Chapter 2
Definition of Environmental Data

The focus of the quality system requirements in EPA Order 5360.1 A2 is on environmental data. In the past, this was often misunderstood to simply mean chemical measurement data collected in the field or in a laboratory, and most quality system documents focused almost exclusively on procedures for assessing the quality of such data. The latest Agency-wide order concerning quality makes it clear that the quality system must address more than just measurement data.

Environmental data - Any measurements or information that describe environmental processes, location, or conditions; ecological or health effects and consequences; or the performance of environmental technology. For EPA, environmental data include information collected directly from measurements, produced from models, and compiled from other sources such as data bases or the literature. EPA Order 5360.1 A2, May 2000

The focus of this quality management plan is on environmental data, whether they are **collected from measurements, produced from models, or compiled from other sources**. Environmental technologies are covered in the event that the Office of Water is involved.

Types or Sources of Data

Many EPA organizations, including the Office of Water, distinguish between "primary uses" of data and "secondary uses" of data, or "primary data" and "secondary data." For the purposes of this plan, the **primary use** of data is defined to the right. The association of the data with the decision at hand is a critical distinction in the context of this quality management plan. If the data are associated with a decision to be made by the Office of Water, then those data are covered by this plan. Examples of primary data include, but are not limited to:

> ***Primary Use*** of data refers to the use of data that are collected by the Office of Water, or under its direction (e.g., by contractors, grantees, or others) for a specific purpose associated with the decision at hand.

1. Field or laboratory data involving the physical, chemical, or biological characteristics of environmental samples
2. Data on the physical location of such samples, including latitude, longitude, city, county, state
3. Field or laboratory data used to assess the performance of treatment systems or technologies
4. Financial information associated with the development of rules, regulations, or guidance documents
5. Engineering and process data
6. Results produced from models

When primary data are collected by the Office of Water, the collection activities must be planned with quality in mind, and **the quality of the primary data must be assessed** against the needs of the project. Chapter 4 of this quality management plan describes procedures for planning, implementing, evaluating, and improving the quality of any activity, including the collection of primary data.

What often distinguishes primary use of data from secondary use is the control that the Office of Water exerts on the generation of the data. This includes the ability to require that the information needed to assess the quality of the data be generated along with the data and delivered to the Office of Water. The control exerted by the Office of Water is often a function of the fact that the Office of Water is paying for the data to be generated, but may also come into play where the Office of Water has direct

approval authority for the generation of data by an external party.

The Office of Water uses the term **secondary use**, as defined to the right. Other terms such as "acquired data," and "data from other sources" have been used to express the same concept. Secondary data may include:

> ***Secondary use*** of data refers to data that were *not* directly generated by the Office of Water to support the decision at hand.

7. Data collected by someone other than the Office of Water and not under the Office of Water's control
8. Data collected by the Office of Water or others for some other purpose than the current intended use
9. Data compiled from a variety of sources and published in the literature
10. Anecdotal information not collected in any organized fashion

This quality management plan is designed to encourage secondary use of data, where appropriate. The secondary use of existing data can preserve budget resources by avoiding redundant data collection activities within the Office of Water and across EPA programs. However, the challenge in using existing data is that their generation is often outside of the Office of Water's control, and as a result, the Office cannot directly manage or control the quality of the data. If the Office of Water lacks the ability to control the quality of the data, then assessing the quality of the data becomes even more important. When the Office of Water uses data that started as primary data generated by another program within the Office of Water or another part of EPA, the quality of the data may be easy to determine by examining the documentation that was produced with the data. In other cases, the data may have to be examined in terms of who originally produced them and the quality may have to be inferred by less direct means. **Whatever the source, the quality of secondary data must be assessed.** Chapter 4 describes planning procedures for any data collection activity, including planning how secondary data may be assessed and used in a manner that provides an acceptable level of risk in making environmental decisions.

The graded approach applies to the quality of environmental data as well. The decision makers and planners must recognize that the quality of data must be defined by the use, and therefore, the decision to be made. It is critical to plan for the use of data from either primary or secondary sources, but the degree of planning and the quality of the data needed should be based on the **importance of the project, the risk of a decision error, the schedule for completion, and the available resources**. Examples of the subjective nature of quality and the use of a graded approach are provided below.

Americans face risks of illness from swimming and other recreational activities in coastal areas, lakes, and rivers that are contaminated with disease-causing microbes. Many of the beaches and lakes are monitored for conditions that present a threat to human health. When the conditions warrant, an advisory may be issued or a beach may closed to swimming. The decision can be expressed as *"Should an advisory be issued or should a beach be closed?"* The data used to make the decision may come from a routine monitoring program. A series of water samples is collected over a 30-day period and analyzed for enterococci, an indicator of sewage contamination. EPA established numerical guidelines for enterococci in water that address the mean concentration of the organism in the samples as well as a statistical protocol for evaluating the results. If the monitoring results exceed the guidelines, then an advisory may be issued or the beach closed. The data require a high degree of quality management. A monitoring program must be designed and implemented, samplers trained, a microbiological laboratory must be hired, and statistical evaluations of the data must be made. This is a significant commitment of time and effort to protect the potential users of the beach from possible exposure and infection.

In contrast, imagine that heavy rains entering a combined sewer system cause large lumps of untreated sewage to wash onto the beach. The decision remains the same, *"Should an advisory be issued, should a beach be closed?"* However, the untreated sewage is visually apparent and is sufficient evidence of a problem without collecting samples or performing statistical analyses. The threat to human

health is immediate. Therefore, the visual observations of lifeguards with minimal scientific training provide data that are of good quality for *this* decision. More importantly, the decision can be made on the schedule required, i.e., *immediately*. Thus, using the graded approach, the decision to close the beach is made without as much effort to manage the quality of the data.