

## Annex 5. Documenting tourism statistics sources at sub-national levels

### Presentation

Because tourism statistics include a wide range of data produced by different type of institutions (both at a national and subnational levels), there is a need for standards for the presentation of *metadata* (which comprise data and other documentation that provide information about the process of producing and using data).

In fact, users are often confronted by some difficulties of interpretation when comparing statistics compiled over time within one agency and by agencies in different countries as well as by different international organizations:

- conceptual differences arising from the use of different variable definitions, units and classifications;
- operational differences flowing out of differences in data collection and processing practices by countries, and;
- different practices in the presentation of data.

This Annex pursues two objectives that complement each other:

- the first one is to help regions to develop a useful database regarding not only the sources of information from which data integrating the Regional System of Tourism Statistics but also other relevant information not necessarily of statistical nature (see chapter 3, Para. 3.3.);
- the second one is to share part of this documentation at international level in such a form that it could be easy to consult, to exchange experiences and to compare the procedures used in the different regions so that best practices might be derived.

For some regions, this project might seem too ambitious or somewhat unnecessary as the objective might seem to be to rapidly put data together and get some idea of what tourism looks like. Nevertheless, UNWTO's experience is that too much haste, using data whose quality has not been sufficiently assessed, might lead to difficult situations in which low-quality data, used carelessly, might show a picture of tourism that is quite far from reality, compared to what more solid statistical work could show. This, in turn, might generate a general distrust towards the whole process of estimation of the economic impact of tourism.

These guidelines are geared toward identifying and discovering existing sources, the information a region has for measuring the economic impacts of tourism activity, the quality of the available data, etc. In other words, the guidelines have to do with the documentation relative to the coverage, temporal reference, distribution and an entire set of technical characteristics of the methodology used in obtaining the data used building up Regional Tourism Information Systems.

The project designed by UNWTO and shared by INRouTe is a medium-term project, and aims at documenting information from two complementary perspectives. It concerns:

- existing statistical and non-statistical source of Regional-Tourism Information Systems. The outcome will be an "Inventory of data sources";
- the type and nature of available information viewed from the perspective of the main tourism variables to be measured. The outcome will be an "Inventory of available data";
- the two perspectives entail a twofold inventory: on sources and available data.

### I. The inventory of data sources

Normally, the term *survey* is used in a general sense to refer to any activity that collects or acquires statistical data. Included as such are:

- a *census*, which attempts to collect data from all members of a population;
- a *sample survey*, in which data are collected from a (usually random) sample of population members;
- collection of data from *administrative records*, in which data are derived from files originally collected for non-statistical purposes;
- *derived statistical activity*, in which data are estimated, modelled, or otherwise derived from existing statistical data sources.

This Annex respected this typology, albeit with a slight terminological adaptation:

- two types of surveys are considered: *surveys applied to travellers or visitors*, and, eventually, household or *surveys applied to enterprises / establishments of those productive activities serving visitors*;
- the term *statistical syntheses* is used to refer to those sources (such as Regional Accounts/Input-Output or other models) that could be included in the category of *derived statistical activity*;
- also included, as another type of source, are *databases* as well as *statistical publications on tourism*, as supports used for the storage/dissemination of the data obtained for the other mentioned sources.

Consequently, seven types of data sources are recommended:

- **T.1 *Surveys applied to the whole population of travellers or visitors or to clearly predefined segments of this population.*** These surveys (derived from the demand perspective) typically occur at the national borders for inbound and outbound visitors; at hotels or other places of collective accommodations; at tourism attractions, or can be modules of household type surveys, etc.
- **T.2 *Surveys applied to enterprises / establishments of those productive activities serving visitors.*** All of them share a supply perspective and might be general purposes surveys, or surveys specifically designed to supply information on tourism productive activities. Usually, it is the National Statistical Office, which shall be in charge of general-purpose surveys encompassing also tourism characteristic activities, whereas the National or Regional Tourism Administrations might develop specific surveys for the activities under its direct responsibility.
- **T.3 *Statistics based on data collection from administrative records.*** These usually concern passenger transport and border control statistics, registration of guests in hotels, etc.
- **T.4 *Census or directories.*** Censuses are complete enumerations of the target population). It must be mentioned that the fact for a variable to be based on a Census, that is, on an exhaustive enumeration, is not a guarantee of having a perfect estimation of the variable as there are many reasons for omission and duplication which have to be identified (errors of observation) and which importance has to be measured.
- **T.5 *Statistical synthesis.*** This category includes those sources (basically Regional Accounts and models) which elaboration requires the use of information stemming from some of those previously enumerated.
- **T.6 *Databases.*** These can be from a unique source or combine various sources.
- **T.7 *Statistical publications on regional tourism.*** These may incorporate quantitative as well as qualitative data from various sources.

A standard format is proposed although it needs to be slightly adapted to each of the seven types of data sources already defined.

**INVENTORY OF DATA SOURCES: BASIC FORMAT**

**Country**

**Region**

**The statistical procedure**

Title

Data source

Description

**Administration**

Institutional framework

    Institution in charge

        Unit responsible within this institution

        Other units and/or institutions involved

Data dissemination

    Name of publication

    Published by

    Frequency

    Length of time series available

    Reference period for the latest available publication

    Timeliness

    Access to data via the Internet

**Coverage**

Geographical coverage

Frequency

Reference period for the first procedure

Reference period for the last procedure conducted

**Concepts**

Specific subjects

Variables of reference

**Data collection, manipulation / accounting conventions, etc.**

Frame

Type of enumeration

Sampling

Distribution of the sample over time

Data collection principles

Geographical breakdown

Reporting units

Other issues

- Geocoding

- others

**Observations on its use**

**Additional documentation**

## ***II. The inventory of available data***

It has been mentioned that this proposed inventory is geared toward identifying the existence or not of a basic set of data and indicators that are essential for the analysis of tourism activity.

The proposal would be to structure a matrix where the rows contain the selected tourism variables and the columns contain the corresponding available information sources. The corresponding cells should be marked to identify where there is information in any or all the existing sources, and the periodicity of the data should also be noted (monthly, quarterly, every six months, annual, or no predefined periodicity).