Some Terminology

Users who are new to the Ecological Metadata Language (EML), or to the concept of metadata generally, may find the following definitions useful.

Data Package – a dataset and the metadata describing it, suitable for upload to a data repository.

Data Table – data in tabular format, usually saved as a CSV (comma-separated value) text file. Such data table CSV files can be read and analyzed by ezEML to assist in generating much of the metadata describing the tables.

Data Table Attributes – the individual variables whose values are stored in a data table. They usually correspond to columns of the table. Since EML uses the term "attributes" rather than columns or variables, ezEML uses that terminology as well.

Dataset – a set of one or more data entities. For ezEML's purposes, these entities will be data tables and "other entities" (see below).

EML (Ecological Metadata Language) – a metadata standard developed by and for the ecology discipline. It standardizes the form and content of metadata documents that describe typical datasets in ecology. The EML standard was designed to handle an enormous variety of data scenarios, so EML is complex and the learning curve to creating it can be steep. Many data scenarios require only a relatively small subset of EML, however, and ezEML is intended to be an easy-to-use, do-it-yourself tool for creating EML documents in such relatively simple scenarios.

Metadata – data that provides information about other data. In the present context, it refers to a document that describes a dataset, giving information about who created it, what it contains, when, where and how it was created, and detailed descriptions of the data tables comprising the dataset. The metadata created by ezEML adheres to the EML standard and makes it possible for you to upload your dataset to a data repository such as the repository managed by EDI, the Environmental Data Initiative.

Other Entity – A dataset described by ezEML may include data tables and "other entities." The latter may be such things as zip files, R or Python scripts, etc., that will be part of the dataset.

Variable Types – ezEML recognizes four variable types: Categorical, DateTime, Numerical, and Text. **Categorical** variables are ones that take on a limited number of possible values; e.g., Low, Medium, High. **DateTime**, **Numerical**, and **Text** variables are just what they sound like.

XML (Extensible Markup Language) – an industry standard language for encoding documents in a format that is both human-readable and machine-readable. XML is used in all sorts of application domains. EML is one such application of XML to the ecological domain.

XML Schema – a description of some particular type of XML document, defining the acceptable form and content of such a document. It gives the "rules" for a valid document of the given type. The EML standard consists of a set of such schemas collectively defining the structure of a valid EML document and the minimum contents such a document needs to contain.