Data Life Cycle

The data life cycle provides a high level overview of the stages involved in successful management and preservation of data for use and reuse. Multiple versions of a data life cycle exist with differences attributable to variation in practices across domains or communities. The DataONE data life cycle was developed by the DataONE Leadership Team in collaboration with the broader DataONE community, and built upon the life cycle model put forward by the National Science Foundation in the original DataNet solicitation. It serves as an underlying framework for the development of tools, services and education materials by DataONE.
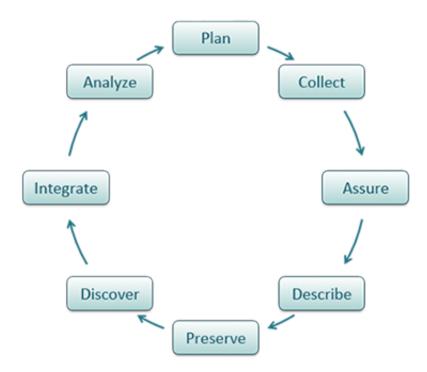

Figure 1: The DataONE data life cycle.

The DataONE data life cycle has eight components:

- Plan: description of the data that will be compiled, and how the data will be managed and made accessible throughout its lifetime
- Collect: observations are made either by hand or with sensors or other instruments and the data are placed a into digital form
- Assure: the quality of the data are assured through checks and inspections
- Describe: data are accurately and thoroughly described using the appropriate metadata standards
- Preserve: data are submitted to an appropriate long-term archive (i.e. data center)
- Discover: potentially useful data are located and obtained, along with the relevant information about the data (metadata)
- Integrate: data from disparate sources are combined to form one homogeneous set of data that

    can be readily analyzed

- Analyze: data are analyzed

Some research activities might use only part of the life cycle; for instance, a project involving meta-analysis might focus on the Discover, Integrate, and Analyze steps, while a project focused on primary data collection and analysis might bypass the Discover and Integrate steps. In addition, other projects might not follow the linear path depicted here, or multiple revolutions of the cycle might be necessary. Further, some scientists or teams (e.g. those engaged in modeling and synthesis) may create new data in the process of discovering, integrating, analyzing, and synthesizing existing data.

Examples of how different users might progress through the data life cycle can be found in our user personas.

The data life cycle serves as a navigation tool for the DataONE Best Practices database, facilitating users in discovering recommendations on how to effectively work with their data across all stages the data life cycle. Software within the Investigator Toolkit has been designed to support researchers at multiple stages of the data life cycle. DataONE has developed data management education modules to aid educators and researchers in their training and self-learning activities.