"Capacity Building and Strengthening Institutional Arrangement / Data Yearbook"

Workshop: "How to produce an Environmental Data Year Book"

# How to prepare the indicators data

## Ms. Mariaconcetta Giunta, Mr. Giovanni Finocchiaro, Ms. Cristina Frizza, Mr. Luca Segazzi

APAT

Agency for Environmental Protection and Technical Services

# Introduction

What input data are required for generating the indicators?

Data acquisition
- External sources
- Self production

Qualification and validation

Representation

# DATA ACQUISITION

**External source**

Do the data I'm looking for already exist?

Who owns them?

How can I get them?

**Sources Catalogue**

# The source catalogue

- Source name → f.i. "National Statistic Office"

- Survey stations → number & dislocation

- Documents → reports/articles/CD/internet references

- Programmes & Projects → data gathering and

- Datasets → accessible data collections

- Contacts → references

# DATA ACQUISITION

**Self production**

Which office should gather the data I need?

Does it actually arrange data gathering?

**Data sharing system**

# The internal data sharing system

## Features

•Data informative contents → environmental matrix or item

•Data owner → the office which owns data

•Data format → magnetic or optic support, papery support, others

•Data contents

•References → people to be get in touch with, in order to ask for the  data if they are not immediately available

# GATHERING METHODOLOGIES

- Monitoring

- Surveys

- Administrative documents

- Statistical models

# Monitoring

**AIM** → environmental qualitative data surveys

## Common Features

- Sample survey;

- Survey stations intercalibration;

- Surveying frequency.

# BRACE database

In Italy the BRACE database includes information about nets, stations and sensors used to monitor air quality and pollutants concentration data.

# Surveys

A survey is a method  of gathering/collecting data from local/national institutions, industries and private subjects in order to produce and elaborate statistical information.

The results coming from different surveys could be elaborate in order to obtain a wider informative content.

An elaboration is a procedure to get new information, rearranging available statistical and non statistical data, coming from former surveys

**Questionnaires are** the main tools used in a survey.

**AIM** → environmental quantitative data surveys

## Features

- Conciseness

- Clearness

- Completeness

# MUD & SACP

**MUD** (Unified Model of Declaration)

- Italian official system to gather waste data

- questionnaire split into two parts: Urban waste and Hazardous and Non-hazardous waste

**SACP** (Structure of Agriculture Companies & Production)

- No environmental data (at the beginning)

- Integration of environmental questions

Two choices: to build the **whole questionnaire** or take advantage of a **pre-existent questionnaire**.

# Administrative Documents

**AIM** → to find environmental quantitative and qualitative data

Examples: Authorizations…

Data supplying → environmental data such as: water and energy consumptions, pollutions emissions, entity of the resources, loss of resources, territorial context data, others.

# Statistical Model

**AIM** → to estimate data

- No monitoring system

- No way to use a questionnaire

- No available or satisfying data

Note: a statistical model can only provide a rough picture of the phenomena.

# Statistical Model

**Key concepts**

An **estimator** is any quantity calculated from the sample data which is used to give information about an unknown quantity in the population.

An **estimate** is the particular value of an estimator that is obtained from a particular sample of data and used to indicate the value of a parameter.

# Corineair

Corineair is a statistical model born to estimate atmospheric pollutants emissions.

In Italy it have been successfully used in the construction of the Italian Greenhouse Gas Inventory.

# DATA QUALIFICATION AND VALIDATION

Quality of data (statistics) is defined by Eurostat with reference to the following six criteria:

- relevance;

- accuracy;

- timeliness and punctuality;

- accessibility and clarity;

- comparability;

- coherence.

**Relevance**
Relevance is the degree to which data meet current and potential users' needs. It refers to
whether all statistics that are needed, are produced and the extent to which concepts used
(definitions, classifications etc.) reflects user needs.

**Accuracy**
Accuracy in the general statistical sense denotes the closeness of computations or estimates to
the exact or true values

# Timeliness and punctuality

Punctuality refers to the time lag between the release date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar, laid down by Regulations or previously agreed among partners.

Timeliness of information reflects the length of time between its availability and the event or phenomenon it describes.

# Accessibility and clarity

Accessibility refers to the physical conditions in which users can obtain data: where to go, how to order, delivery time, availability of micro or macro data, various formats (paper, files, CD-ROM, Internet…), etc.

Clarity refers to the presence of appropriate metadata, illustrations, whether information on their quality are also available (including limitation in use…) and the extend to which additional assistance is provided.

# Comparability

- Over time

- Between geographical areas

- Between domains

Comparability **over time** refers to comparison of results, derived normally from the same statistical operation, at different times.

The **geographical component** of comparability emphasises the comparison of statistics between countries and/or regions in order to ascertain, for instance, the meaning of aggregated statistics at European level.

Comparability **between domains** refers to non-geographical domains, for instance between industrial sectors, between different types of households, etc.

**Coherence**

Data adequacy to be combined in different ways and for various uses.

Quality → **multi-faceted concept**.

Even if data is accurate, they cannot be said to be good if they are produced too late, or cannot be easily accessed, or appear to conflict with other data.

The most important quality characteristics depend on user perspectives, needs and priorities as well as available economic resources.