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

Contents

1.0 Introduction

In 2012, the U.S. Army Corps of Engineers (USACE) Portland District initiated development of a web-accessible geoscientific database (called *Oncor*) for analysis and synthesis of action effectiveness and related data from monitoring and research efforts for the Columbia Estuary Ecosystem Restoration Program (CEERP). Pacific Northwest National Laboratory (PNNL) researchers are developing this estuary-wide data management and information discovery/retrieval system to provide an intuitive user environment and the necessary resources and tools to standardize and upload/download data (see the 2012 annual report by Coleman et al. 2013). The intent is for *Oncor* to enable synthesis and evaluation of data generated by multiple entities, the results of which can then be applied in subsequent CEERP adaptive management and decision-making processes. The database is called *Oncor* after the genus *Oncorhynchus*, which includes Pacific salmon and steelhead, the focus of CEERP estuarine and tidal freshwater habitat restoration efforts in the lower Columbia River and estuary (LCRE).

Action-effectiveness monitoring and research (AEMR[1]) and other relevant data are being collected at CEERP restoration and reference marshes, shrub-dominated wetlands, forested wetlands, and other habitats and AEMR study sites (Figure 1.1). Where applicable, data are collected using protocols developed by Roegner et al. (2009), called the *Data Collection Protocols*. Many regional entities are involved in this data collection, including the following organizations: Columbia Land Trust, Columbia River Estuary Study Taskforce, Cowlitz Tribe, Lower Columbia Estuary Partnership, National Oceanic and Atmospheric Administration, Oregon Department of Fish and Wildlife, PNNL, USACE, and the Washington Department of Fish and Wildlife. The CEERP prioritizes AEMR measurements and metrics pertaining to juvenile salmon and their habitats, because the focus of CEERP is on ecosystem improvements to support juvenile salmon emigrating from the Columbia River basin (BPA/USACE 2013; Thom et al. 2013). In addition, CEERP prioritizes select habitat data important for the assessment and adaptive management of ecosystem restoration, such as plant communities, channel cross sections, photo points, sediment accretion/erosion rate, water-surface elevation, water temperature, and inundation. But, while there are standard protocols for data collection, none exist for the next phase of the scientific process for reducing and uploading these data to a regional database in support of CEERP.

Figure 1.1. From left: tidal freshwater marsh, beaver dam in shrub-dominated wetland, riparian forest, and Sitka spruce swamp.

[1] Action-effectiveness *monitoring* involves spatially extensive sampling of basic restoration indicators, whereas action-effectiveness *research* involves locally intensive sampling at restoration and reference sites to characterize ecosystem structures, processes, and functions.

1.1 Purpose of This Handbook

The purpose of this handbook is to provide the regional partners that are collecting AEMR and other data supporting CEERP with the procedures and tools to reduce and upload their field-collected data into *Oncor* (Figure 1.2; Table 1.1). The data categories included in this handbook include physical and biological characteristics. Not all data categories covered by the *Data Collection Protocols* are covered herein and, likewise, not all data categories covered herein were included in the *Data Collection Protocols.*

Figure 1.2. Data flow for AEMR and other data intended to support CEERP. The red boxes signify material that is covered in this handbook.

Table 1.1. Data categories covered in the *Data Collection Protocols* and the *Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates.* Green shading signifies the data category is covered to some degree.

Data Category	Collection Protocols	Reduction Handbook
Water-surface elevation	covered; referred to as Hydrology	covered
Water quality	water temperature and salinity	water temperature; included within the data reduction procedure for water surface elevation
Elevation	elevation survey methods	covered as related to sediment accretion
Sediment accretion	covered	covered as a distinct data category
Channel cross section	covered	covered as a distinct data category (under construction)
Landscape features	aerial photography and photo points	photo points
Plant communities	herbaceous vegetation	covered for herbaceous wetland vegetation cover (plant community composition and percent cover)
	forested wetlands	covered for tree plots and site summaries (tree densities)
	shrub/scrub vegetation communities	shrub/scrub vegetation densities
Vegetation plantings	pertained to planting success	not covered
Fish catch	variety of gear types covered to sample species composition, size distribution, catch per unit effort (CPUE)	covered for beach seine methods only
Fish density	expressed as CPUE	covered as a distinct data category pertaining to beach seine collections
Fish size	included under fish community	ibid
Chinook salmon genetics	not covered	ibid
Fish diet	not covered	ibid
Fish prey	not covered	covered as a distinct data category

The handbook presents a set of data reduction[2] procedures (DRPs) that data generators can follow to process raw data collected in the field, for the purposes of analysis and reporting and standardization for uploading to *Oncor*. In association with the DRP documentation in this handbook, we created Microsoft Excel workbooks for actual data reduction for specific data categories (Table 1.1). These workbooks are termed data reduction workbooks (DRWs), and the most current versions are available for download on the *Oncor* website: http://oncor.pnnl.gov/ (link for external, non-PNNL users). A subset of the sheets in each workbook constitutes a data exchange template (DET), which provides a standard mechanism for importing data into *Oncor*. The selected subset is customized for each data category, e.g., fish density. Guidance to support data quality and management efforts is also incorporated in the DRPs. To summarize, the DRPs, DRWs, and DETs provided in this handbook and accompanying Excel files help ensure *Oncor* data integrity while maximizing ease of data reduction, quality control, and uploading. The ultimate intent is to have a web-accessible, comprehensive geodatabase of AEMR and other data to facilitate synthesis and evaluation of the collective effectiveness CEERP restoration actions.

1.2 Organization of This Handbook

The next section of this handbook contains basic material that apply to all data categories covered in the handbook. After the basic material, detailed DRPs for select data categories are presented (Table 1.1). Each DRP has an associated Excel workbook, the DRW. Because there may be multiple definitions of the same term and that the terminology involved in the entire data flow process can be confusing, there is a glossary of the terms in Appendix A. Other appendices include technical information about *Oncor* such as data workflow (Appendix B) and data standardization (Appendix C). Appendix D demonstrates outcomes from data analysis to answer analysis questions with specific temporal or spatial parameters (e.g., seasonal temperatures and fish densities), and how monitored indicators from different data categories can be combined to answer analysis questions (e.g., topography and plant species presence).

1.3 How to Use This Handbook

For a given data category, complete the following steps:

- Visit the *Oncor* website and download the latest version of the *Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates*. Open the handbook PDF.
- Download the latest version of the DRW for the data category(s) of interest, save it to a local computer, and open it. Having a split screen of the handbook PDF and the DRW Excel workbook will facilitate the work.
- Study the basic material on general procedures and requirements.
- Go to the appropriate data category in the handbook and read the section completely before working with data.
- Familiarize oneself with the content and structure of the particular DRW.
- Create a trial data set, enter the data in the DRW, perform the data reduction procedure, and inspect the results for the desired outcome.
- Use the Help button on the *Oncor* website to request assistance.

[2] Data reduction is simply the process of transforming raw data by statistical or mathematical functions into a structured format.

2.0 Basic Material on DRPs, DRWs, and DETs Relevant to All Data Categories

Data undergo a series of steps from collection to uploading into *Oncor* (Figure 2.1). The collective term for data reduction and uploading to *Oncor* is "data processing." In research and monitoring, the measurements made by scientists in the field or laboratory are colloquially referred to as "raw data." This section describes the effort by the *Oncor* development team to provide detailed and efficient methods for those collecting raw data for the CEERP ("data generators") to reduce data as needed to meet typical reporting requirements and upload data into *Oncor* (Figure 2.1). We understand and recognize that there is not a single best way to accomplish this goal. However, it is important to standardize the process to enable data to be integrated for estuary-wide analysis. The procedures apply to existing data (both *Oncor*-formatted and custom-formatted) and new data. Once data are in the *Oncor* database, they will be available for *Oncor* users to analyze across data collection sites, times, and monitoring programs in the LCRE and answer specific CEERP analysis questions.

This section includes the relationship between this handbook and the data collection protocols, key terminology, quality assurance and quality control procedures, and detailed information about data reduction workbooks. The section culminates with the DET uploading procedure.

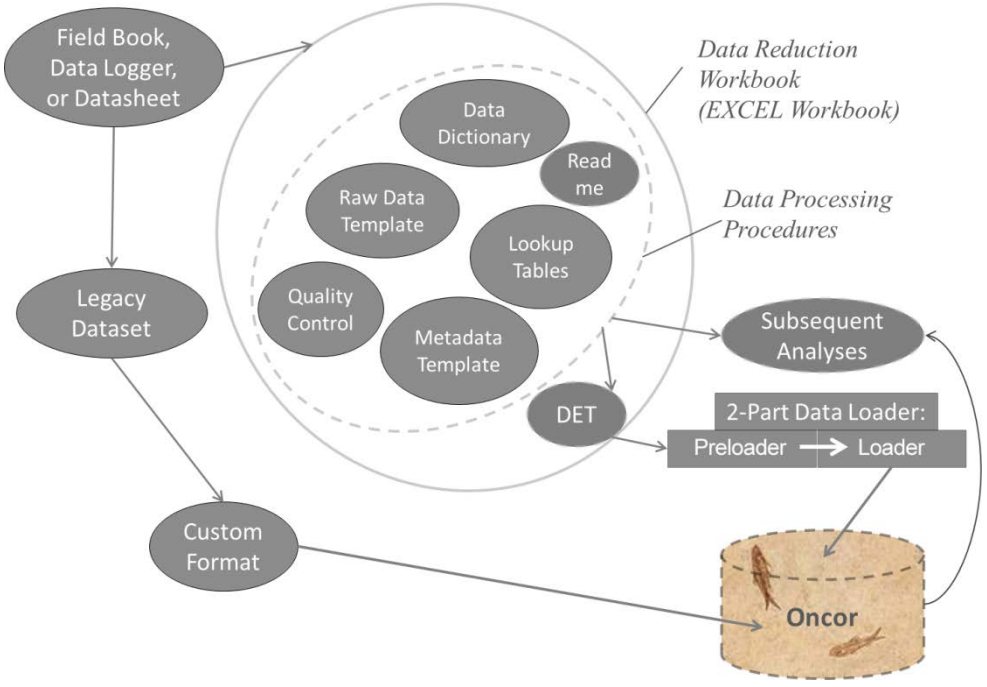


Figure 2.1. Schematic of data flow from field data to *Oncor*. Data flow includes data reduction and entry into a Data Exchange Template (DET) by the data generator, prior to uploading to the *Oncor* database. The data reduction workbook also includes spatial data.

2.1 Relationship to Standard Data Collection Protocols

In the LCRE, many people who monitor restoration and reference sites use the *Data Collection Protocols* described by Roegner et al. (2009), a document developed by the USACE's Cumulative Effects study (study code EST-P-04-04). The *Data Collection Protocols* were developed to support the CEERP and are available at www.nwfsc.noaa.gov/publications/. The individual methods presented by Roegner et al. (2009)—e.g., hydrology, vegetation—are also available for selection at www.monitoringmethods.org, where according to the Pacific Northwest Aquatic Monitoring Partnership (PNAMP) terminology, they are termed "methods" not "protocols."

In the 2013 geographic review of the Bonneville Power Administration (BPA)/Northwest Power and Conservation Council's Fish and Wildlife Program by the Independent Scientific Review Panel, proposals for five "umbrella projects" implementing CEERP restoration actions—Columbia Land Trust, Cowlitz Indian Tribe, Columbia River Estuary Study Taskforce, Lower Columbia Estuary Partnership, and Washington Estuary Habitat Memorandum of Agreement (Washington-Action Agencies 2009)—state that the *Data Collection Protocols* by Roegner et al. (2009) will be the basis of monitoring in 2014.

The *Data Collection Protocols* focused on field data collection with few references to the procedures required for data reduction, analysis, and reporting. Not surprisingly, questions have arisen about how to ensure data quality and standardization to provide comparable results across the CEERP (see for example Borde et al. 2012). In addition, several metrics that are frequently calculated for salmon today, including density, genetics stock identification, and diet, were not detailed in the protocols (Table 1.1). To address these needs, the *Oncor* development team created this *Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates* and associated Excel workbooks providing detailed instructions and demonstration examples to help users efficiently transform raw data collected with these methods into measurements, metrics, and indicators in formats for uploading to *Oncor*.

2.2 Terminology

This section contains frequently used terms that users of the handbook must understand. These and additional terms are defined in the glossary (Appendix A).

As a matter of principle, *Oncor* adopts whenever possible the relevant PNAMP terminology, which is available from www.monitoringmethods.org/Glossary/Definition/, including the following three terms:

- *Measurement*: A value resulting from a data collection event at a specific site and temporal unit.
- *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.
- *Indicator*: A 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. According to the inference design, metrics are combined or reduced to produce indicators.

Other terms that are essential to understand warrant definitions here because they are used throughout this handbook.

- *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, organizations, etc.
- *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 ensuring data standards are met and that data are uploaded into the database correctly.
- *Data generator*: The person, or agency/organization, providing data to *Oncor*.
- *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 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 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.
- *Original data*: Measurements made by scientists or technicians in the field or laboratory. Data that are not quality control checked, reduced, or mathematically transformed. Also called “raw data.”
- *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.
- *Worksheet*: The same as a single worksheet/tab in an Excel workbook. Also “datasheet” or “spreadsheet.”

2.3 Quality Assurance and Quality Control

The steps in the data reduction process include electronic data entry, subsequent calculations, and quality control to reduce the data into a usable format. (Additional details are provided in Appendix C, Data Quality.) Data reduction can be as simple as changing the unit—a common example being changing survey feet to meters. Typically, a series of reduction steps needs to be performed on AEMR and other data, with quality control (QC) checks at each step to ensure that the original values are correctly reduced.

Final data used for reporting or uploading into *Oncor* should be traceable from raw data through all processing steps that were performed. Ultimately, data reduction should produce documentation sufficient to permit an independent data auditor to determine whether data are accurate, complete, traceable, and meet specifications. These types of procedures also help to ensure that data are not lost when staff members change at a data-generating organization.

For the purposes of *Oncor*, the focus of QC is on *data verification*. In general, data verification is the process used to determine if data are accurate, complete, traceable, and meet specified performance criteria or control limits. In the metadata for a given data category DRW, the user is requested to indicate whether QC was done to the level described in this section.

A review of logbooks and other data collected and recorded in the field should occur as soon as practically possible after the data are collected. In fact, completing this process in the field helps to ensure that any missing, unclear, or incomplete data can be corrected without making additional field trips. The design and use of datasheets and logbooks should ensure maximum legibility, accuracy, traceability, validity, and clarity of meaning. While the use of pencil in the field is standard, scanned or printed backup copies should be made in the office and archived separately from the raw datasheets as soon as possible upon return from the field. Datasheets and logbooks require the date of collection, time of collection if appropriate to the indicator, the initials of the recorder(s), and in most cases, the place of collection (e.g., the site, plot, global positioning system [GPS] point, etc.). These data are typically included in *Oncor*. Because *Oncor* is a geospatial database and all data are organized based on place of collection, the most accurate measurement of location that can be made in the field should be made. The indicator-specific procedures specify whether points, lines, or polygons are the standard.

For *Oncor* data, we recommend that independent verification of the data in the reduction process comprise the following, at a minimum:

- 100% check of data transcription (data entry)
- 10% check of calculations, with a 100% check required when errors are found to determine the extent of the problem and correct it.

Data verification must be independent. That is, it can be completed by a peer reviewer or quality assurance staff member, if available, but it should not be completed by the scientist or technician who conducted the data entry or the calculations that are being verified. To make the QC process traceable, all activities must be recorded, and therefore it is customary to conduct checks on paper copies in ink. For example, one standard method for documenting the process involves 1) denoting all transcribed data that have been checked with a mark made in ink; 2) crossing out incorrect values and writing the correct ones instead; and 3) on each page, noting the level of data review performed (e.g., 100%, 10%) and the initials and date of the checker. Once this level of check has been performed, the data in the electronic file must be corrected; when correction is completed, the initials and date of the person who has done so are also noted in ink on the QC record sheet. Ideally, a final copy is printed and verified for 100% correctness. While the QC process is typically conducted using paper copies of datasheets, an alternative approach is to perform data checks and documentation of errors using electronic files. Data generators should ultimately select an approach that is most suitable for project needs and programmatic requirements.

2.4 Data Access Control: Permissions and Security in *Oncor*

For a variety of reasons, owners of data in the *Oncor* database may not wish to share parts of their data sets with other users or the public. Data could be provisional and not ready to be publicly disseminated or results could be of a type that might be easily misinterpreted. To address this anticipated need, *Oncor* includes a mechanism that allows data generators to control access to data they have uploaded.

Data access control is implemented using a permission-based system. Every measurement record in the database is associated with a data owner and an access group. The DB_Access column in each DET is where the user can define the level of access for the individual measurements, metrics, or indicators in the data set. Every *Oncor* user will have a list of access groups to which they have owner-permitted access. To access a record, the user must have permission from the owner of the data for the access group associated with the record.

When data are uploaded via a DET, *Oncor* stores the metadata supplied for that DET with the data set. Two metadata fields will be used for establishing the accessibility of the data: DET_Owner and Access_Group. The DET_Owner field represents the person who owns the data in the DET. This person is referred to here as the data owner. The Access_Group field is a column in each of the DET worksheets of the DRW workbook that allows the data owner to assign each table row to a group for the purpose of restricting access. An access group is a collection of owned records that have similar access properties. Access groups are specified with a positive integer, with the group 0 representing the default publicly accessible group. A blank value in the Access_Group field will assign the record to the 0 access group.

To view data beyond the standard *Oncor* base data, users must have an account on the system. Part of the information associated with *Oncor* user accounts is a table of group permissions. The table lists owner and access-group pairs to which the user has access. The absence of a permission record for a data owner is equivalent to having access only to his/her public (access group 0) records. Data owners will have access to all their data without an explicit entry in the permission table.

2.5 Standard Values

To ensure proper integration of all data in *Oncor*, the maintenance of a single set of data standards is critical. Prior to entering data into DRWs, users must first provide standard values via the StandardValues_v0.4.xlsm file which is structured according to the following topic areas:

- **Agency.** Organizations associated with lower Columbia and estuary (LCRE) data (e.g., Lower Columbia River and Estuary Partnership, Columbia Land Trust, etc.).
- **Person.** Full name and affiliation of people associated with LCRE projects.
- **Program.** Highest organizational level guiding a data collection effort; missions, research questions, and protocols are defined at this level (e.g., Columbia Estuary Ecosystem Restoration Program; CEERP)
- **Project.** Specific research activity that supports one or more of the research questions of the Program (e.g., Reference Site Study).

- **Document.** Reference document citation provided as supportive material for methods that may be referenced in various DRWs.
- **Instrument.** Specifications for equipment used to collect data (e.g., data logger, fish net, etc.).
- **Method.** Data collection method.
- **Species.** Specific information for a plant or animal (e.g., scientific name, species code, common name, etc.).
- **Unit.** Measurement units reported.

Appendix C provides additional background information as well as an expansion on the purpose of standard values. During the beta testing phase, users should populate each of worksheets within the StandardValues_v0.4.xlsm file, to the extent possible. Ultimately, the standard value file will be linked to the metadata and other relevant locations within each of the individual DRWs. In the future, coupling the standard values file with the DRW will facilitate a more automated process for data entry while assuring proper integration of data into *Oncor*.

2.6 Data Reduction Workbooks and Data Exchange Templates

This section describes the DRWs, data dictionary, metadata, entering data into DRWs, validation mechanisms, and key fields.

2.6.1 Description of DRW

We are calling the activities that occur after data collection and before *Oncor* data loading "data reduction" (Figure 2.2), which includes implementing DRPs and preparing DRWs and DETs (Figure 2.1). As noted above, we provide as a companion to this handbook example data reduction workbooks in Excel corresponding to key data categories: water-surface elevation and temperature, sediment accretion rate, photo points, channel cross sections,[3] herbaceous wetland vegetation cover, tree plots and site summaries, fish catch and density, fish size, fish diet, fish prey, and Chinook salmon genetic stock identification.

The DRWs (Figure 2.2) are available on the *Oncor* website to data generators and may be adopted at will, but are not required because many data generators may have preferred data reduction workbooks already in use. As depicted in Figure 2.1, *Oncor* can also accept "custom-format" data given appropriate coordination between the data generator and *Oncor* data custodian. Data generators should follow the stepwise process outlined in the DRPs to ensure data are formatted according to the DET, which is a subset of worksheets in the DRW workbook containing final data for upload into *Oncor* (see Figure 2.2, red boxes). The DETs will be required for all data uploaded to *Oncor* in the future. The DETs contain fields that represent data collection in the LCRE that have been identified as a priority by the CEERP managers. DETs include fields for each measurement, metric, and indicator that will be entered into the *Oncor* database.

[3] Under construction.

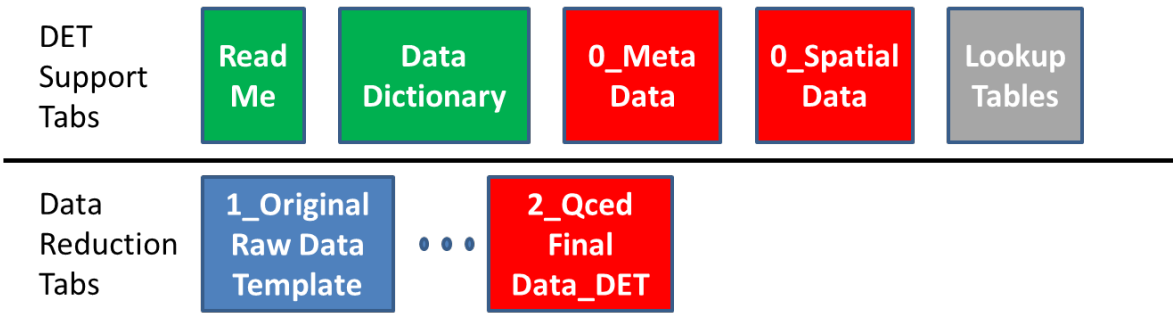


Figure 2.2. Generic Data Reduction Workbook structure including the Data Exchange Template (in red). The three-part general processing sequence is represented by 0, 1, and 2. Each box represents a worksheet tab in the workbook. Green tabs are informational. Blue tabs consist of a template with column headers for the entry of raw data. The ellipsis represents the data reduction process, e.g., QC checks. Grey tabs provide standardized naming conventions. Red tabs are quality-control and data-reduction end points and constitute the Data Exchange Template. The red tabs are required for uploading information into *Oncor*. Only quality-control-checked data may be uploaded into *Oncor*.

When ready to reduce the field data, data generators will download blank DRWs from the *Oncor* web site. The DRWs will not only be specific to the data category requested, they will also contain the latest standard-value lookup lists with alias values associated with the requesting user. To access *Oncor* for upload, data generators will need to identify themselves so that pre-loaded information specific to them can be incorporated with the upload, e.g., name and agency. Data generators will be presented with a choice of DRW data category(s) to download. By accessing DRW/DETs through the web interface, format updates will be easily disseminated. After a DRW/DET modification, the loader would remain backward compatible for some set period of time to allow data generators to complete old DETs that they may have already started.

Data reduction space (Figure 2.1) includes data processing procedures (e.g., mathematical transformations and QC checks), as well as several key elements that can be stored in Excel data reduction workbooks (Figure 2.2). The DRW contains a set of worksheet tabs, which are described as follows:

- **Read Me.** Contains general information about the DRW, including a list of the names of each worksheet tab that describes its purpose, contents, and use.
- **Data Dictionary.** Defines each field (i.e., each column header) found in subsequent sheets; information includes a description of what goes in the field, the data type, whether the field is required, what automatic validation occurs, any applicable standards, and whether it is a calculated field.
- **0_Metadata.** Lists associated metadata (e.g., data owner, contact, instrumentation, etc.); contains the same columns for all data categories, which must be filled in by the data generator.
- **0_SpatialData.** Two tabs for entering spatial data. One tab can be used to enter coordinates for points the other can be used for to enter information for an uploaded shapefile with points, polylines, or polygons.
- **Raw Data Template.** One or more tabs containing an empty table into which the data generator enters or copies data.

- **DET Tabs.** One or more tabs into which the data generator copies quality-control−checked and reduced data; for some data categories, the DET tabs may be automatically populated using formulas and/or macros.
- **Lookup Tables.** Tabs that, through the use of lookup lists, are used to help populate standard-value fields (e.g., species, instruments, organizations, sites, etc.) to ensure compliance with *Oncor* data standards; lookup tables may be viewed and edited by pressing the Standard Values button available on most tabs.

Each DRW includes as many datasheets as necessary to accommodate associated measurements, metrics, and indicators, and data generators can add sheets for their own purposes as needed, which will not be delivered to *Oncor*. The data generator may choose to perform QC checks within sheets in the DRW, on a printed copy, or in another electronic file as preferred. Raw data are not delivered to *Oncor*. Only the DET, a subset of the DRW, is delivered to *Oncor*.

2.6.2 Data Dictionary

The data dictionary in each DRW contains the following information about each worksheet in the DRW:

- **Sheet**. Name of worksheet where the field appears.
- **Col_Num.** (hidden column) Column to help sort the fields.
- **Column**. Letter identifying a column in the worksheet.
- **Field Name**. Name in the column header cell.
- **Description**. Description of what the field is used for.
- **Data Type**. Storage category of the field in the database (text, integer, Boolean, etc.).
- **Key Field**. Yes, if the field is required to uniquely identify the record.
- **Required**. Yes, if the field must be nonblank.
- **Std Value**. If nonblank, the standard value must be used to populate the field.
- **Validation**. Description of the automatic validation performed by the workbook.
- **Conventions**. Data standards applicable to the field (e.g., units),
- ***Oncor* Attribute Name**. Name of *Oncor* attribute to which field will be mapped.
- **Formula**. Excel formula for calculation of a value in a worksheet cell.

2.6.3 Validation Mechanisms

The data-entry worksheets in the DRW incorporate automated validation mechanisms to help assure data integrity. The types of validation performed include assuring that standard-value fields only contain values in appropriate lookup lists, checking that data in certain numeric fields fall within an appropriate value range, and verifying compliance with other rules. See Appendix C - - Data Quality and Format: Standards and Enforcement for more details about data validation and standards. Specific validation rules for each field are given in the Data Dictionary.

Data-entry validation is initiated by clicking on a custom button titled “Validate” that is located on the last worksheet where data are manually loaded. This button is powered by underlying Visual Basic for Applications (VBA) code. Because the DRW must include VBA code, the filename extension of the DRW file is .xlsm, to indicate the presence of macros. In addition to data validation, the integrity of the DRW file is protected by selective locking of cells to prevent inadvertent changes to the sheets.

The spatial data in a DRW is validated by checking all fields that require the entry of a location against a set of allowable values. The validation process searches three places in the DRW to determine whether a specified location is valid. First, it looks at a list in a hidden tab called LU-Location, which contains all currently existing locations in *Oncor*. Next, it looks at the Alias_Name field of any records in the tab 0_SpatialData_Coordinates. Finally, it looks at the FileRelate_Field attribute of any shapefiles specified in the tab 0_SpatialData_Shapefile.

2.6.4 Key Fields

Most data-entry worksheets contain key fields, which are indicated in the data dictionary. The presence of key fields ensures that individual records can be distinguished from each other. Every combination of key-field values must be unique in the data-entry table. For example, if a data set such as water-surface elevation consists of a location, measurement date/time, and measurement value, the key fields include both location and measurement date/time. This ensures that records can be distinguished from each other in the database. Compliance with the key-field rule is checked when the Validation button is pressed.

2.6.5 Entering Data into DRWs

The data-entry sheets of the DRWs are designed to make entry of data as easy as possible, while ensuring data integrity and completeness. Data are first entered into the worksheet called “0_Metadata” (Figure 2.2). (See the following subsection for more details on metadata.)

Next, raw data are entered into the tab or tabs that have the suffix “_Template.” If field data are collected electronically, most data entry should be able to be accomplished with a small number of copy-and-paste operations. For manually recorded data, the Excel data-table structure should mirror the form of the field notes. The number, format, and content of DRW data-entry sheets will differ for each data category.

The raw data in the template worksheets are then processed according to the specific DRP, each step documented under the appropriately numbered tab, until, at the end, one or more red-tabbed DET worksheets is populated. The red-tabbed DET worksheets initially contain gray-shaded example records to assist with understanding the expected contents of the fields. These example records must be deleted before submitting the DET for uploading to *Oncor*.

The data dictionary will provide guidance on whether a particular field in a DRW must be populated. Examples of required fields include site name and sample date. However, it is plausible that not all fields will always contain data. For example, while water temperature is a routine measurement that occurs in conjunction with fish sampling, data may not be available if there the meter malfunctions. In lieu of blank cells, missing data should be indicated using one of two approaches. The code ‘ND’, meaning ‘no data’

can be used when the data type associated with a given field is TEXT. Missing data in fields whose data type is defined as INTEGER or DOUBLE should be populated using '-9999'.

2.6.6 Metadata

Every DRW includes a data-entry sheet with the name "0_Metadata" for the purpose of entering metadata about the DET. Currently, required information includes the following:

- **DET_Type**. Standard value for the data category (read only).
- **DET_Version**. Version number of the DET electronic file (read only).
- **DET_Creation_Date**. The date the data generator created the DRW/DET (read only).
- **DET_Generator.** Name of person generating (creating) the DET.
- **DET_POC**. Standard value for the name of the contact person for the data.
- **DET_File_Name**. Name of the DET file.
- **DB_Alias_Owner**. Standard value for the name of the person associated with Alias_Groups included in this DRW (read only).
- **Sponsor_ProjectID.** Project identifier used by project sponsor.
- **Generator_ProjectID.** Project identifier used by data generator.
- **Project_Name**. Standard value for the name of the project that generated the data.
- **Method.** Method used to collect the data.
- **QC_Procedure.** Note percentage of data transcription and percentage of data calculations quality-control checked (e.g., 100% transcription, 10% calculations).
- **Instrumentation**. If applicable to a given data category, the metadata sheet in the DRW will also identify instrumentation.
- **Vert_Datum.** If any elevation data is included apart from that in the Spatial Data tabs then the vertical datum should be specified here (e.g., North American Vertical Datum of 1988, NAVD88)
- **Units.** If applicable, units for the data reported in the DET should be specified here (e.g., water surface elevation data is in meters or feet).

2.6.7 Spatial Data

Spatial data, the information that describes the geographic location of a feature, is an essential component of the *Oncor* database. Because *Oncor* is a geodatabase, all measurements, metrics, and indicators in the DRWs must be associated with a location on the ground. *Oncor* recognizes three types of geographic features: points, lines, and polygons. A point is defined by a single pair of horizontal coordinates, such as latitude and longitude or easting and northing. A line is defined by a set of two, or more, points that form an open shape, and a polygon is defined by a set of three, or more, points that form a closed shape. To define a location, one must specify the feature type, the point coordinates, and the geographic coordinate system used.

Oncor manages locations as standard values. This means that locations referenced in new data are validated against a list of standard, locations maintained in the database to ensure consistency. The procedure for data entry differs from that for other standard values, because users can specify new location data in two ways, by providing either explicit coordinates (point locations only) or a reference to external GIS shapefile.

Users may define new point locations by explicitly providing coordinates into the table located on a tab called 0_SpatialData_Coordinates, which is available in every DRW. New locations defined this way must include information in the following fields:

- **PointID.** The name of the point. The point name needs to be the same as the names in the DET that identify a spatially distinct data record (e.g., a point ID field or a Site ID field).
- **SiteID.** If the location of a specific point is not available then the site location can be used instead. The name of the site is then identified in the SiteID field.
- **Alias_Group.** Grouping category for Standard-Value Name alias (see Appendix C - Data Standardization for more on Alias Names).
- **Easting and Northing.** The fields associated with the x (e.g., easting or longitude) and the y (e.g., northing or latitude) coordinates.
- **Horiz_Coord_System.** Horizontal coordinate system associated with Easting and Northing values.
-

These fields are further described in the Data Dictionary of the DRW. The user may optionally provide a vertical elevation (Elev), including the referenced vertical datum (Vert_Datum).

While the point-entry method described above provides a way to load point data for those without access to a GIS, the preferred method for defining new locations is through the use of an ESRI shapefile[4]. This method is preferred because it is versatile, allowing the user to upload points, lines, and polygons into *Oncor*; compact, needing only one record for a shapefile that may contain many individual locations; and extendible, allowing the user to include additional attributes about each location.

Users specify a shapefile as the spatial data source in the tab called 0_SpatialData_Shapefile, which is available in every DRW. Every new shapefile definition record must contain values for the following fields:

- **File_Path.** File path to File_Name on user hard drive; this is needed for the validation step described in Validation Mechanisms above.
- **File_Name.** Name of zipped file containing ArcGIS shapefile-format files with GIS data.
- **Alias_Group.** Grouping category for Standard-Value Name alias (see Appendix C - Data Quality for more on Alias Names).

[4] The shapefile format (see http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf) 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

- **File_RelateField.** The name of the field (attribute) in the shapefile that contains the feature name. The feature name needs to be the same as the names in the DET that identify a spatially distinct data record (e.g., a point ID field).

These are further described in the Data Dictionary of the DRW. The shapefile may contain additional attributes that can be associated with each feature. If elevation data is part of the shapefile, the name of the elevation attribute is specified in the field File_ElevField. The attribute with that name is thereby mapped to the *Oncor* attribute Elev (elevation). Similarly, if the shapefile contains an attribute identifying the site, then File_SiteField is specified.

2.7 DRW/DET Uploading Procedure

Users of this document must access the *Oncor* website to verify that they have the most recent version of the DRW and therefore the DET. This will ensure compatibility with the latest *Oncor* data loaders. Data generators will submit completed DET files via the *Oncor* web interface. Initially, this will be in the form of electronic mail, but ultimately there will be a data upload button for automated uploading. Only QC-checked data may be uploaded into *Oncor*. The data generator is responsible for assembling, verifying, and formatting the data, uploading the file to *Oncor*, and following up with the data custodian as necessary. A detailed explanation of the current plan for DET uploading procedure is presented in Appendix B.

3.0 Data Reduction Procedure Overview

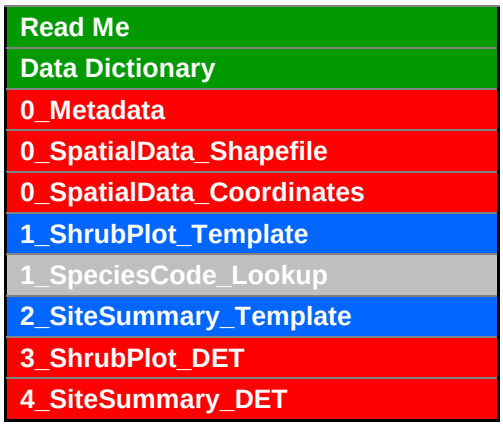
The DRP for a given data category is intended to provide stand-alone guidance for the data generator. Each DRP includes guidance for data QC, stepwise data reduction, and formatting for the DET. As a reminder, there are common features in the process across data categories. For example, there is a DRW in Excel to lead users through data processing that is specific to each DRP. Data generators are responsible for the quality assurance (QA)/QC of their data. It is suggested that the workbooks be saved separately in a file structure to aid in the tracing of data processing steps. The uploading of the final data (DETs, coded in the red-tabbed sheets) is completed through the *Oncor* website (http://oncor.pnnl.gov), where the entire DRW or independently saved DET Excel files can be uploaded along with any associated spatial data sets saved in the Shapefile format (*.shp).

4.0 Shrub/Scrub Vegetation Density[5]

This DRP describes an approach for preparing shrub plot data and subsequent site summary calculations for uploading to the *Oncor* database. It corresponds to the shrub field data collection protocol described by Roegner et al. (2009), which involves collecting data on tree species and diameter in 10x1-m rectangular plots. This procedure refers to a corresponding DRW—DRW_Shrub_v0.2.xlsm—This DRP addresses methods for using the DRW for field data entry, QC, calculations, and final data formatting for uploading into the *Oncor* database.

4.1 DRW Structure

The tab structure of DRW_Shrub_v0.2.xlsm is described below and each of the associated worksheets is further described in the Read Me tab of the DRW.



The **green** tabs are information only, the **blue** tabs provide templates for the data processing steps, the **gray** tab provides species codes and name, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

Cell Color Codes. Within the DRW, gray highlighting is used for cells containing example data. When gray is seen in cells with blue-highlighted tabs, it is example data that the data generator may view but does not need to delete. When gray is seen in cells with red-highlighted tabs, the data generator must delete those data before entering final data into the DET.

Tab Numbering. Tab numbers convey the steps used to complete the DRW. When two tabs have the same number, they are paired for one step in the process. The data generator copies raw data into the template. The example in the DET is for the data generator to review for guidance. The examples may also contain formulas, where applicable. These may be copied into the templates as needed.

[5] Prepared by Heida Diefenderfer and Amy Borde (360-681-3619; heida.diefenderfer@pnnl.gov).

Definitions. The column headers or "fields" of each tab in the DRW are defined in the data dictionary.

Required Fields. Required fields are indicated in the Data Dictionary tab. The key-field column of the Data Dictionary tab also specifies whether a field is a key field or has standard values. Validation methods, if applicable, are also given in the Data Dictionary tab.

4.2 Data Reduction Steps

The Read Me tab in the DRW also contains an abbreviated version of these instructions.

1. Enter metadata in the tab 0_Metadata and delete the corresponding example data with gray fill.
2. Using the 1_SpeciesCode_Lookup tab, identify the four- or six-letter species codes for all species included in your data set. To customize your datasheet, enter each species and size class if appropriate, in a column in 1_ShrubPlot_Template, beginning with column "H" to the right of the Notes column. Size classes are defined in the data dictionary and as follows:
   1 = 0 - 1.0 cm
   2 = 1.0 - 2.5 cm
   3 = 2.5 - 5 cm
   4 = 5 - 10 cm
   5 = 10 - 15 cm
   6 = 15 - 20 cm
   7 = 20 - 25 cm
   The number associated with the size class should be entered after the species code, for example "COSE-2."
3. Transcribe shrub plot data from field datasheets into the tab 1_ShrubPlot_Template. To reference the example data as needed, see 3_ShrubPlot_DET.
4. Quality-control check the shrub plot data.
5. Calculate the site summary data in the tab 2_SiteSummary_Template. To reference the example data as needed, see 4_SiteSummary_DET.
6. Quality-control check the calculations of site summary data.
7. To populate the DET, copy the shrub plot data over the example data in the tab 3_ShrubPlot_DET. Paste special, values only.
8. Similarly, copy the site summary data over the example data in the tab 4_SiteSummary_DET. Paste special, values only.
9. Quality-control check the copy-paste operations.

4.3 Spatial Data

Enter spatial data associated with the site or with the plots in either the 0_SpatialData_Shapefile tab or the 0_SpatialData_Coordinates tab as described in the section 2.6.7 Spatial Data, above. Specifically, for the shrub data, the location of each plot where data were collected is the preferred spatial data; however, these may not be readily available or collected by all projects. At a minimum, coordinates for the site must be provided. The location data is required for associating the shrub data with a sampling

location; if location data has not been collected in the field using GPS or another survey method, it can be estimated using GIS. The point data associated with each plot or site may be uploaded to *Oncor* using the DRW in one of two ways: 1) as coordinates or 2) as a shapefile.

Coordinates. If coordinates are entered, the 0_SpatialData_Points tab is used and the names in the "PointID" column must match those in "POINT ID" found in column F of tab 3_ShrubPlot_DET. If point data for the plots are not available then points for the site are used and the names in the "SiteID" column must match those in the "SITE ID" columns of tabs 3_ShrubPlot_DET and 4_SiteSummary_DET.

Shapefile. If a shapefile is submitted, the 0_SpatialData_Shapefile tab must be completed. The "File_RelateField" column is used to identify the field in the shapefile that has the point names for each plot; the points must be named the same as those in "POINT ID" found in column F of the tab 3_ShrubPlot_DET. The "File_SiteField" column is used to identify the field in the shapefile which has the point names for the sites; the points must be named the same as those in the "SITE ID" columns of tabs 3_ShrubPlot_DET and 4_SiteSummary_DET.

4.4 Quality Control

Quality control is the responsibility of the data generator. Recommended methods described in this report include a 100% check of transcribed data and a 10% random check of calculations. Quality control may be conducted by adding additional worksheets to the DRW, or using printed copies. In either case, a record of the person who conducted the check and the level of the check (e.g., 10%, 100%) must be kept in the electronic or paper file to ensure a transparent and complete data package. For additional details on QA/QC, see Section 2.3 and Appendix C.

4.5 Uploading Data to *Oncor*

Please see Section 2.6 in the basic material for a description of these procedures.