

2 DATA FLEXIBILITY

Environmental data record phenomena in the physical environment. Some of these recordings are factual, for example the grid reference of the place where a species was observed, the dimensions of a tree, the weight of a log, the annual precipitation at a site, or the absorptive capacity of a soil profile. These are all **primary data** based on facts which can be measured against stable, widely accepted standards.

Secondary or derived data are obtained from primary data by a process of classification or interpretation, either at the time of measurement or later. Examples include species name, vegetation type, forest canopy extent and climatic zone. Derived data are not a substitute for primary data, and should not be stored permanently unless the primary data used to create them are also available. This is because derived data slowly degrade in value and, ultimately, become useless as concepts and paradigms shift. For example, if the only data on species distribution is an outline drawn on a map, this may become redundant if the species is split or otherwise disaggregated following a taxonomic revision. A better approach would be to store the primary data relating to the identification and location of the species in the field, so that new outlines can be derived as necessary.

Primary data are much more flexible than derived data. They can be used for a wider range of applications because they have not been modified for a specific function. For instance, daily rainfall measurements in millimetres from a local weather station can be used to assess local climatic fluctuations. They can also be fed into national or international climate monitoring programmes, or be integrated with other data to assess the capability of an area to support biodiversity. If the rainfall measurements had been classified at time of collection into, say, five secondary categories (e.g. very low, low, medium, high, very high), then this flexibility would have been lost, resulting in fewer potential applications for the data (such categories may be too coarse or meaningless in other contexts).

To ensure that data remain flexible, they should be collected and stored in their primary form, not classified, aggregated or otherwise interpreted. However, this rule does not need to be implemented rigidly; it may be subjected to intelligent assessment in each case. No one, for example, would refuse to store the names of species, even though they are susceptible to change. The process of deciding which type of data to store involves risk assessment. Given the high costs of collecting and managing data, the benefits of doing so needs to be balanced against the risk that data will become obsolete or prove to be inflexible. To assist with this judgement, Box 1 highlights the

key characteristics of primary and derived data, and compares these with the information that both can be used to generate (see Volume 3).

It should be noted that perceptions of primary data and derived data vary considerably, according to the particular individuals or organisations concerned. For example, derived data to a scientific researcher may be regarded as primary data by a policy-maker, and be subjected to further analysis and interpretation. Despite differences in perception, the principle of storing primary data holds true within any particular domain, although between domains it may not.

Box 1 The nature of data and information

- **Primary data**

These are facts which result from measurements or observations about the world, referenced to stable, widely accepted standards. The latter include absolute measures, such as units of length, volume or density.

- **Derived data**

These are data obtained from primary data by a process of classification or interpretation, either at the time of measurement or later. They may be referenced to absolute measures but more commonly relate to professionally agreed conventions and products, for example maps which comply with an accepted structure and format.

- **Information**

This is altogether different to data: it is the knowledge derived from the analysis, integration and interpretation of data, including 'expert' opinion. Unlike data, which may be applied to a range of purposes, information is produced for a specific purpose and has a short shelf life. Because of its transient nature, information should not be stored in databases unless this is judged to be cost-effective.