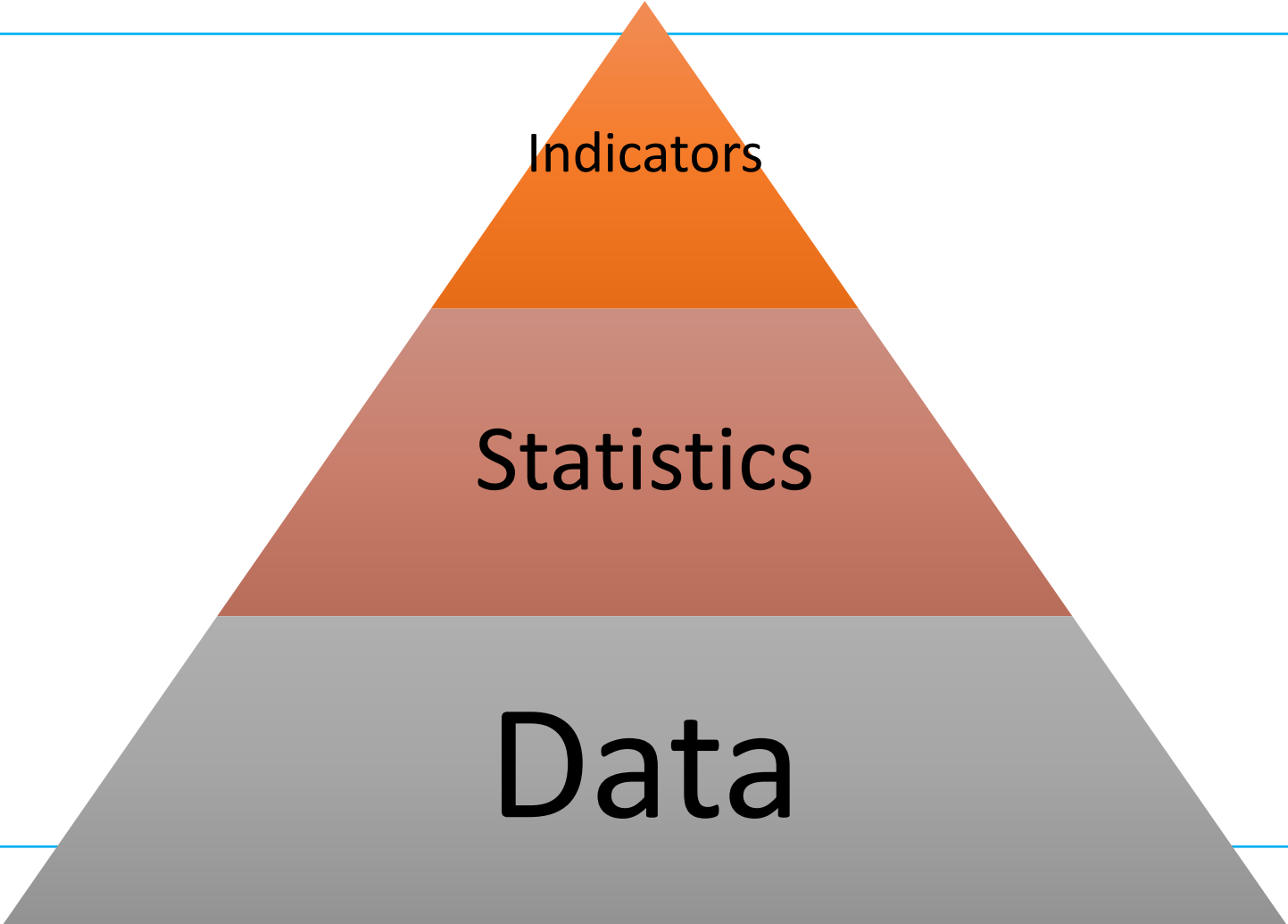
Quality of official statistics

❖ The quality of official statistics must be respected at all stages of the production process, as well as the final product.

❖ Over the past 20 years, NSOs have arrived at the consensus that quality is multidimensional and that **there is no single measure of quality**.

❖ At the same time, NSOs and other producers of official statistics produce many statistical products every year. These products are required to fulfil several criteria as regards their relevance, impartiality, accuracy, and timeliness. These criteria are derived from the UN Fundamental Principles of Official Statistics, codes of practice based on the UNFPOS, specific international standards as well as statistical laws of the different countries.

UN
environment
programme

UN Fundamental Principles of Official Statistics

- ❖ Principle 1 – Relevance, impartiality and equal access
- ❖ Principle 2 – Professional standards, scientific principles and professional ethics
- ❖ Principle 3 – Accountability and transparency
- ❖ Principle 4 – Prevention of misuse
- ❖ Principle 5 – Sources of official statistics
- ❖ Principle 6 – Confidentiality
- ❖ Principle 7 – Legislation
- ❖ Principle 8 – National coordination
- ❖ Principle 9 – Use of international standards
- ❖ Principle 10 – International cooperation

UN environment programme

UN National Quality Assurance Framework for Official Statistics

- At the global level, the **UN National Quality Assurance Framework for Official Statistics** (NQAF) was first adopted by the UNSC **in 2012**

- The NQAF was subsequently revised by the UNSC with the issue of the **NQAF Manual in 2019**.

  - ✓ The Manual offers comprehensive guidance for an NSO on adopting and operating a quality assurance framework to help implement quality management policy.
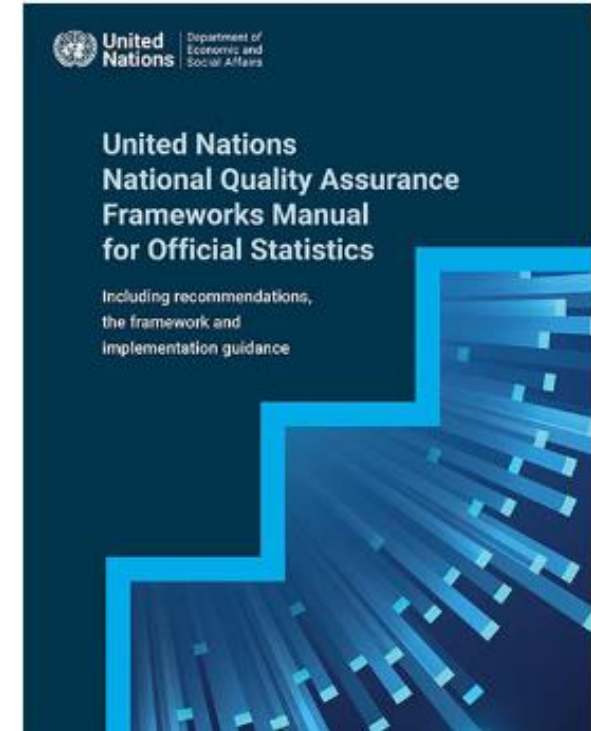
here

African Charter on Statistics

- The **African Charter on Statistics** was adopted in 2009 and entered into force in 2015

- It presents six quality principles expressed in the form of 25 quality statements, adapted to the African situation

Here

Professional independence

Quality

Mandate for data collection and resources

Dissemination

Protection of individual data, information sources and respondents

Coordination and cooperation

Metadata

The first and most fundamental purpose of **metadata** is to help users of statistical data to interpret, understand, and analyse statistical data.

Metadata provides information about the background, purpose, content, collection, processing, quality, and related information of a statistical dataset that a user needs to find, understand and manipulate statistical data.

All published or released statistics should be accompanied by metadata.

UN environment programme
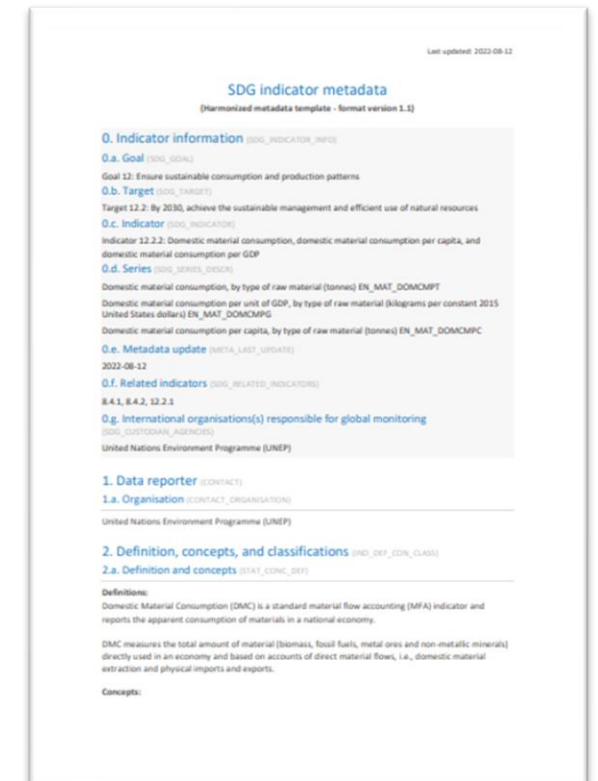
Managing statistical data and metadata

Data and metadata management involves the management of documents and records, archiving, managing knowledge, standards and access rights as well as metadata and data management.

A data and metadata management policy should consider the entire statistical production process from data collection to dissemination and archiving *(derived from the Generic Activity Model for Statistical Organizations (GAMSO))*:

UN
environment
programme
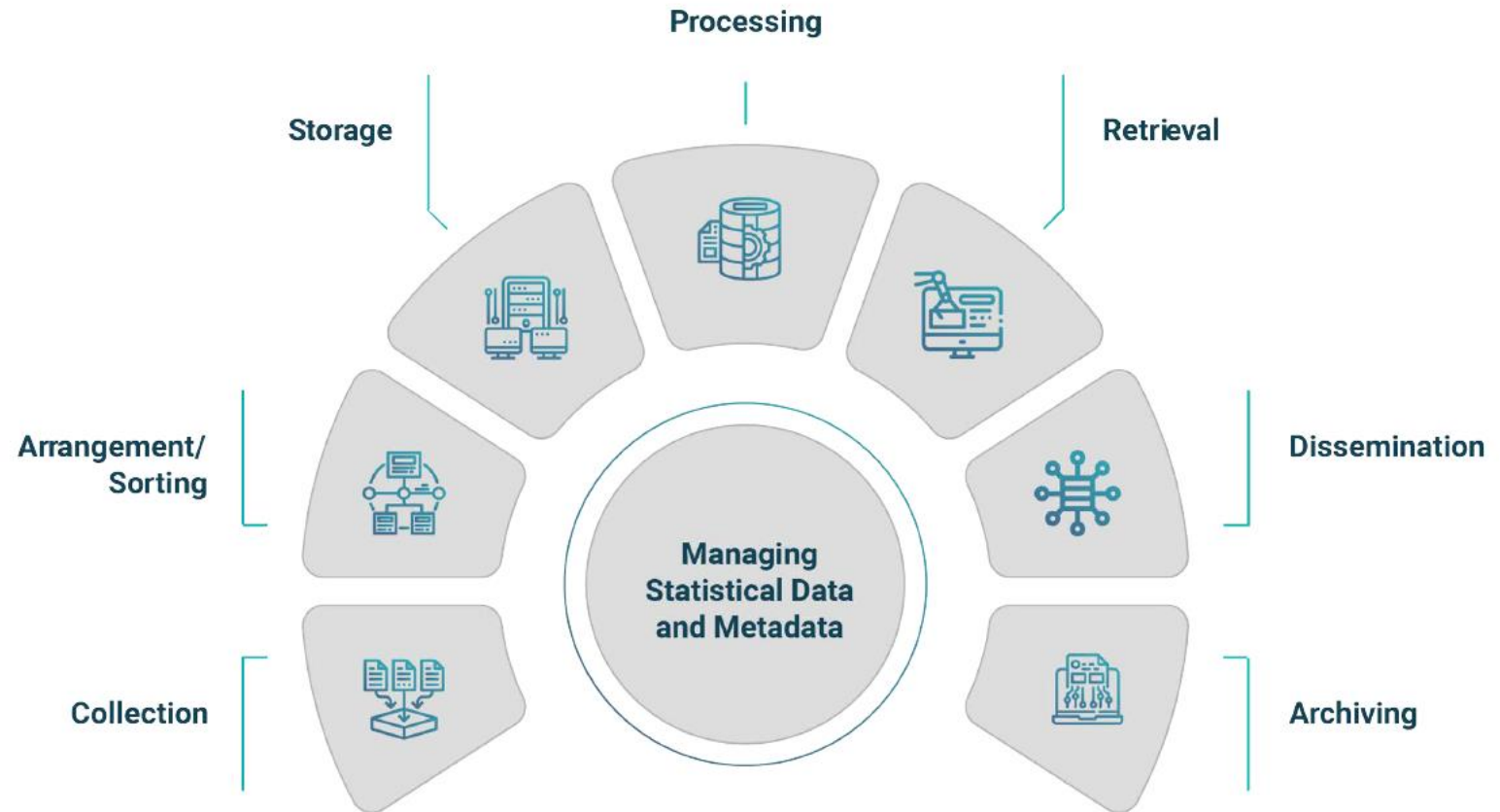
Data users

National Data Ecosystem:

- Privacy and Data Protection Act

Governs personal data collected, processed and/or shared by public and private operators
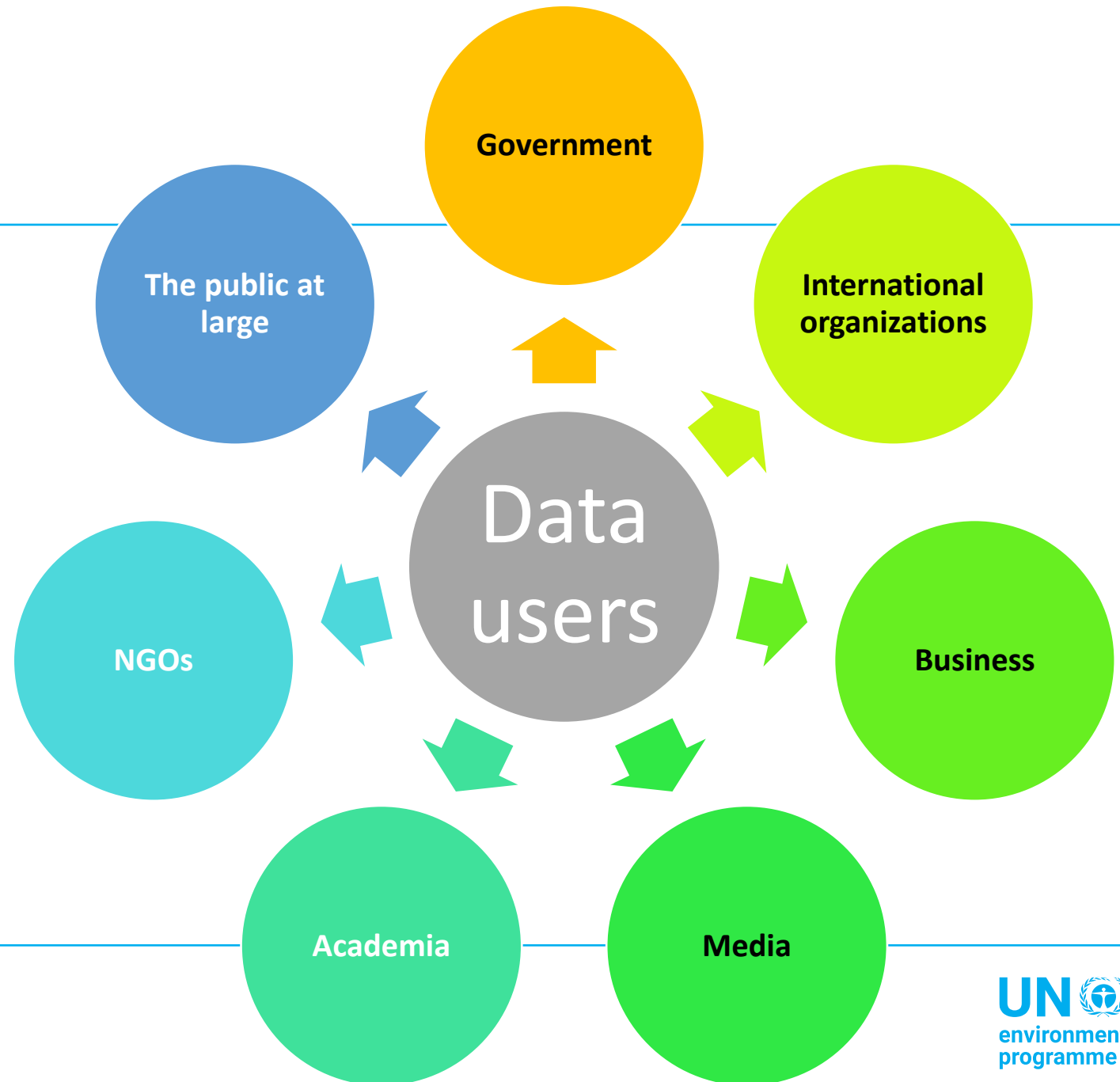
- Public Information Access Act

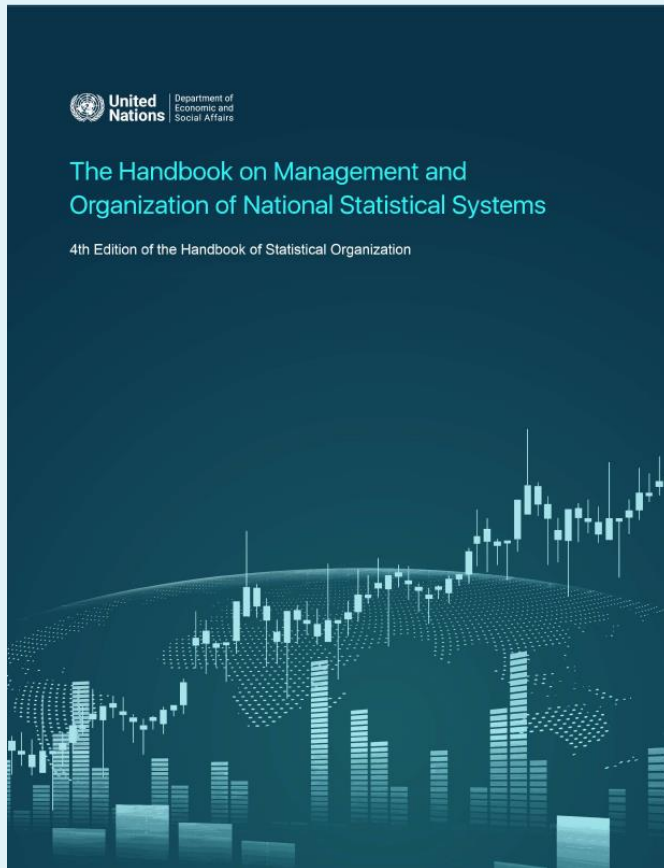Governs the procedure which ensures free access to and right to reuse information held by public authorities

- Archiving Act

Rules the procedure for archiving data of national interest

- Code of Data Ethics

UN
environment
programme

More information

The Handbook on Management and Organization of National Statistical Systems

4th Edition
UN, 2022

Thank you

UN environment programme

www.unep.org