

2 BASIC PRINCIPLES

As with other complex issues relating to the management of scientific information, effective progress requires the underlying principles of data custodianship to be understood and widely accepted. These principles, which are presented below, can then guide operational developments:

- Data should be managed by the organisation in the best position to do so.
- Data should be managed cost-effectively by staff who understand the data: what it purports to represent, what its characteristics are, how it was collected, what quality-assurance procedures have been applied and its limitations.
- Data should not be duplicated or fragmented in different places.
- Data should be widely available to those that have a need to make better informed decisions.

The above principles are not meant to be interpreted as rigid rules. There may be sound operational reasons why, for example, a copy of a dataset may be duplicated in another place, such as in a network hub (see Volume 4) for purposes of distribution. This may be done because the custodian is not accessible on-line, which could impede access to the original dataset. However, the risk (and cost) of violating custodianship principles needs to be understood. In the above example, the risk is that the distributable copy will become out of date, while the cost implications are that the hub will have to monitor developments in the original dataset and periodically obtain an updated copy.

The key is to manage data in such a way that they can be converted into a variety of information products, for a variety of users, thus ensuring that they are flexible enough to respond to the demands of decision-making. Ideally, every dataset has a recognised custodian although, if priorities need to be made, the custodians of **essential datasets** should be determined first. These datasets are sufficiently vital to the day-to-day operation of an organisation or network that they justify the effort and expense incurred in their collection, storage and quality-assurance (see Volume 3).

Essential datasets underpin the development of multiple information products, for multiple users, and are, thus, permanent, or at least of a lasting nature. Non-essential datasets, on the other hand, are produced by *ad hoc*, undocumented processes for quick results, and may be transient or of uncertain quality (after Janzen 1993, 1995). Unfortunately, many organisations manage essential datasets as if they were non-essential datasets.