The *Oncor* Geodatabase for the Columbia Estuary Ecosystem Restoration Program: Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates

NK Sather
AB Borde
HL Diefenderfer
JA Serkowski
AM Coleman
GE Johnson

May 2014

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Appendix A

Glossary

Appendix A

Glossary

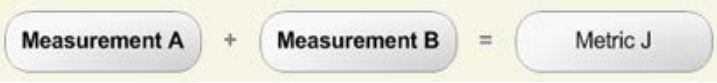
* Definitions obtained from https://www.monitoringresources.org/Resources/Glossary/Index.

** Definitions obtained from http://www.gbif.org/informatics/discoverymetadata/ (the source used by the Pacific Northwest Aquatic Monitoring Partnership)

action	Restoration activity or project, e.g., dike breach.
alias	A user-specific name for a standard value in *Oncor*. For example, more than one project or organization may collect data at the same site. A standard name (value) for that site will exist in a lookup table in *Oncor*. Users (data generators) may also enter an alias, or more typically used name, for that value, for their convenience. This may also apply to instruments, etc.
alias table	A conversion table for translating aliases to standard values.
analysis question	A query of the database addressing a particular research or management need.
attribute	A value in a database that defines a characteristic. For example, the attributes of an instrument might be its serial number, manufacturer, model, etc.
basemap data sets	Mapping data that contains basic reference data, such as roads, cities, prominent landscape features, etc. to orient the user.
category*	A classification rank used for summarizing and reporting that is below subject, above subcategory. For example, Fish or Water Quality. Also "data category."
convention	Adopted standards for measurements, metrics, units, etc. included in the Data Exchange Template.
custom format	An alternative format to the Data Exchange Template developed for large data sets that predate *Oncor* and are deemed important to include within the database.
data category	A set of data collected under a particular field-data collection method. Each data category has a corresponding data reduction workbook. For example, all data associated with water-surface elevation monitoring using data loggers would constitute a single category. This includes data about the instrumentation used and the time-series measurements. Multiple data categories may be combined to answer analysis questions. A data category may

	include multiple metrics and indicators, e.g., mean accretion as well as annual sediment accretion rates.
data custodian	An *Oncor* administrator responsible for enforcing data standards and ensuring that data are loaded into the database correctly. The data custodian interfaces with the data generators to resolve issues regarding data in *Oncor*. Enforcement occurs through manual spot checks and special validation software. Additional responsibilities include maintaining the alias table, and regularly updating lookup tables as needed (i.e., additions and changes).
data dictionary	A worksheet in a Data Exchange Template that defines all fields included in that Data Exchange Template (i.e., field names, data type, description, etc.).
data event	The lowest grouping of data in the data model that includes a unique combination of measurement type, place, and time. A data event is an organizational structure of the data model.
Data Exchange Template (DET)	The Excel file format used to transfer data from a data generator to *Oncor*. The DET is a subset (one or more worksheets) within the DRW.
data generator	The person, or agency/organization, providing data to *Oncor*.
data layer	An individual geographic information system (GIS) data file representing a theme or parts of a theme such as land cover, elevation, or hydrography.
data model	A conceptual database design process structure that includes a life cycle of end-user needs assessment, data type definitions, linkages of data, design review, implementation, and testing of design.
data provider	See data generator.
data reduction	The process of transforming raw data by statistical or mathematical functions into a more usable format.
Data Reduction Procedure (DRP)	A data-category-specific stepwise description of how to reduce data for *Oncor*.
data reduction space	Includes a set of data processing procedures, including quality assurance/quality control and a Data Exchange Template.
Data Reduction Workbook (DRW)	The Excel workbook, corresponding to the DRP, which contains informational, data reduction, and data loading worksheets documenting the data reduction process for a specific data set.
data standard	The set of rules that applies to the contents of fields and records stored in a database.
data steward	An individual user or entity that maintains control over a data set.
data table	Numerical and/or textual information structured into rows and columns and may or may not be linked to spatial features.

data theme	A categorization of GIS data, sometimes synonymous with data layer, into groupings of geographic objects that share a common purpose, function, or type such as vegetation type, soil texture, dike, and tide gate locations.
data type	The attribute of a variable, field, or column in a table that determines the kind of data it can store. Common data types include character, integer, decimal, single, double, and string.
data verification	The process used to determine if data are accurate, complete, traceable, and meet specified performance criteria or control limits.
database	A collection of structured, interrelated information stored as a series of tables in a commonly accessible information system and can include both spatial and non-spatial data.
data set*	A collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the data set in question. It lists values for each of the variables, such as the height and weight of an object. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows. Nontabular data sets can take the form of marked up strings of characters, such as an XML file.
derived data	Using a base set of data, either tabular or spatial, multiple variables and/or mathematical functions are used to convert data to another form, revealing additional metrics or information.
domain table	A table within a database defining unique allowable values for a given column of data to aid in reduction of data errors; e.g., a user can only include one of the following values: 1, 6, 12, 18, 24 for column *X* in Table *Y*.
end user	The individual, organization, or entity that is using a developed product.
estuary-wide scale	Spatial and ecological scale represented by the area of the historical floodplain from Bonneville Lock and Dam to the mouth of the Columbia River.
error file	A file generated by the Preloader that shows why a DET submitted to *Oncor* failed to load.
feature class layer	A collection of geographic features that share a common feature geometry (i.e., point, line, polygon).
feature geometry	Spatial representation of geographic objects within a data theme that are represented by a point, line, polygon, or distributed grid/mesh.

field description	The background information for a column of data in a data table that corresponds to the field name. It may include what the data in the column represent and what the units are for the data.
field name	The given identifier for a column of data in a data table; i.e., the header.
foreign key	Within data tables in a database, an attribute or set of attributes in one table that match the primary key attributes in another table with the intent of joining one or more data tables together. Also see primary key.
indicator*	Value resulting from the data reduction of metrics across sites and/or temporal periods based on applying the procedures in the inference design. A reported value used to indicate the status, condition, or trend of a resource or ecological process; intended to answer questions posed by the objectives of the protocol. Contrast with metric. 
key field	A field that is required to uniquely identify the record, such as sampling date and location.
keyword	A single word or short phrase that describes the context and content of a given data set.
landscape scale	A spatial and ecological scale that makes use of site scale and regionally available information within a larger system to consider process and function in a more system-based and holistic manner.
legacy data	Historical data and/or data collected and structured using an older protocol.
loader	A program that loads data in *Oncor* format into the *Oncor* database.
managed data	Spatial and tabular data held in the local *Oncor* database located on a server managed by the *Oncor* development team. These data may come from multiple entities who have agreed to store data locally within the *Oncor* database. Also see unmanaged data.
measurement*	A value resulting from a data collection event at a specific site and temporal unit. Measurements can be used to produce metrics using a response design. 
metadata**	Metadata are literally "data about data." They provide information about the "who, what, where, and when" of data and can be considered from the perspective of both the data producer and the data consumer. For the producer, metadata are used to document

	data in order to inform prospective users of their characteristics, while for the consumer, metadata are used to both discover data and assess their appropriateness for particular needs. The *Oncor* database has been structured such that metadata for each data category are identified in the DRW.
metric*	A value resulting from the reduction or processing of measurements taken at a site and temporal unit at one or more times during the study period based on the procedures defined by the response design. Metrics can be used to estimate an indicator using an inference design. Note that a variety of metrics can be derived from original measurements.
monitored indicators	Values resulting from data reduction of metrics sourced from a time-series of field-collected data around specific themes of 1) water-surface elevation, 2) water temperature, 3) channel cross-section surveys, 4) sediment accretion, 5) vegetation, and 6) fish. Also see indicators.
non-spatial data	Information structured without reference to a geographic object, these types of information would typically be stored in data tables.
normalization	As used in *Oncor*, the process of eliminating duplication of data in a database structure.
Oncor	A geodatabase for storage and retrieval of data generated by the Columbia Estuary Ecosystem Restoration Program.
Oncor format	An intermediate file format generated from a DET by the Preloader for use by the Loader.
Oncor data standard	See data standard.
pedigree	The recorded source and history of a given data set for the purpose of understanding the integrity of the data and appropriate use and application of the data.
original data	Measurements made by scientists or technicians in the field or laboratory. Not quality control checked, reduced, or mathematically transformed. Also raw data.
preload file	A file generated by the *Oncor* data loader that shows how a successfully submitted DET will appear when uploaded to *Oncor*.
primary key	Within data tables in a database, an attribute or set of attributes in a database that uniquely identifies each record with the intent of joining one or more data tables together. A primary key allows no duplicate values and cannot be null.
published services	A means of making available, publicly or privately, data that can be accessed and transferred over the Internet using web-enabled applications in a seamless, behind-the-scenes manner using one of a

	number of established protocols such as REST, WSDL, or JSON; also commonly known as "web services." See web services.
raster data	A type of GIS file format that represents a data theme for a geographic area in a continuous manner using an equal size, cell-based, row-column structure (i.e., a matrix). Examples of data in this format include imagery and digital elevation models. These types of data can be layered into "bands" that represent different phenomena, for example, different ranges in the electromagnetic spectrum, as are found in multi-spectral satellite imagery.
raw data	Measurements made by scientists or technicians in the field or laboratory. Not quality control checked, reduced, or mathematically transformed.
reach	A common hydrogeomorphic area typically using a number of criteria including floodplain boundary, landforms and geology, presence and location of tributaries, gradient, and in the case of estuary systems, salinity and tidal influence.
regional data	Spatial data consistently representing an area with similar physical characteristics or a system or component of a system. In general, regional data are often represented at a coarser spatial scale, but cover a broader geographic area.
relationship	In the context of databases, data from two or more data tables are joined through a common data field, referred to as a primary key or foreign key. The linkages to other tables can be set as one-to-one or one-to-many.
sampling location	The spatial area where one or more measurements are taken. Usually smaller than a site.
SDE	A software technology from Environmental Systems Research Institute (ESRI) for managing spatial data in a Relational Database Management System (RDMS) allowing for enterprise use (large multi-user environment) of geographic data. The technology makes accessing spatial data from the RDMS seamless to the end-user. Also referred to as ArcSDE and Spatial Database Engine.
shapefile	The shapefile format is a commonly used GIS file format that consists of several individual files and, while native to ESRI products, can be generated using a variety of GIS software. The shapefile can include points, polylines, or polygon features.
site*	The spatial area that encompasses one or more sampling locations. Sampling designs in the Columbia Estuary Ecosystem Restoration Program include restoration, reference, control and other sampling sites. Examples include natural features such as Karlson Island, and ownership boundaries such as Julia Butler Hansen National Wildlife Refuge.

source data	The origin of a particular set of information, whether it is tabular or spatial.
spatial data	Representation of information in a geographic context stored using either one feature geometry type or in simple X/Y or longitude/latitude in a data table, thus data may or may not be in a standard GIS file format.
standard value	A frequently referenced person, place, or thing that has been assigned a single term in the database. This standardized set of values is managed by the data custodian. Examples include people who collected the data, sampling locations, and instruments. For example, Jane Doe may be the standard value and the initials JAD may be an alias used by a data generator.
standard CEERP metrics	a set of metrics and indicators that have been established for each data category and are included in Data Reduction Procedures and Data Exchange Templates.
study area	A conglomeration of sites. Also see site.
subcategory*	A classification rank used for summarizing and reporting that is below category. For example, Fish Abundance or Turbidity.
survey data set	A collection of information sourced from a survey instrument such as a Total Station, theodolite, or global positioning system. For research, monitoring, and evaluation work, these types of data are usually collected for cross sections, transects, surface or feature elevation points, instrument calibration, or boundary definitions.
tab	A worksheet in an Excel workbook.
tag	See keyword.
temporal data	Any spatial or tabular data consistently and repeatedly collected over a regular or irregular time interval. This form of data will have date/time stamps associated with the observation value.
temporal unit*	The interval during which measurements are made at the site, and subsequently the interval for which metric values could be determined.
unmanaged data	Data owned and maintained by others through a special means of live data access over the Internet referred to as “web services.” These data are formally referred to as “unmanaged” because the data are not stored within the *Oncor* database nor does the *Oncor* team have control over the data.
use case	Container for analysis questions. Typical application of the database, e.g., Expert Regional Technical Group project template where individuals would have a specific use for the database. A means to organize the analysis questions.

user interface	The aspects of a computer system or program with which a software user can interact, and the commands and mechanisms used to control its operation and input data. In the case of *Oncor*, the user interface is a web-based interface.
vector data	Spatial data taking the form of points, lines, or polygons and stored as a single coordinate pair (in the case of a point) or an ordered list of coordinate pairs representing the vertices of a geographic feature (in the case line or polygon). Compare to raster data.
web services	A means of communicating and transferring data over the Internet using web-enabled applications in a more seamless, behind-the-scenes manner using one of a number of established protocols such as REST, WSDL, or JSON.
widget	An interactive graphic component of a user interface (such as a button, scroll bar, or menu bar), its controlling program, or the combination of both the component and program. Also see user interface.
worksheet	The same as a single tab in an Excel workbook.