U.S. DEPARTMENT OF ENERGY

PNNL-22922

Prepared for the U.S. Army Corps of Engineers, Portland District,
under an Interagency Agreement with the U.S. Department of Energy
Contract DE-AC05-76RL01830

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

October 2013

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

This document was printed on recycled paper.

(8/2010)

PNNL-22922

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

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

October 2013

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

Pacific Northwest National Laboratory
Richland, Washington 99352

Preface

This *Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates* is designed to support the *Oncor* geodatabase for the Columbia Estuary Ecosystem Restoration Program (CEERP). The CEERP is implemented by the Bonneville Power Administration and U.S. Army Corps of Engineers Portland District (USACE) to help mitigate for effect of operation of the Federal Columbia River Power System on salmonid populations listed under the Endangered Species Act. The handbook and geodatabase are being developed by the Pacific Northwest National Laboratory (project number 61657) with funding from the USACE Portland District, technical lead Ms. Cynthia Studebaker. The effort is coordinated regionally in the lower Columbia River and estuary through the USACE's Anadromous Fish Evaluation Program (study code EST-P-12-01) and the Lower Columbia Estuary Partnership's Science Work Group. The following data categories are covered: water-surface elevation and temperature, sediment accretion rate, photo points, 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 handbook is intended for use by scientists collecting monitoring and research data for the CEERP. The ultimate goal of *Oncor* is to provide quality, easily accessible, geospatial data for synthesis and evaluation of the collective performance of CEERP ecosystem restoration actions at a program scale.

A suggested citation for this document follows: Diefenderfer HL, NK Sather, AB Borde, JA Serkowski, AM Coleman, and GE Johnson. 2013. *The Oncor Geodatabase for the Columbia Estuary Ecosystem Restoration Program: Handbook of Data Reduction Procedures, Workbooks, and Exchange Templates*. PNNL-22922. Report version October 2013 submitted to the U.S. Army Corps of Engineers, Portland, Oregon, by the Pacific Northwest National Laboratory, Richland, Washington.

Acronyms and Abbreviations

AEMR	action-effectiveness monitoring and research
BPA	Bonneville Power Administration
CEERP	Columbia Estuary Ecosystem Restoration Program
cm	centimeter(s)
CoC	Chain of Custody
CPUE	catch per unit effort
CRD	Columbia River Datum
DET	data exchange template
DRP	data reduction procedure
DRW	data reduction workbook
ft	foot(feet)
GIS	geographic information system
GPS	global positioning system
ID	identification
kPa	kilopascal(s)
LCEP	Lower Columbia Estuary Partnership
LCRE	lower Columbia River and estuary
LRR	Lower River Reach
m	meter(s)
m^2	square meter(s)
NAVD88	North American Vertical Datum of 1988
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
PNAMP	Pacific Northwest Aquatic Monitoring Partnership
PNNL	Pacific Northwest National Laboratory
psi	pound(s) per square inch
QC	quality control
USACE	U.S. Army Corps of Engineers Portland District
VBA	Visual Basic for Applications
WSE	water-surface elevation

Contents

Figures

Tables

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.*

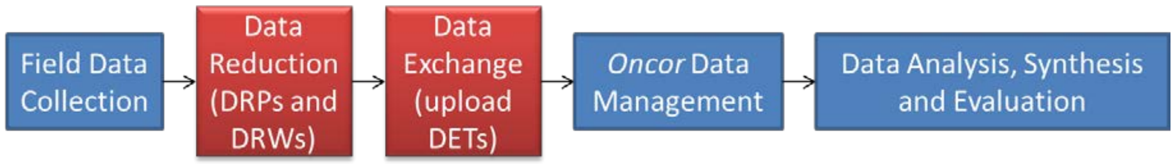

Figure 1.2. Data flow for AEMR and other data intended to support CEERP. The red boxes signify material that is covered in this handbook.

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 densities (under construction)
Vegetation plantings	pertained to planting success	not covered
Fish community	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 quality (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.

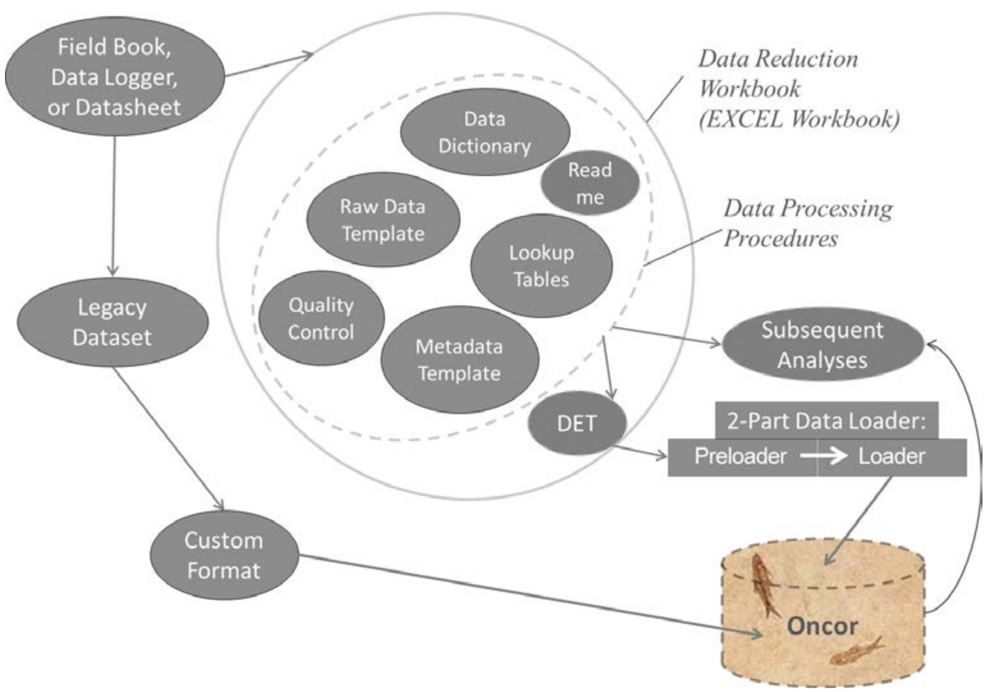

Figure 2.1. Schematic of data flow from field data to *Oncor*. Data flow includes data reduction and entry into a Data Exchange Template (DET) by the data generator, prior to uploading to the *Oncor* database. The data reduction workbook also includes spatial data.

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 Data Reduction Workbooks and Data Exchange Templates

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

2.5.1 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.

[3] Under construction.

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.

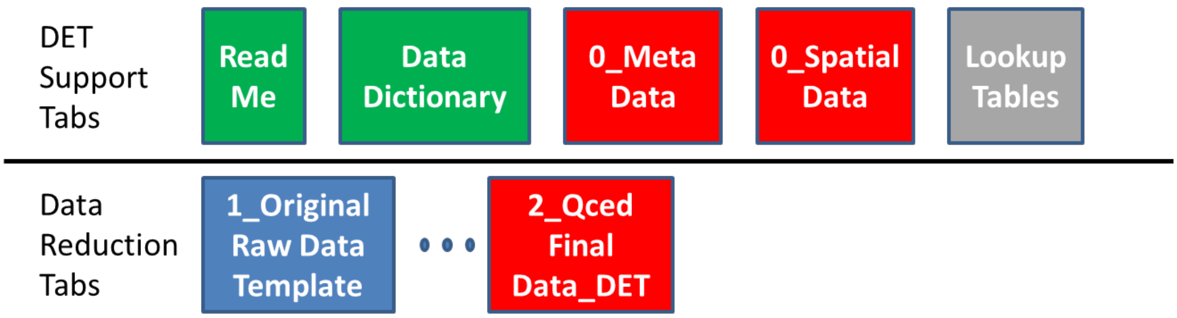

Figure 2.2. Generic Data Reduction Workbook structure including the Data Exchange Template (in red). The three-part general processing sequence is represented by 0, 1, and 2. Each box represents a worksheet tab in the workbook. Green tabs are informational. Blue tabs consist of a template with column 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.** Hidden 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*.

2.5.2 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.

2.5.3 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.

2.5.4 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.

2.5.5 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*.

2.5.6 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_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).
- **Program_Name**. Standard value for the name of the program associated with the data.
- **Project_Name**. Standard value for the name of the project that generated the data.
- **QC_Meth**. Yes/no indicates whether the data were QC-checked.
- **Instrumentation.** If applicable to a given data category, the metadata sheet in the DRW will also identify instrumentation.
- **DET_Type** and **DET_Version**. Additional DET metadata, such as DET_Type and DET_Version, are harvested from the Read Me tab.

The fields described as "read only" are automatically assigned by *Oncor* at the time of DRW creation and cannot be modified by the data generator. Fields that require a standard value on this and all other DRW tabs use a dropdown list of allowed values to restrict data entry. If the user types an invalid value into such a field, an error will occur when the user exits the cell. The dropdown list is accessible by clicking on the cell and then clicking on the down-arrow that appears to the left of the cell. If a desired item is not in the dropdown list, the user may enter it using the Standard Values Form, which appears when the Standard Values button is pressed.

2.5.7 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:

- **Alias_Name.** User defined point name (see Appendix C - Data Quality for more on Alias Names).
- **Alias_Group.** Grouping category for Standard-Value Name alias (see Appendix C - Data Quality 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).
- **File_RelateField.** The name of the field (attribute) in the shapefile that contains the feature name. The feature names need to be the same as the names in the field in the DET that identifies a spatially distinct data record (e.g., a point ID field or a Site 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

[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

mapped to the *Oncor* attribute Elev (elevation). Similarly, if the shapefile contains an attribute identifying the site, then File_SiteField is specified and the attribute with that name is mapped to the *Oncor* attribute Related_Location.

2.6 DET Uploading Procedure

To ensure compatibility with the latest *Oncor* data loaders (responsible for extracting data from the Excel workbooks and importing to an enterprise database), users of this document must access the website to verify that they have the most recent version of the DRW and therefore the DET.

Data generators submit completed DET files via the *Oncor* web interface. Only QC-checked data may be uploaded into *Oncor*. A detailed explanation of the stepwise DET uploading procedure is presented in Appendix B. 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.

Step 1. Data generator processes field-collected data according to the appropriate DRP, using the DRW to produce a DET file.

Step 2. Data generator submits completed DET to *Oncor* web site and receives either an acceptance in the form of a preload file or a rejection in the form of an error file.

Step 3. If the DET is rejected, the data generator must correct problems, with the help of the data custodian (if needed), and resubmit the DET until it is accepted.

Step 4. Data generator approves the preload file, submits the load, and receives an upload confirmation.

Step 5. Data generator queries the database for selected new records to verify their presence.

Step 6. If the data generator created new standard values, data associated with those records will be marked as provisional in *Oncor* until the data custodian approves the new entries.

After a DET is uploaded, the file is processed by a Preloader and a Loader, as described below. To verify successful data upload, the Preloader performs the following actions on the DET:

- Runs more-sophisticated validation checks not performed in DRW file.
- Translates the file to a standard *Oncor* format. This format provides a one-to-one correspondence to the *Oncor* data model.
- Generates either a preload file or an error file, depending on whether errors were detected. If errors occur during validation or translation, the error file provides information about offending records. If no problems are found, the preload file consists of the DET in *Oncor* format.
- If an error file is produced, the data generator must resolve the identified issues, correct the DET, and resubmit the file to the Preloader. Resolution may involve interaction with the data custodian.
- Once the Preloader creates a preload file, the data generator, after reviewing its contents, may submit it to the Loader, which then populates the web-accessible database.

The Loader generates a load-confirmation file providing load statistics. It is suggested that data generators perform queries (through the *Oncor* website) on their newly loaded data to further verify their data were uploaded correctly.

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).

The ensuing sections contain DRPs, DRWs, and DETs for the following data categories:

- water-surface elevation and water temperature
- sediment accretion rate
- photo points
- herbaceous wetland vegetation cover
- tree plots and site summary
- fish catch and density
- fish size
- fish diet
- fish prey
- Chinook salmon genetic stock identification.

4.0 Water-Surface Elevation and Temperature DRP[5]

The *Data Collection Protocols* identify hydrology, and specifically, measures in the variation of water-surface elevation (WSE) as part of the "Core Monitored Metrics" (see Section 2.2.1 – Hydrology of Roegner et al. 2009). Using pressure transducers with automated hourly logging is the recommended approach for capturing the variability in water levels. Section 4.1 – Hydrology (Protocol 1) of Roegner et al. (2009) provides the field deployment and general calculation and analysis methodology. The procedure presented below covers processing data collected with a data-logging pressure transducer (hereafter water-level logger), methods for calculating WSE, QC measures, and final data formatting for entry into the *Oncor* database.

4.1 DRW Structure

This procedure refers to the DRW_WSE_v0.X.xlsm Excel workbook, which has the following tab colors and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_AtmCorrected_Example
2_ElevCorrected_Example
2_TimeCorrection_Lookup
3_Deployment_DET
4_Measurement_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

The metadata and spatial data entry steps are described in the section on basic material of this handbook. The process converting field collected water depth data to water surface elevation is described below.

An organized file structure is important for keeping track of the various files created throughout the data correction process for WSE. A suggested file structure includes the following folders:

- Sensor Log
- Step1_Original Files (downloaded from the sensor)
- Step2_Corrections

[5] Prepared by Amy Borde, Shon Zimmerman, Andre Coleman, and Ron Kaufmann. For more information, please contact Amy Borde (360-681-3663; amy.borde@pnnl.gov).

- Step2a_AtmCorrection
- Step2b_ElevCorrection
- Step3_QAQC
- Step4_Final.

Suggested filenames and locations in the file structure are provided in the sections below, with the following notation to indicate nested folders: \Step2_Corrections\Step2a_AtmCorrection. Note: a specific directory structure for WSE data files is prescribed because of the typically large amount of data and the complexity of the data reduction steps for WSE. Example filenames are also provided in the steps below in the following format:

[SITE]_elev_correct_[*mmddyyii*].xlsx

where, [SITE] refers to a user defined site name or code and [*mmddyyii*] refers to the six digit date in month, day, year format, with two-letter initials at the end for the person who conducted the correction, QA/QC, or finalization step.

4.2 Data Reduction Steps

The steps taken to process the WSE data are described below.

Step 1. Save and read the Sensor Log.

The Sensor Log is a way of tracking and organizing the information collected as part of the field deployment and retrieval of water-level loggers. The Sensor Log usually contains information such as the following:

- serial number of each sensor for data tracking
- deployment and retrieval date and time
- initials of personnel conducting deployment/retrieval
- water depth above the sensor at the time of deployment and retrieval. This information is used to verify that the sensor is measuring correctly.
- distance from the sensor to the top of deployment post (or some way of marking the position of the sensor) at deployment and retrieval. This information is critical to determining the sensor elevation if elevation is surveyed at the top of the deployment post. This distance is also useful for determining if the sensor moved during the deployment period.
- distance from the sediment to the top of the post. This measurement is not needed for the processing, but indicates whether conditions surrounding the sensor changed during the deployment (e.g., erosion, deposition) or if the post moved.
- elevation of the top of the deployment post and/or elevation of the sensor
- notes about the deployment or retrieval.

Step 2. Perform the atmospheric pressure correction.

Water-level loggers record absolute pressure, which is converted to water depth by processing software that uses atmospheric pressure. Absolute pressure includes atmospheric pressure and water pressure. Atmospheric pressure is nominally 100 kPa (14.5 psi) at sea level and must be separated from the water pressure to determine water depth. However, atmospheric data fluctuate with weather and altitude; left uncompensated, barometric variations could result in errors of 0.6 m (2 ft) or more.

To compensate for barometric pressure changes, hourly atmospheric pressure data need to be acquired from either a local meteorological station (i.e., from the nearest station listed under the National Oceanic and Atmospheric Administration's [NOAA's] National Climatic Data Center [NCDC]) or from another pressure sensor that is deployed near the site and out of the water, specifically to collect atmospheric pressure. In either case, the time period during which atmospheric pressure data were collected needs to span the same time period the water level logger was deployed and, ideally, should have the same temporal resolution (i.e., hourly). Some software allows for the direct input of atmospheric data when they are collected with the same kind of data logger. If this is the case, follow the manufacturer's instructions for the correction process. If atmospheric are needed they can be downloaded from the NCDC at:
http://cdo.ncdc.noaa.gov/pls/plclimprod/poemain.accessrouter?datasetabbv=DS3505

Follow these steps at the NCDC website:

1. Click "I Agree to these terms (continue)".
2. Click "Continue With SIMPLIFIED Options".
3. Select the "Country – United States" radio button and then click "Continue".
4. Select "Oregon or Washington" from the dropdown menu and click "Continue".
5. Select the station of interest that has a comparable data range (e.g., Portland Troutdale, WBAN# 24242) and click "Continue".
6. Choose your date and time window from the dropdown menus and click "Continue". (Note: hourly-only data can be selected; however, this can sometimes result in an incomplete data set. Check the resulting data set to be sure adequate data exist; if not, re-submit the request without hourly-only data being selected.)
7. Check the "Inventory Review" box, type in your email address, and click "Submit request". You will receive an email from NCDC with links to your data. Save the data as a text file and proceed to the next step.

Format the data file according to the software's specifications for using an imported atmospheric data file. For example, Onset Computer HOBOware® software requires a tab-delimited text file with only three columns: "Date", "Time (in same time zone as data was collected)", and "Sea level pressure (SLP) (mbar)", with the data in the following format: mm/dd/yyyy, hh:mm:ss, and the SLP should have one decimal place visible. Formatting can be done in Excel or any other worksheet.

Follow the manufacturer's software instructions for converting pressure data to water depth using the atmospheric data file and specifying the water density. Water density is determined based on salinity and/or the water temperature if freshwater. After the water depth is calculated, export the data for further manipulations in a worksheet.

> Example filename: [*SITE*]_atm_correct_[*mmddyyii*].xlsx
> Located in: \Step2_Corrections\Step2a_atm_corrected

The file exported from the water depth processing software will look something like the tab provided in the DRW titled: 1_AtmCorrected_example.

Step 3. Perform the elevation correction.

Note: In the future, a script will be written to automate the process in this section and will be included as part of the DET package.

Water-surface elevation is determined by the sum of the sensor depth below the water-level surface and the surveyed elevation of the water-level logger (see Roegner et al. 2009, Section 4.1 Hydrology). The WSE will be relative to the datum in which the elevation was collected and can be converted into different data if desired.

For the following steps, please also refer to the “2_ElevCorrected_Example” tab in the DRW.

1. Save the file “atm_correct” as “elev_correct”.

   > Example filename: [*SITE*]_elev_correct_[*mmddyyii*].xlsx
   > Located in: \\Step2_Corrections\Step2_elev_corrected

   Remove unnecessary data columns, insert a new column at column A, and populate it with the site code. Ensure that the first row of the worksheet contains the column headers, with no extra rows above it. The only needed columns are as follows:

   > “Site Code”
   > “Date time”
   > “Temp, °C”
   > “Sensor Depth, meters”

   If the atmospheric pressure and the absolute pressure do not occur at the same logging interval, blank cells will occur in the Sensor Depth column and need to be removed.

2. Often the data are collected in the time zone in which the sensor is launched (e.g., daylight savings time, which is GMT-7 in the Pacific time zone). If the deployment period spans a time change, it will be necessary to convert to the “local” time, which will switch between daylight savings and standard time on the correct date.

   a. If converting to local time then the following steps can be followed:

      i. From the data reduction workbook, copy the TimeCorrection_Lookup worksheet from its original workbook to the WSE workbook you are currently editing.

      ii. Insert two columns for “Date Time, GMT+00:00 and “Date Time, local”.

iii. The time must first be standardized to GMT+00:00 by populating the first new column with the sum of the GMT-7 date/time and 7hrs/24hrs (0.292) for converting from Pacific Daylight Time or 8hrs/24hrs (0.333) for converting from Pacific Standard Time.

iv. Copy the formula found in cell C3 from the example worksheet 2_ElevCorrected_Example in the DRW.

v. Paste the copied formula into the first blank cell in your second new column below the Date Time, local header. Populate the rest of the column with the formula.

3. Remove the data with time stamps that are outside the deployment period (i.e., times when the logger is recording, but has not been put in the water) by referring to the deployment/retrieval times in the Sensor Log. Be sure time zones are the same in the Sensor Log and in the data file.

4. In the sensor depth column, if the sensor was out of the water at any time the values will be negative. These negatives need to be changed to 9999 so they are easily identifiable during later processing.

5. To calculate the WSE, get the sensor elevation value from the Sensor Log and in a column named "Water_Surface_Elevation," sum the sensor elevation value and the values from the "Sensor Depth, meters" column. The resultant value is the WSE relative to the vertical datum, NAVD88 (North American Vertical Datum of 1988; assuming that was the datum used to collect the sensor elevation).

6. An optional step is to convert the WSE from NAVD88 to the Columbia River Datum (CRD), a low-water datum developed by the USACE. The conversion varies with location in the river and has been calculated by the USACE for each river mile. The conversions are available in NOAA Technical Memorandum NOS CS 22 (Xu et al. 2010).

Step 4. Quality control check.

This step is necessary to ensure the data was collected accurately and that the corrections and data processing steps were conducted properly. See QA/QC of Water Surface Elevation Data below for details.

Step 5. Finalize the WSE data.

This step completes the preparation of the WSE data for use in analysis or by others.

1. If uncorrectable errors are found then the QA/QC version of the file should not be finalized. If there are no errors or any detected errors are corrected then save the file in the Final folder.

   Example filename: [*SITE*]_FINAL_[*mmddyyii*].xlsx
   Located in: \\Step4_Final

2. Copy all data and paste as values so there are no formulas in the cells.

3. Delete unnecessary data:

   a. Delete the elevation conversion values.
   b. Delete any of the date/time columns except the desired time zone.

A time-series of WSE values is the final output from the procedure. The time-series data show plainly the hydrologic patterns at the site over time. Also, derived data such as the sum exceedance value and area-time inundation index use these WSE data.

Step 5. Enter the data into the DRW and prepare the DET.

The finalized WSE and temperature data will be uploaded into the *Oncor* database. Data can be copied and pasted into the pre-formatted red DET example tabs provided in the DRW or the Final data file created in the previous step can be formatted to be identical to the Measurements_DET example (i.e., with the necessary columns added). The three tabs required for uploading data to *Oncor* are as follows:

- 0_Metadata. This tab is where information about the DET is entered: the data point of contact, the filename, the program and project under which the data were collected, and the method used for data collection are input here.

  Note: All data must be entered using the dropdown arrows at the right of the cell. If the needed option is not available, it can be added by clicking on the Standard Values button at the top of the page.

- 3_Deployment_DET. Each deployment period should be entered as an individual record (row). Some fields can be filled out by typing directly in the cell, while others require the use of dropdown arrows to enter a previously specified value. Again, if the needed option is not available in the drop down, it can be added by clicking on the Standard Values button at the top of the page. Descriptions of the fields can be found in the data dictionary. The Water_Elevation_Location field refers to the name of a point in the 0_SpatialData_Coordinates tab or a GIS shapefile. Information from the QA/QC procedure can be noted in the Instrument _Deployment_Notes column.

- 4_Measurement_DET. This tab is where the WSE and temperature data are entered, as follows:

1. In the first two columns enter the Water_Elevation_Instrument ID and Instrument_Deployment_Date from the 1_Deployment tab. These values need to be copied down the entire length of the data so that these identifying fields are associated with each individual record (row.

2. The next four columns (D−G) can be directly copied from the Final data file(s) ([*SITE*]_FINAL_[*mmddyyii*].xlsx) and pasted into the following columns in the DET:

DET Column Header	Final WSE Data Description
Water_Measurement_Date	The date/time column in “local” time
Water_Temperature	Temperature data in deg C
Temperature_Sensor_Exposed	Temperature exposure flag
Water_Surface_Elevation	WSE data in meters relative to NAVD88.

3. The Instrument_Measurement_Notes column can be used to note anything that was observed about individual measurements; for example, the field water depth measurements taken at deployment and retrieval could be entered in this column.

4. The DB_Access column is where the user can define the level of access for the individual deployments in the data set.

4.3 Quality Control

The QA process and the QC steps in place for the WSE data are designed to ensure the quality of data for any ensuing tasks involving these data. The following steps should be completed by someone other than the person who conducted the data processing. In addition, all QA/QC steps should be listed in a QA/QC Log regardless of the results. Any issues found during QC should also be marked and commented on in the data file so they can be corrected for the final version. For additional details on QA/QC, see Section 2.3 and Appendix C.

1. Complete an atmospheric data correction check.

   The atmospheric correction check is designed to ensure the atmospheric data were appropriate for the site and to identify possible sensor malfunction. Pre- and post-deployment of the sensor should record a "depth" value close to 0 and not vary more than a few centimeters.

   Open the [*SITE*]_atm_correct_[*mmddyy*].xlsx file from the location at:
   \\Step2_Corrections\Step2a_atm_corrected

   Using the data in the Sensor Log determine the date/times of deployment and retrieval. Check the sensor depth column before and after the dates listed for deployment and retrieval. In the QA/QC Log, record the greatest value within 2 hours prior to deployment or after retrieval.

2. Verify water depth.
   Field water depth measurements are essential to confirming that the pressure sensor is measure accurately.

   Open the [*SITE*]_elev_corrected.xlsx file from the location:
   \\Step2_Corrections\Step2b_elev_corrected

   Find the times in the Excel file closest to the times of deployment and retrieval. Compare the field water level measurement in the Sensor Log with the sensor depth measurement in the water-level file (make sure the times and time zones are the same). Note any differences in the QA/QC Log. See the tab 1_AtmCorrected_Example in the DRW and notes highlighted in blue for an example.

   If the Sensor Log has notes about potential issues with this measurement then copy the note into the QA/QC Log. If the field measurement was not taken at exactly the same time as the sensor measurement then tide tables may need to be referenced to determine whether the tide could have resulted in a difference in the measurements.

3. Perform an elevation correction check.

   a. Check the depth sensor elevation value in the Sensor Log to make sure the correct value was used.

   b. Check the CRD_conversions_Lookup tab to make sure the correct conversion value was used.

c. Check the equations where these values are used to make sure that they are correctly calculated for the entire column.

4. Check for sensor movement during deployment.

   Within the Depth Sensor Log, check for differences in the measurement of the sensor to top of post ("Dist to top of post" in the log) between deployment and retrieval. Note any difference in the QA/QC Log and differences greater than 10 cm require an investigation of the water-level data to see if the time of change can be determined and corrected.

5. Check the Sensor Depth column for 9999 values.

   All negative values, which were logged when the sensor was above the water surface, should have been replaced with 9999 values. This check can easily be done using the Filter function in Excel.

6. Check the Temperature Sensor Exposed column.

   Ensure that the temperature values are flagged appropriately when the temperature sensor was exposed. Again, this can be done using the Filter function in Excel for all depth values less than 7 cm.

7. Create a hydrograph.

   The last step of the QA/QC process is the creation of the hydrograph to check for any anomalies or odd trends in the data.

   a. Create a new tab in the workbook and label it "graph".

   b. Copy the "Time" and "WSE" columns into the new tab. Select the relevant time period to plot.

   c. In the ribbon, go to the Insert tab and select a Scatter with Smooth lines Plot type. Select the time column for the x value and select the WSE column for the dependent variable.

   Note any anomalies in the QA/QC Log. This hydrograph is also useful for looking at hydrologic patterns in the data.

If there are errors in the water-level file, make sure they are highlighted and commented on, then save the file and change the file name from "elev_correct" to QA/QC with date and initials (ii) at the end of the filename in the format mmddyyii.

Example filename: [*SITE*]_QAQC_[*mmddyyii*].xlsx
Located in: \\Step3_QAQC

4.4 Steps for Reducing Temperature Data

If temperature data were collected at the same time as the WSE data then the following procedure applies. No correction is necessary for the temperature data; however, measurements taken when the temperature sensor was out of the water should be flagged for removal prior to any analysis. The data can be either replaced by 9999, similar to the WSE data, or the data can be simply flagged so the in-air measurements are still part of the data set. The latter method allows the user to evaluate air temperature relative to water temperature, but care must be taken to ensure these measurements are not included in any water temperature analyses.

1. Determine the location of the temperature sensor relative to the pressure sensor in the sensor housing or as measured in an adjacent sensor.

2. Use the distance value between the sensors to determine when the temperature sensor is exposed. For example, with Onset HOBO data loggers, the temperature sensor is located 6 cm above the pressure sensor. Therefore, any temperature data collected when the water depth was less than 7 cm should be flagged. See the 2_ElevCorrected_Example tab in the DRW for an example of how this can be calculated.

4.5 Uploading Data to *Oncor*

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

5.0 Sediment Accretion Rate DRP[6]

This DRP describes an approach for preparing sediment accretion stake data and subsequent mean and sediment accretion rate calculations for uploading into the *Oncor* database. It corresponds to the sediment accretion rate field data collection protocol provided by Roegner et al. (2009), which is a simple approach to obtain a coarse rate by using two stakes driven into the ground a meter apart that serve as markers for seasonal or annual data collection. This procedure refers to a corresponding DRW. The first step in creating a DRW is to download the most recent version of the Excel file, the first version of which was named DRW_SedAccretion_v0.X.xlsm, from the *Oncor* website: http://oncor.pnnl.gov/. This version has the most recent *Oncor* conventions and is compatible for uploading data to the database. This DRP addresses methods for using the DRW for field data entry, QC, calculations, and final data formatting for uploading into the *Oncor* database.

5.1 DRW Structure

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

Read Me
Data Dictionary
0_Metadata_Template
0_Metadata_Example
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_SurveyData_Template
1_SurveyData_Example
2_MeanDistance_Template
2_MeanDistance_Example
3_SedRates_Template
3_SedRates_Example
4_SurveyData_DET
5_MeanDistance_DET
6_SedRates_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, 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.

[6] Prepared by Heida Diefenderfer (360-681-3619; heida.diefenderfer@pnnl.gov).

Tab Numbering. Tab numbers convey the steps used to complete the DRW. When two tabs have the same number, they are paired template and example data tabs. The data generator copies raw data into the template. The example is for the data generator to review for guidance. The examples also contain formulas, where applicable, to calculate mean distance between ground and the plane at the top of the stakes, and the annual sediment accretion rate. These may be copied into the templates as needed.

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" also specifies whether a field is a key field[7] or has standard values. Validation methods, if applicable, are also given in this field.

5.2 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_Template. To reference the example data as needed, see the 0_Metadata_Example.
2. Transcribe survey data from field datasheets into the tab 1_SurveyData_Template. To reference the example data as needed, see the 1_SurveyData_Example.
3. Quality-control check the survey data.
4. Calculate the mean distance in the tab 2_MeanDistance_Template. To reference the example data as needed, see the 2_MeanDistance_Example. The example also contains the formulas for calculated fields.
5. Quality-control check the calculations of mean distance.
6. Calculate the sediment accretion rate in the tab 3_SedRates_Template. To reference the example data as needed, see the 3_SedRates_Example. The example also contains the formulas for calculated fields.
7. Quality-control check the calculations of sediment accretion rate.
8. To build the Data Exchange Template tabs in the DRW, copy survey data over the example data in the tab 4_SurveyData_ExampleDET and use "Paste special, values only."
9. Copy the mean distance data over the example data in the tab 5_MeanDistance_ExampleDET. Paste special, values only.
10. Copy the sediment accretion rate data over the example data in the tab 6_SedRates_ExampleDET. Paste special, values only.
11. Quality-control check the copy-paste operations.

[7] A field that is required to uniquely identify the record, such as sampling date and location.

5.3 Spatial Data

For sediment accretion data, the spatial data for each set of stakes consists of one point, because the distance between the stakes is only 1 m. This location point is required data for associating the sediment accretion data with a place, and if the location point has not been collected in the field using GPS or another survey method, it can be estimated using a GIS. In some cases, an elevation at ground level will also have been collected. The XY point data associated with each set of sediment accretion stakes may be uploaded to *Oncor* using the DRW in one of two ways: 1) as coordinates in columns K and L of tab 4_SurveyData_DET, or 2) as a shapefile. If a shapefile is submitted, columns O and P of tab 4_SurveyData_DET must be completed. The shapefile must contain a column named "SS ID" that contains the latitude/longitude points for each SS ID found in column F of tab 4_SurveyData_DET. This is how location data from the shapefile are associated with sediment accretion data from the DET in *Oncor*. Therefore, the SS ID numbers for every set of sediment accretion stakes must be unique from all others contained in that shapefile. Use Columns M and N of tab 4_SurveyData_DET to upload elevation data, if available.

5.4 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.

5.5 Uploading Data to *Oncor*

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

6.0 Photo Points DRP[8]

Photo points are used as a qualitative means to document site conditions and can be used as a tool to examine site-level changes through time. Procedures for collecting photo points are documented by Roegner et al. (2009).

6.1 DRW Structure

This procedure refers to DRW_PhotoPoint_v0.X.xlsm that has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_PhotoPoint_Template
2_PhotoPoint_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

6.2 Data Reduction Steps

Step 1. Merge individual photos as appropriate.

Depending on site conditions, photo points may be collected as a single photo or as several photos that are merged into a panoramic photograph of the site. Several software programs are available for photo merging. In addition, many digital cameras and cellular phones are equipped with panoramic capabilities that can readily stitch multiple images together "on the fly."

Step 2. Populate the DRW with supporting data, including location coordinates.

Before uploading photo points into *Oncor*, the 1_PhotoPoint_Template must be populated with required information such as the site, date, and time the photo was collected. In addition, each photo must be referenced to a geographic location. Coordinates for the actual location at which the photo point was collected are preferred; however, these may not be readily available or collected by all projects. At a minimum, coordinates for the site must be included in the1_PhotoPoint_Template. An additional data requirement is to include the photo file name. This ensures appropriate linkage between the photo and the data within the 1_PhotoPoint_Template worksheet. It is recommended that the photo file name include the site and date code to provide additional clarity and prevent duplicative naming of photo files. For

[8] Prepared by Nichole Sather, Heida Diefenderfer, and Amy Borde. For more information, please contact Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

example, the photo file name “C0209.jpg” denotes a photo point collected at site C during February 2009. Additional information such as project codes could also be included, but these naming conventions are at the discretion of the data generator.

Step 3. Create the photo point DET.

The PhotoPoint_DET is created by following QA procedures (see section below), an example of which is found in the 2_PhotoPoint_DET.

6.3 Quality Control

An independent party must review the merged photo files and the supporting data to ensure accuracy. The reviewer must also verify that photos correspond to correct file names and verify that there are no transcription errors. For additional details on QA/QC, see Section 2.3 and Appendix C.

6.4 Uploading Data to *Oncor*

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

7.0 Herbaceous Wetland Vegetation Cover[9]

The information contained in this section outlines an approach for preparing herbaceous wetland cover data and subsequent average percent cover calculations for uploading into the *Oncor* database. The procedure outline here is focused on data collected using 1-m^2 quadrats and a random systematic sampling design as described by Roegner et al. (2009). While there are similarities between this kind of data and the data and accompanying field codes from other wetland vegetation types (e.g., shrub or forested wetlands), the data collection methods and the resulting calculations would differ and therefore require distinct DRWs and DETs. The procedure described below addresses methods for field data entry, QC measures, calculating species averages, and final data formatting for entry into the *Oncor* database.

7.1 DRW Structure

This procedure refers to DRW_HerbVegetation_v0.X.xlsm that has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_HerbCover_Template
1_SpeciesCode_exampleLookup
2_HerbCover_exampleDET
3_HerbAvgCover_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

The metadata and spatial data entry steps are described in the section on basic material of this document. The process for entering and reducing field collected herbaceous wetland vegetation data are described below.

7.2 Data Reduction Steps

The following steps and accompanying DRW outline the process for preparing herbaceous wetland vegetation species cover data collected in the field to average cover estimates from a given site. The steps describe an approach for managing data using Excel worksheets and workbooks. Data managed in other software programs will need to export data accordingly to be successfully uploaded into *Oncor*.

[9] Prepared by Amy Borde (360-681-3663; amy.borde@pnnl.gov).

Step 1: Electronic transcription from field datasheets.

Use the 1_HerbCover_Template from the DRW_HerbVegetation_v0.1.xlsm to enter cover data collected from the field. The template includes columns associated with typical vegetation data collection procedures, not all of which are required for uploading data to the *Oncor* database. The column headers are defined in the data dictionary of the herbaceous vegetation DRW. The following steps refer to additional worksheets that may be useful within a data-entry workbook and provide some specific information about the columns within the data-entry worksheet template.

1. Copy the **1_HerbCover_Template** into a new workbook and save it with a name that includes RAW in the filename, designating it as the raw data file.
2. Create a worksheet titled "Data status" to enter information about the progression of the data processing, such as who entered the data and on which date, who QA/QC-checked the data, and notes about any changes, such as species name changes, that occur with data.
3. Enter the field data according to template provided in the DRW. The fields in the template are as follows:
   a. ID is a unique identification created for each individual quadrat of data collected. For example, it can be a combination of the sample period, the site code, the transect, and the quadrat. If locations of individual quadrats are collected the quadrat ID must match the ID field in the shapefile or coordinates provided in the Spatial Data tab.
   b. The SamplePeriod is needed to ensure that the database can query based on individual sample events even if they spanned more than 1 day or a site was sampled more than once a year, for example. The sample period should be indicative of the sample periodicity whether monthly (June2009), seasonally (Summer2009), or annually (2009).
   c. If elevation was collected as part of the vegetation survey, enter the elevations associated with each quadrat in the appropriate row. A lookup table can be used to populate this column based on the ID column.
   d. Enter species codes in the column headers to the right of the Elevation column. In the template, there is only one column with a header "Species Codes…"; this should be expanded to numerous columns, each with the header for an individual species. It does not matter in which order the species are listed, but it is easier to enter and to QC-check content when the species are in the same order as the physical datasheet.
   e. Enter notes about individual quadrats in the Notes column.

1. Enter notes about the monitoring site as a whole into a separate worksheet titled "Site Notes."

Step 2: Finalize herbaceous cover data.

Upon completion of the QA/QC process (described below), populate the **2_HerbCover_DET** worksheet with the QC-checked data using the simple copy-and–paste value buttons.

Step 3: Average species cover calculations.

Once the final cover data have been pasted into the **2_HerbCover_DET** worksheet, click the Calculate Cover button located in the upper left corner of the DET. This feature calculates the average cover of each species per site in the subsequent worksheet **3_HerbAvgCover_DET** by recognizing the new data that have been input into the **2_HerbCover_DET** worksheet.

7.3 Quality Control

To ensure there were no errors when transcribing field data to electronic format, QC measures should be implemented following initial data entry. Using the field datasheets, a 100% check of data entry and a 10% check of formulas is recommended. This check must be performed by an individual who did not perform the initial data-entry tasks described in Step 1 above. For additional details on QA/QC, see Section 2.3 and Appendix C.

It is possible for a data generator to maintain all QA/QC files as separate worksheets in the DRW. However, we find creating separate QA/QC files provides an efficient way to track the progression of the data from raw to final format. The process for implementing data QA should occur in a manner that is best suited to the project needs. We recommend the following steps:

1. Save the Raw data file created in Step 1 as a new Excel file and rename it using ”QAQC” in the filename designating it as the QA/QC file.
2. In the Data Status worksheet, insert the name of the person performing the QA check and the date the check was completed.
3. Check 100% of the entered data. Highlight all transcription errors and formula errors in the electronic QA/QC file. The fill color and comment features provide simple and efficient means of highlighting the location of errors in the electronic file. An alternative to creating an electronic file is to print a hard copy of the electronic sheet and check the paper copy against the field datasheet, noting any errors on the hard copy.
4. Check the entered plant species codes against the most recent version of the *Oncor* plant list found in the DRW. Changes to plants names and codes frequently occurs and duplicate codes can exist for different species, therefore the entered species codes need to be checked to ensure that the codes entered are indicative of the species that the user intended.
    a. To facilitate this check, the row above the codes in the 2_HerbCover_DET sheet are automatically populated with the scientific name associated with the species code. Check that all the codes result in the expected species name.
    b. If any of the values return as “#N/A” then the plant code is either new or has changed. Correct any codes that are incorrect on the datasheet.
    c. If it appears that there is a new plant code check the USDA PLANTS Database (http://plants.usda.gov/java/) to confirm it is the most up to date classification, and add the species to the 1_SpeciesCode_Lookup using the Standard Value button on the tab in the DRW.
5. Check the elevation data to ensure they were entered in the correct rows. If a lookup table was used, only 10% of the data needs to be checked. If the data were hand entered, 100% of the data needs to be checked.

Those implementing QA/QC using the paper copy method need to have an alternative project tracking system for ensuring data quality. Regardless of one's preferred method, the **2_HerbCover_DET** worksheet should be populated with —QC-checked data for the subsequent data reduction steps.

7.4 Uploading Data to *Oncor*

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

8.0 Tree Plots and Site Summary[10]

This DRP describes an approach for preparing tree plot data and subsequent site summary calculations for uploading to the *Oncor* database. It corresponds to the tree field data collection protocol described by Roegner et al. (2009), which involves collecting data on tree species and diameter in 10-m−diameter circular plots. This procedure refers to a corresponding DRW—DRW_Tree_v0.1.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.

8.1 DRW Structure

The tab structure of DRW_Tree_v0.X.xlsm is described below and each of the associated worksheets is further described in the Read Me tab of the DRW.

Read Me
Data Dictionary
0_Metadata_Template
0_Metedata_Example
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_TreePlot_Template
1_TreePlot_Example
1_SpeciesCode_Lookup
2_SiteSummary_Template
2_SiteSummary_Example
3_TreePlot_DET
4_SiteSummary_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, 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 template and example data tabs. The data generator copies raw data into the template. The example is for the data generator to review for guidance. The examples also contain formulas, where applicable, to calculate mean distance between ground and the plane at the top of the stakes, and the annual sediment accretion rate. These may be copied into the templates as needed.

[10] Prepared by Heida Diefenderfer (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.

8.2 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_Template. To reference the example data as needed, see 0_Metadata_Example.
2. Using the 1_SpeciesCode_Lookup tab, identify the six-letter species codes for all species included in your data set. To customize your datasheet, enter these codes into row 1 of 1_TreePlot_Template, beginning with column "I" to the right of the Notes column.
3. Transcribe tree plot data from field datasheets into the tab 1_TreePlot_Template. Enter elevations for any quadrat for which they are available. To reference the example data as needed, see 1_TreePlot_Example.
4. Quality-control check the tree plot data.
5. Calculate the site summary data in the tab 2_SiteSummary_Template. To reference the example data as needed, see 2_SiteSummary_Example. The example also contains the formulas for calculated fields.
6. Quality-control check the calculations of site summary data.
7. To build the DET, copy the tree plot data over the example data in the tab 3_TreePlot_DET. Paste special, values only.
8. Copy the mean distance data over the example data in the tab 4_SiteSummary_DET. Paste special, values only.
9. Quality-control check the copy-paste operations.

8.3 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.

8.4 Uploading Data to *Oncor*

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

9.0 Fish Catch and Density[11]

The information contained in this section outlines an approach for preparing fish catch data and subsequent density calculations for uploading to the *Oncor* database. The procedure outlined here is focused on data collected using beach seines in shallow water habitats. While there are similarities in the data types and accompanying field codes between various fishing techniques, data collected using other methods (i.e., trap net, electrofishing) would require distinct DRWs and DETs. The procedure described below addresses methods for QC measures, calculating densities, and final data formatting for entry into the *Oncor* database.

9.1 DRW Structure

This procedure refers to DRW_FishDensity_v0.X.xlsm that has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_FishCount_Template
2_FishCount_Example
3_FishDensity_Example
4_FishDensityAvg_DET

> The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

Furthermore, the **blue** tabs in the fish catch and density DRW provide 1) a template that can be used to enter or paste in raw data and 2) example worksheets that include data used to convert fish count (number of fish caught per haul) to mean density (number of fish caught per square meter). The fish catch and density DRW differs slightly from other data categories and their corresponding DRWs because the final _DET which is uploaded into *Oncor* includes calculated (i.e., mean density) values.

9.2 Data Reduction Steps

The following steps and accompanying DRW outline the process for reducing fish catch data collected at each haul to density (#fish/m^2) estimates from a given site. The steps describe an approach for managing data using a series of Excel worksheets and workbooks. Data managed in other software programs will need to export data according to the structure of the DET to ensure successful uploading and inclusion of data in *Oncor*.

---

[11] Prepared by Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

Step 1: Enter fish catch data into the DRW.

Data on counts of fish by species that are collected in the field should be entered into the 1_FishCount_Template worksheet, located in the DRW_FishDensity_v0.1.xlsm DRW. The template includes columns associated with typical fish data collection procedures, not all of which are required for uploading data into the *Oncor* database. Required fields are denoted in the data dictionary of the fish density DRW.

In addition to entering data from field worksheets, it may be necessary to incorporate data from calculations into the DET at this phase. Examples include area swept with the beach seine, and if subsampling was implemented in the field for any haul, the catch data should be adjusted accordingly. Methods for calculating the area swept and subsampled catch are described by Roegner et al (2009) and Sather et al. (2011). The formulas used for calculations should be retained in the worksheet because they will become part of the QC check described in Step 2.

Step 2: Finalize fish catch.

Upon completion of the QA/QC process, populate the 2_FishCount_Example worksheet with the QC-checked data (see QC procedure below) using simple copy-and-paste options.

Step 3: Calculate fish density.

Once the final catch data have been pasted into the 2_FishCount_Example worksheet, click the Resize Tables button, located in the upper left corner of the DET. This feature calculates the density per a haul in the 3_FishDensity_Example worksheet and mean density per site in the 4_FishDensityAvg_DET worksheet by recognizing the new data that have been input into the 2_FishCount_Example worksheet.

9.3 Quality Control

To ensure there were no errors when transcribing field data to electronic format, QC measures should be implemented following initial data entry. Using the field datasheets, a 100% check of data entry and a 10% check of formulas are recommended. This check must be performed by an individual who did not perform the initial data-entry tasks described in Step 1 above.

The process for implementing data QA should occur in a manner that is best suited to the project needs. We recommend saving the 1_FishCount_Template as a new Excel file and renaming it using “QAQC” in the filename designating the new file as the QA/QC file. In the first blank cell at the bottom of the worksheet, insert the name of the person performing the QA check and the date the check was completed. Highlight all transcription errors and formula errors in the electronic QA/QC file. The fill color and comment features provide simple and efficient means of highlighting the location of errors in the electronic file. An alternative to creating a workbook is to print a hard copy of the worksheet and check the paper copy against the field datasheet.

It is possible for a data generator to maintain all QA/QC files as separate worksheets in the fish catch DRW. However, we find creating separate QA/QC files provides an efficient way to track the progression of the data from raw to final format. In addition, those implementing QA/QC using the paper copy method likely have an alternative project tracking system for ensuring data quality. Regardless of one’s preferred method, the 2_FishCount worksheet should be populated with QC-checked data for the subsequent data reduction steps described below. For additional details on QA/QC, see Section 2.3 and Appendix C.

9.4 Uploading Data to *Oncor*

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

10.0 Fish Size[12]

The information contained in this section describes an approach for preparing fish size data for uploading to the *Oncor* database. The procedure outline here is focused on data collected using beach seines in shallow water habitats (e.g., Sather et al. 2011). Data collected using other methods (e.g., trap net, electrofishing) may require slightly different information to be included in the DRPs and DETs. Not all of the field codes in the fish size DET will be necessary. For example, some projects may not collect diet samples or use passive integrated transponder tags. The dictionary in the fish side DET will elucidate the purpose of field codes and identify which are necessary for uploading data to *Oncor*. The procedure described below addresses methods for QC measures and final data formatting for entry into the *Oncor* database.

10.1 DRW Structure

This procedure refers to DRW_FishSize_v0.X.xlsm, which has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_FishSize_Template
2_FishSize_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

10.2 Data Reduction Steps

Step 1: Electronic transcription from field datasheets.

Use the 1_FishSize_Template from the DRW_FishSize_v0.1.xlsm to record size data collected from the field.

Step 2: Finalize fish size data.

The individual who completed Step 1 reviews the QA/QC file and examines errors. Once errors have been rectified a final size file is generated. This corrected file becomes the ‘FishSize_DET’, which is uploaded to *Oncor*. Example data are portrayed in the 2_FishSize_exampleDET worksheet.

[12] Prepared by Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

10.3 Quality Control

To ensure there were no errors when transcribing field data to electronic format, QC measures should be implemented following initial data entry. Using the field datasheets, a 100% check of data entry is recommended. This check must be performed by an individual who did not perform the initial data-entry tasks described in Step 1 above.

The process for implementing data QA should occur in a manner that is best suited to the project needs. We recommend saving the 1_FishSize_Template as a new Excel file and renaming it using ”QAQC” in the filename designating the new file as the QA/QC file. Below the last row of data in the worksheet, insert the name of the person performing the QA check and the date the check was completed. Highlight all transcription errors and formula errors in the electronic QA/QC file. The fill color and comment features provide simple and efficient means of highlighting the location of errors in the electronic file. An alternative to highlighting cells would be to create a QA/QC worksheet and/or separate file to document identified errors.

For additional details on QA/QC, see Section 2.3 and Appendix C.

10.4 Uploading Data to *Oncor*

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

11.0 Fish Diet[13]

This section describes data-entry processes for uploading fish diet data to the *Oncor* database. Diet data are typically collected via lavage techniques during field sampling events and samples of individual gut contents are transported to the laboratory for analysis. Field and laboratory methods for the fish diet data category described by Storch and Sather (2011).

11.1 DRW Structure

This procedure refers to DRW_FishDiet_v0.X.xlsm, which has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_FieldCollectedData_Template
2_DietTaxaID_Template
3_MassMeasurements_Template
4_FieldCollectedData_DET
5_DietTaxaID_DET
6_MassMeasurements_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

11.2 Data Reduction Steps

Step 1. Enter data from field datasheets into the DRW.

Field datasheets associated with collecting fish diets are used to record basic information about the site as well as information about the fish sampled. Upon returning to the laboratory, the field datasheet is transferred into the electronic format found in the 1_FieldCollectedData_Template. It is important to note that every diet sample must be associated with a unique diet sample ID. These names could be generated automatically in Excel using macros; however, they can also be generated manually in the field by combining date, site, and species information. For example, a Chinook salmon diet sample collected during May 2010 at site B can be coded as 0510BCH-03. The numeric value at the end of the code is used to designate the particular diet sample number collected from a specific date/site/species combination. In the preceding example, '-03' indicates the sample was the third Chinook diet collected from site B during May 2010.

[13] Prepared by Nichole Sather (PNNL) and Adam Storch (Oregon Department of Fish and Wildlife). For more information, please contact Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

Step 2. Transfer the diet samples to the laboratory and commence processing.

The analysis of fish diet samples conducted in the laboratory will generate two additional worksheets that reflect the fish diet: DETs – 2_DietTaxaID and 3_DietMassMeasurements. In the example DETs provided, prey items are identified to the lowest taxonomic category possible for each diet sample. Within a given diet sample, individual prey items are segregated and placed into individual vials (labeled with alpha and/or numeric codes). In the 5_DietTaxaID_DETexample worksheet, each row represents an individual taxon within a particular diet sample. For example, diet sample 0510BCH-03 contains two prey items; therefore, this sample ID appears in two rows of the worksheet.

Step 3. Weigh diet samples.

After taxonomic identification of diet contents, samples are weighed for further analyses; for example to calculate an index of relative importance. The 6_DietMassMeasurements_DET worksheet is structured similarly to the 5_DietTaxaID_DETexample worksheet in that each row represents an individual taxon within a particular diet sample.

11.3 Quality Control

Quality control actions are included in the description of the data reduction steps above. Please note that QA/QC procedures should be implemented on the field datasheet that was transferred into the 1_FieldCollectedData_Template to find transcription errors.

For additional details on QA/QC, see Section 2.3 and Appendix C.

11.4 Uploading Data to *Oncor*

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

12.0 Fish Prey[14]

The data-entry processes for uploading fish prey data into the *Oncor* database are described in this section. Prey data are typically collected by several different methods (see Storch and Sather 2011) during field sampling events and samples are transported to the laboratory for analysis.

12.1 DRW Structure

This procedure refers to DRW_FishPrey_v0.X.xlsm, which has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_FieldCollectedData_Template
2_PreyTaxaID_Template
3_FieldCollectedData_DET
4_PreyTaxaID_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

12.2 Data Reduction Steps

Step 1. Transfer data from field datasheets to the DRW.

Field datasheets associated with collecting fish prey are used to record basic information about the site as well as information about the type of sample collected (e.g., benthic, drift, fallout). Upon returning to the laboratory, the field datasheet is transferred into the electronic format found in the 1_FieldCollectedData_Template. It is important to note that each sample must be associated with a trap ID. The trap ID is linked to dimensions of sampling gear (recorded in the 0_Metadata tab), which are ultimately used to calculate the densities of prey items.

Step 2. Conduct quality control. See Section 12.3.

Step 3. Populate the DRW with QC-checked data.

The 1_FieldCollectedData_DET worksheet is populated with QC-checked data.

---

[14] Prepared by Nichole Sather (PNNL) and Adam Storch (Oregon Department of Fish and Wildlife). For more information, please contact Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

Step 4. Generate the DET.

The laboratory processing of fish prey samples uses the 2_PreyTaxaID_Template to generate the final DET in this workbook. In the example data provided, prey items were identified to lowest taxonomic category possible for each sample and enumerated.

Step 5. Finalize the DRW.

Create the prey taxa DET using the 4_PreyTaxaID_DET worksheet.

12.3 Quality Control

Quality control procedures to cross-check for accurate transcription should be implemented on the field datasheet that was transferred to the 1_FieldCollectedData_Template. Similarly, checks for accurate translation for the final DET should be undertaken.

For additional details on QA/QC, see Section 2.3 and Appendix C.

12.4 Uploading Data to *Oncor*

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

13.0 Chinook Salmon Genetic Stock Identification[15]

Tissue samples from juvenile salmon sampled with beach seines or other gear are collected in the field for subsequent genetic stock identification of Chinook salmon in a laboratory. Tissue samples for genetic stock identification are typically collected via fin clipping techniques during field sampling events. Tissue samples are transported to the laboratory for analysis. Field and laboratory methods for the data category for Chinook salmon genetic stock identification are described by Sather et al. (2011). This DRP covers treatment of the samples from their collection in the field to transfer to the laboratory for processing and receipt of the laboratory results followed by incorporation into the DRW.

13.1 DRW Structure

This procedure refers to DRW_FishGenetics_v0.X.xlsm, which has the following tab color and structure:

Read Me
Data Dictionary
0_Metadata
0_SpatialData_Coordinates
0_SpatialData_Shapefile
1_FishGenetics_Template
2_FishGenetics_DET

The **green** tabs are information only, the **blue** tabs provide examples from the data processing steps, and the **red** tabs provide the format for the DET that will be used for uploading data to *Oncor*.

13.2 Data Reduction Steps

Step 1. Prepare and deliver fin clip samples to the genetic laboratory.

Details pertaining to the delivery of samples to the genetics laboratory should be thoroughly coordinated prior to shipment. Approaches for sending samples and receiving analyzed data may differ according to project needs and/or laboratory standard operating procedures and protocols. Regardless, all samples should be accompanied by a Chain-of-Custody (CoC) form. At minimum, the CoC should include the project name and number, project manager contact information, field-generated sample IDs, collection date, and time of each sample. The name and contact information for the sample custodian should also be included on the CoC forms. The original CoC accompanies the sample shipment, and a copy should be retained for the project files. In addition to the CoC form, it is recommended that an electronic datasheet containing additional details about the samples be sent to the geneticist. The 1_FishGenetics_Template in the DRW_FishGenetics_v0.1.xlsm can be used as a template. The data pertaining to the various field codes can be derived from the FishSize DET in the

[15] Prepared by Nichole Sather (360-681-3688; Nichole.sather@pnnl.gov).

DRW_FishSize_v0.1.xlsm. This eliminates the need for entering new data, which would necessitate a new QA/QC process.

Step 2. Enter the results obtained from genetics processing in the laboratory.

The genetics lab populates the last five rows of the worksheet. The electronic results of the genetic stock assignments should be reviewed for completeness prior to uploading the FishGenetics_DET to *Oncor*.

13.3 Quality Control

Quality control actions were included in the preceding steps. For additional details on QA/QC, see Section 2.3 and Appendix C.

13.4 Uploading Data to *Oncor*

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

14.0 References

Borde A, S Zimmerman, V Cullinan, J Sagar, H Diefenderfer, K Buenau, R Thom, C Corbett, and R Kaufmann. 2012. *Lower Columbia River and Estuary Ecosystem Restoration Program Reference Site Study: 2011 Restoration Analysis*. PNNL-21433. Pacific Northwest National Laboratory, Richland, Washington.

BPA/Corps (Bonneville Power Administration and US Army Corps of Engineers). 2013. *Columbia Estuary Ecosystem Restoration Program: 2013 Strategy Report*. Final report, Bonneville Power Administration and U.S. Army Corps of Engineers, Portland, Oregon.

Coleman AM, GE Johnson, AB Borde, HL Diefenderfer, NK Sather, TE Seiple, and JA Serkowski. 2013. *Synthesis and Evaluation of Research, Monitoring and Evaluation and Restoration Effectiveness Data in the Lower Columbia River and Estuary, 2012*. PNNL-22405, draft annual report submitted to the U.S. Army Corps of Engineers, Portland District, Portland, Oregon, by Pacific Northwest National Laboratory, Richland, Washington.

Roegner GC, HL Diefenderfer, AB Borde, RM Thom, EM Dawley, AH Whiting, SA Zimmerman, and GE Johnson. 2009. *Protocols for Monitoring Habitat Restoration Projects in the Lower Columbia River and Estuary*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration Technical Memorandum NMFS-NWFSC-97, Northwest Fisheries Science Center, Seattle, Washington.

Sather NK, DJ Teel, AJ Storch, GE Johnson, ES Van Dyke, EM Dawley, DL Kuligowski, TA Jones, A Bryson, and KL Sobocinski. 2011. "Juvenile Salmon and Fish Community Characteristics," pp. 2.1–2.38, in *Ecology of Juvenile Salmon in Shallow Tidal Freshwater Habitats of the Lower Columbia River, 2007-2010*, Johnson et al. 2011. PNNL-20083, Pacific Northwest National Laboratory, Richland, Washington.

Storch AJ and NK Sather. 2011. "Feeding Ecology." In *Ecology of Juvenile Salmon in Shallow Tidal Freshwater Habitats of the Lower Columbia River, 2007-2010*, Johnson et al. 2011, pp. 4.1–4.20, PNNL-20083, Pacific Northwest National Laboratory, Richland, Washington.

Thom RT, NK Sather, D Bottom, and GC Roegner. 2013. *Columbia Estuary Ecosystem Restoration Program: 2012 Synthesis Memorandum*. Final report submitted to the U.S. Army Corps of Engineers, Portland, Oregon, by the Pacific Northwest National Laboratory, Richland, Washington, and the National Marine Fisheries Service, Seattle, Washington.

Washington-Action Agencies. 2009. Estuary Habitat MOA. Memorandum of Agreement on Columbia River Estuary Habitat Actions Between the State of Washington, the Bonneville Power Administration, The U.S. Army Corps of Engineers, and the U.S. Bureau of Reclamation. Available at www.salmonrecovery.gov/.

Xu J, EP Myers, and SA White. 2010. *Vdatum for the Coastal Waters of North/Central California, Oregon and Western Washington: Tidal Datums and Sea Surface Topography*. NOAA Technical Memorandum NOS CS 22. National Oceanic and Atmospheric Administration, Silver Springs, Maryland.

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.

Appendix B

Detailed Workflow
from Data Collection to Successful *Oncor* Entry

Appendix B

Detailed Workflow from Data Collection to Successful *Oncor* Entry

The steps of the detailed workflow are listed below and an illustration of the workflow is displayed in Figure B.1.

Step 1. The data generator collects data according to field procedure. The critical aspect of this step with respect to uploading data to *Oncor* is that all required data for the relevant data category are obtained during field collection (e.g., sampling location, date).

Step 2. The data generator processes field data according to the data reduction procedure (DRP) to produce a data exchange template (DET). Included in this step are the quality control measures which are the responsibility of the data generator. Every DRP, at some stage, will require that the data generator download an appropriate DET from the *Oncor* web site. The DET permits the data generator to add new aliases to existing standard values or provisionally add new standard values. The DRP provides detailed instructions for how to populate the DET.

Step 3. Where standard values are required in the DET, lookup lists will be present containing alias names associated with the data generator. Any new standard values used in the data collection (e.g., new people, new instruments, etc.) that are not present in *Oncor* should be recorded on the Metadata tab at this stage. Previously entered values will appear in dropdown menus within the Metadata tab. If the values needed are not available the data generator uses the Standard Values button. Similarly, if a desired alias for an existing standard value is not present, the data generator can add the new alias using the Standard Values button of the Metadata tab of the DET. The addition is provisional because new standard values must be approved by the data custodian (see Step 4 below).

Step 4. The data generator submits the completed DET for uploading to the *Oncor* web site. The data generator uploads the completed DET to the *Oncor* web site and receives one of the following responses:

a. If the DET is valid, a preload file is created that shows how the data will appear in *Oncor* after it is loaded and receives either an acceptance in the form of a preload file (to be described later) or a rejection in the form of an error file (to be described later). Problem resolution may involve the data custodian.

b. If the DET has errors, an error file is created that shows where errors occurred. At this point, the data generator must correct problems and resubmit the DET until it is accepted. Problem resolution may involve the data custodian, especially when a new standard value is involved.

Step 5. The data generator approves the preload file, submits the load, and receives a load confirmation. Once satisfied with the preload file, the data generator approves the loading of the DET and the loader performs the task automatically. After successful completion of the load, the loader produces a load confirmation report, which may include some statistics that summarize the load. If a failure occurs at this point, the data custodian must be contacted for resolution of the problem.

Step 6. The data generator may query the database for selected new records to verify their presence. It is suggested that data generators perform some spot checks on their newly uploaded data to confirm its successful entry. This is done through the use of standard queries available in the *Oncor* interface. The data generator or user may also generate their own queries.

Step 7. The data custodian checks newly entered data to ensure compliance with standards. The data custodian has a number of tools at his or her disposal for verifying and validating data after uploading has taken place. These may include anything from manual spot checks to statistical analyses.

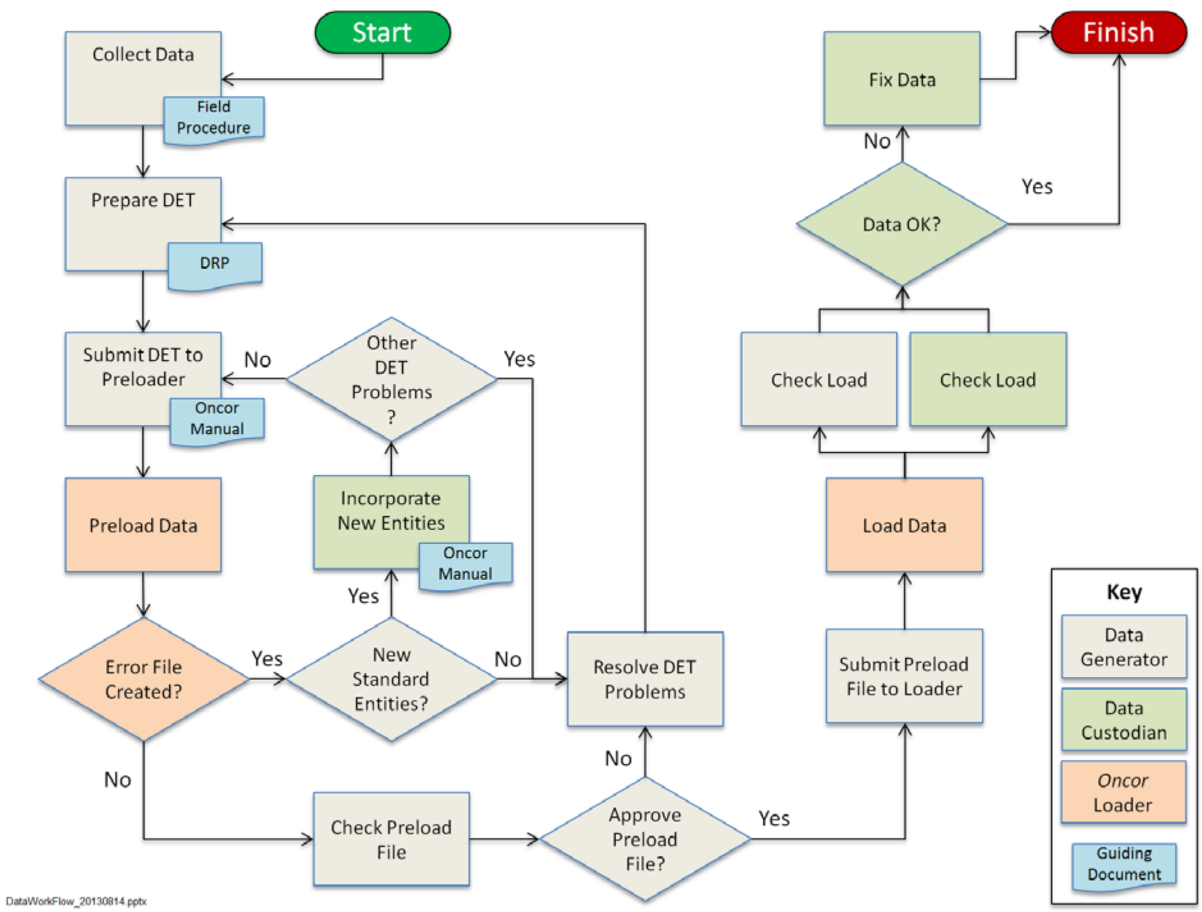

Figure B.1. Detail of data flow for *Oncor*.

Appendix C

Data Quality and Format: Standards and Enforcement

Appendix C

Data Quality and Format: Standards and Enforcement

Oncor will store data generated by multiple people from multiple agencies, each likely with its own set of data standards. To ensure proper integration of all data in *Oncor*, the maintenance of a single set of data standards is critical. Data quality will be ensured through the use of data and formatting standards.

C.1 Data Standards

Data standards help ensure that data from disparate sources have consistent meaning and can be properly compared. For *Oncor*, data standards are defined as the set of rules that applies to the contents of fields and records stored in a database. Standards include field type specifications (e.g., integer, date/time, Boolean, etc.), content constraints (e.g., times are local, etc.), and consistency in nomenclature.

Oncor implements data standards in several ways. At the lowest level, certain standards are controlled by the definitions of the table structures in the database. These definitions physically prevent certain improper data from entering the database. Every field in the table structure has a specific data type to which new values must adhere. For example, numeric measurement values are always placed in a field with a specific format—double precision floating point number—even if reported as an integer. Dates use must be in date/time format. Thus, string values in a numeric field or invalid dates in a date field would be automatically rejected by the database. Database tables may also have fields defined as *required*, so records cannot be saved unless values are present for these fields. In summary, incomplete records or records that do not conform to the required data type would be rejected from loading into such tables.

A higher level of data standardization concerns the prevention of loading values that are "legal" so far as the data structure is concerned, but undesirable for one reason or another. Standard value naming conventions fall under this category. Inconsistent naming of identical entities can obfuscate the data and significantly reduce its utility. When different data owners refer to the same standard value, e.g., a location, by different names, retrieving all data for that value becomes very difficult. To ensure consistent nomenclature, *Oncor* uses a coded-field scheme for certain standard values, such as location, instrument name, sampling method, person name, and others. An integer value, rather than a text string, uniquely defines these standardized entities. This allows users to maintain multiple string names (aliases) for identical entities. For example, one user may refer to a location as "BBM," while another may use the term "Baker Bay Marsh" and a third just "Baker Bay." *Oncor* assigns a single Location_ID integer value for this site, but allows each user to refer to it, both in their load files and in their data queries, using his or her own name or "alias."

In addition to maintaining naming conventions, the *Oncor* loader performs various checks on incoming data to ensure its validity before being entered into the database. These checks include verifying that required fields are present. For example, most measurement values are meaningless if they lack a measurement date associated with them. Values are also checked for "reasonableness," such as whether measurements dates are in the past or water temperatures are below boiling. In addition, the data exchange templates (DETs) provide a structure that enforces certain standards by supplying the fields and record types required for each data category. The DRPs and data exchange template (DETs) document the overall requirements for data loaded into the database. These rules include standards that cannot be enforced through automated checks, but that must be adhered to in order to maintain consistency in the database. For example, in specifying the time standard used for measurement dates or the units used for specific measurements, software can detect if a measurement date erroneously occurs in the future, but it cannot determine if the date uses local daylight savings time, as it should. Data generators are responsible for providing their data in compliance with these rules.

C.2 Standards Enforcement

Failure to comply with many of the *Oncor* data standards will result in rejection of records by the data loader, leading to data-entry delays and possible frustration for the data generator. The purpose of this report is to provide all information needed to avoid rejection and ensure the consistency of data throughout the database. *Oncor* will use the following mechanisms to enforce data standards:

1. Data exchange template. The DET is the primary data-standards enforcement mechanism. DETs allow users to use their own data standards (aliases) to the greatest degree possible. The DET enforces standards by specifying the fields required for each subject area, defining how they may be filled, and automatically checking selected fields using simple validity tests before being accepted for loading.
2. *Oncor* data standards. Data generators will be responsible for adhering to the *Oncor* Data Standards. While *Oncor* can facilitate many aspects of data standardization, it is impractical to guarantee accurate data entry without the cooperation of the data generator.
3. Data custodian. An *Oncor* administrator will maintain a position of data custodian. The data custodian is responsible for enforcing data standards, ensuring that data are loaded into the database correctly, and assisting data generators with loading issues.

C.3 Standardized Value Management

Standard values are identified and defined in this appendix and on the *Oncor* website, and managed by the data custodian. The *Oncor* coordination process will manage the definitions of standard values and may help resolve issues relating to values. Current standard values are as follows:

- **Agency.** Organizations associated with lower Columbia River and estuary (LCRE) data (e.g., Lower Columbia River and Estuary Partnership, Columbia Land Trust, etc.).
- **Document.** Reference document citation.
- **Instrument.** Equipment used to collect data (e.g., data logger, fish net, etc.).
- **Location.** Geographical point, line, or polygon.
- **Method.** Data collection method.
- **Person.** Full name and affiliation of people associated with LCRE projects.
- **Program.** Highest organizational level of a data collection effort; missions, research questions, and protocols are defined at this level (e.g., LCRE Ecosystem Restoration Program, Cumulative Effects

Study). Note: the distinction between program and project needs to be better defined, or, perhaps, the two categories should be combined.

- **Project.** Specific research activity that supports one or more of the research questions of the Program (e.g., Reference Site Study). Note: the distinction between program and project needs to be better defined, or, perhaps, the two categories should be combined.
- **Species.** Scientific name for a plant or animal.

Standard-value records are stored in *Oncor* as "groups." In the *Oncor* data model, a group record provides the hierarchical data structure that relates associated data together. For example, storage of a **Person** consists of the group name **Person_ID** having a unique long-integer code for each distinct person in the database. Associated with every **Person_ID** are the attributes **Person_LastName**, **Person_FirstName**, **Person_MI**, **Person_Agency**, and possibly others.

C.4 Alias Names

An important component of standard-value management is the concept of alias names. It is recognized that users will have preferred names for many of the standard values. So that users are not forced to adopt standard names, the *Oncor* data model allows access to standard-value records by multiple reference names. This is accomplished using an alias table, an example of which is shown below.

Table C.1. Example Alias Table

Group_Name	Code_Value	Alias_Group	Alias_Name
Instrument_ID	333	*Oncor*	HOBO #2
Instrument_ID	444	*Oncor*	HOBO #6
Instrument_ID	555	*Oncor*	HOBO #1
Instrument_ID	666	*Oncor*	HOBO #12
Instrument_ID	555	PNNL Default	HOBO #1
Instrument_ID	555	PNNL Short	H1
Instrument_ID	666	PNNL Default	H2
Instrument_ID	555	USGS	Old Logger
Instrument_ID	666	USGS	New Logger
Instrument_ID	444	USACE	Data Logger
Instrument_ID	333	USACE	H2
Instrument_ID	555	USACE	H1
Location_ID	10001	*Oncor*	Baker Bay Marsh
Location_ID	10001	PNNL Short	BBM
Location_ID	10001	USGS	Baker Bay
Location_ID	10001	USACE	Baker Bay Marsh

The columns in the alias table contain the following information:

- **Group_Name**: group name of the standard value (e.g., Person_ID, Instrument_ID, etc.).
- **Code_Value**: long-integer value of specific instance of the group.

- **Alias_Group**: a group of related aliases belonging to a specific data user that may be requested for output in a query. Every standard value has at least one alias that is in the **Alias_Group** called "*Oncor*", which is the default name used in query results. Associations between Alias_Groups and users
- **Alias_Name**: name associated with **Code_Value** for given **Alias_Group**.

Note that a single user may refer to a specific standard value by multiple names, which are distinguished by the Alias_Group value. For example, a single user at Pacific Northwest National Laboratory (PNNL) may be associated with both the "PNNL Default" and "PNNL Short" Alias_Groups and have the option of referring to instrument 555 by either "HOBO #1" or "H1", respectively. Different users may call different entities by the same name. For example, users associated with "PNNL Default" and "USACE" both have instruments called "H2", but they refer to different actual units: 666 and 333, respectively. *Oncor* will always maintain one default name under the Alias_Group "*Oncor*" for each standard value. This will be the preferred name and will appear in query results unless a user specifically requests a different Alias_Group.

Upon entry into the *Oncor* community, every data generator will provide information about his or her user-specific standard-value information to the data custodian. This will be done via the Metadatasheet of the DET using a Standard Values button. The data custodian will verify that the necessary information for each type of standard value is complete and will resolve duplicates, making sure that no two standard values point to the same object. The data custodian will be responsible for building and maintaining the alias table. The alias table will not only translate user-specific nomenclature at load time, but will also provide lookup lists for appropriate fields in the DET that will force data generators to enter only valid names. Data generators may continue providing the data custodian with new aliases and standard values via the same method.

Appendix D

Analysis Questions

Appendix D

Analysis Questions

This appendix demonstrates example outcomes from data analysis, after data are uploaded to *Oncor*, 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).

D.1 Fish Analysis Questions

The following questions and subsequent figures provide examples of queries that managers and data generators may wish to perform with *Oncor*. The examples below are from data collected during 2009 near the Sandy River delta study area (river kilometer [rkm] 188−205) and the Lower River Reach (LRR) study area in the lower Columbia River and estuary. Fish were sampled monthly from nine sites using beach seines. Water temperature data were collected in conjunction with beach seine activities. For additional information about sampling methods, see the report by Sather et al. (2011).

What was the seasonal, site-specific water temperature in a given study area?

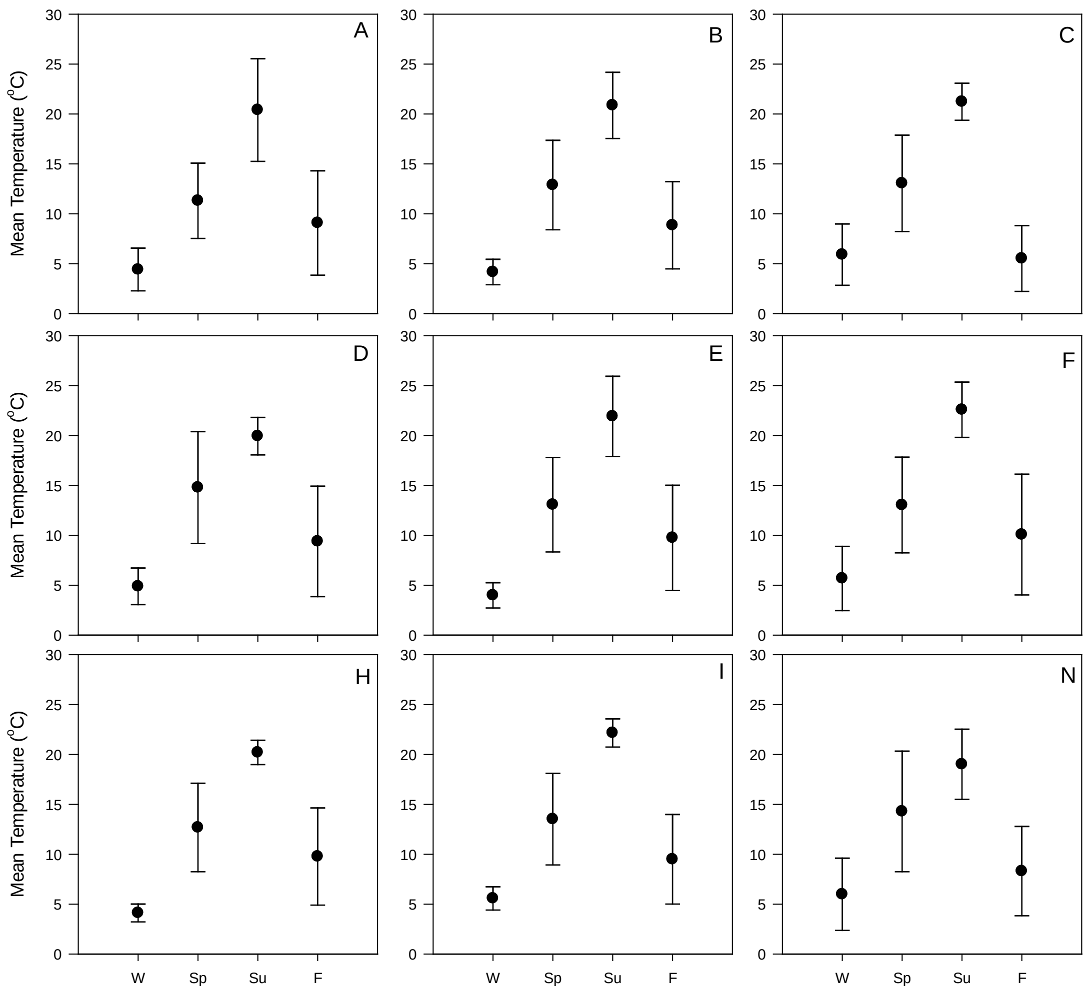

Figure D.1. Mean seasonal temperature collected from nine sites in the Sandy River delta study area. Site names correspond to the letter code in the upper right corner of each panel. Data were collected during 2009 as point measurements (from a YSI instrument) and taken in conjunction with fish seine hauls.

What were the site-specific mean densities of selected fish species over a given year, month, and/or season?

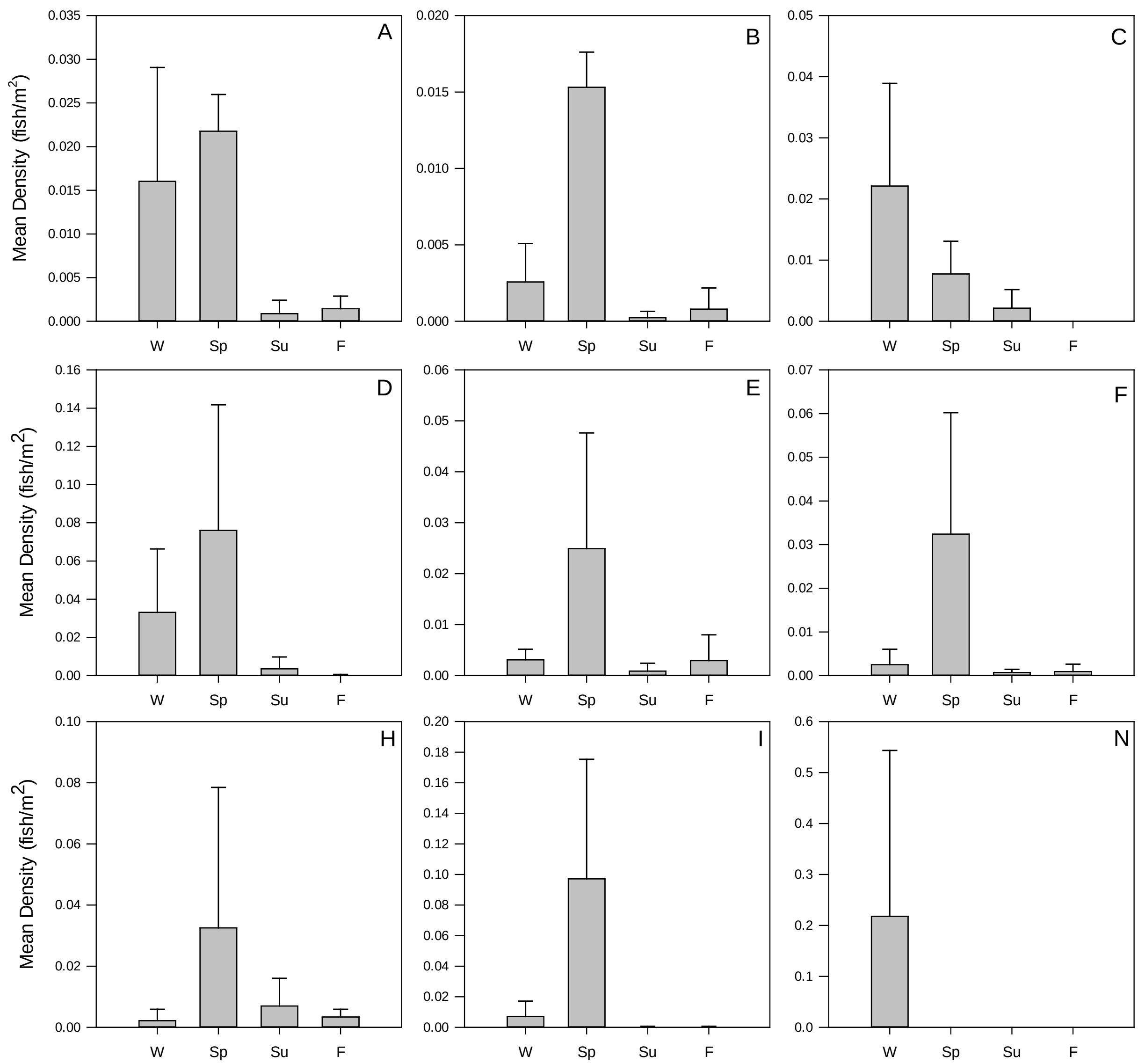

Figure D.2. Mean seasonal densities of unmarked Chinook salmon estimated from nine sites in the Sandy River delta study area during 2009. Site names correspond to the letter code in the upper right corner of each panel. Error bars are one standard deviation.

What were the seasonal mean densities for salmon, non-native, and other native (excluding salmon) taxa in a given study area over a given time period? What is the standard deviation?

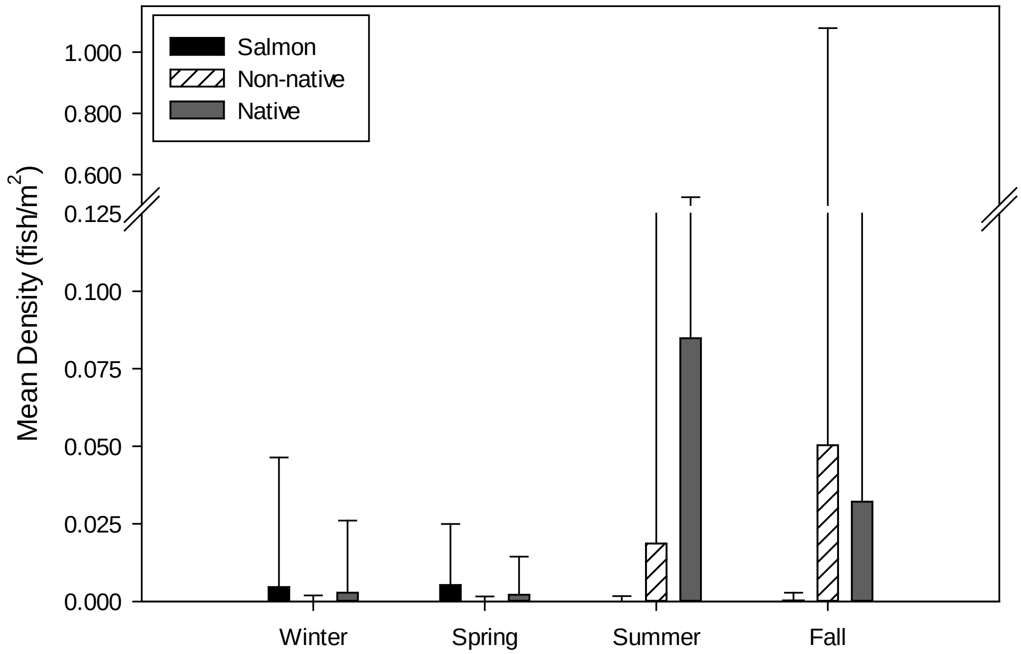

Figure D.3. Mean seasonal densities for different groups of fish sampled at the Sandy River delta study area during 2009. Fish groups include native and non-native status, and salmon have been extracted from the native group to be represented as their own category. Error bars are one standard deviation.

What are the seasonal mean densities for selected salmon species in a given study area?

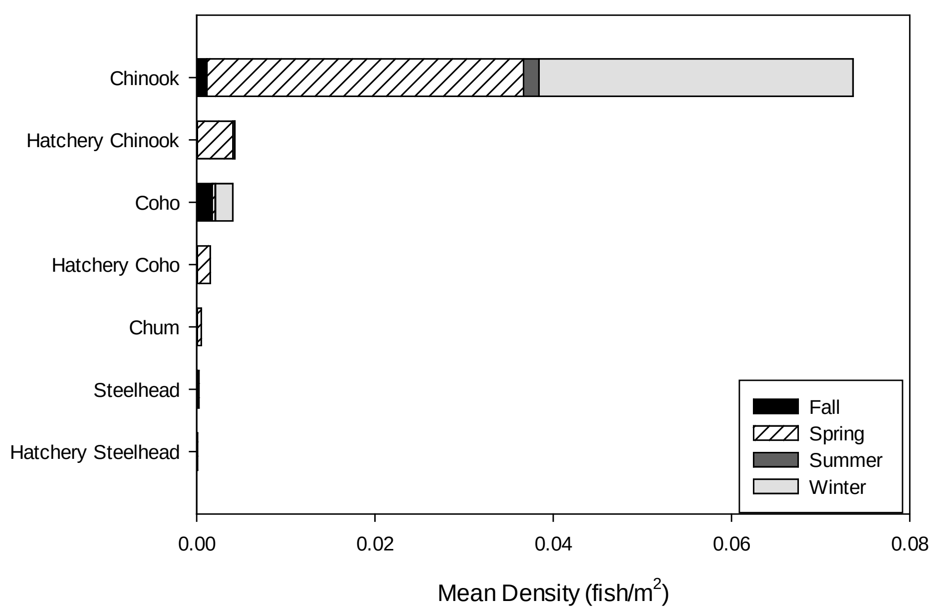

Figure D.4. Mean densities of individual salmon species seasons. Data were collected from the Sandy River delta study area during 2009.

What are the seasonal densities of a given species within a particular habitat type?

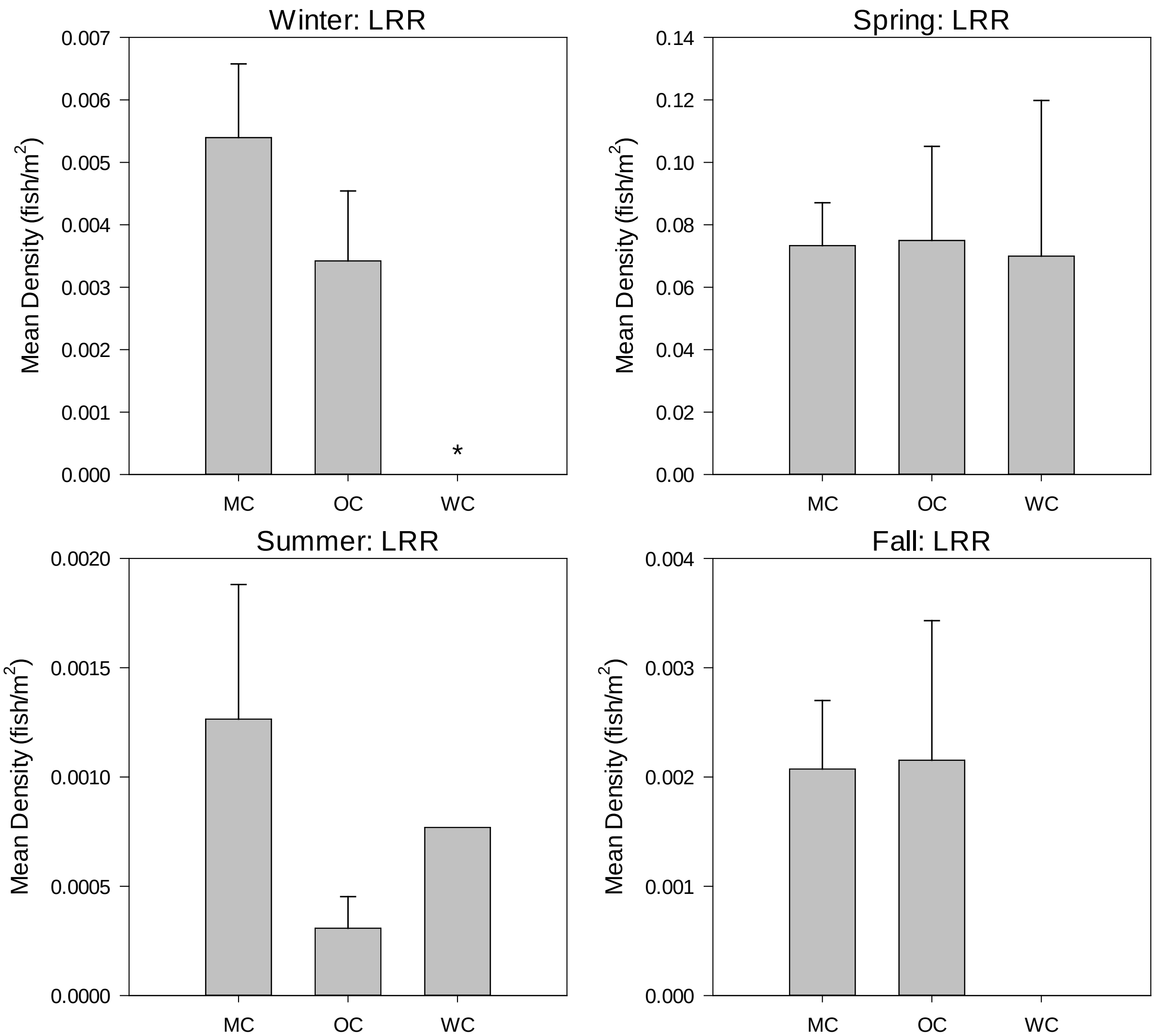

Figure D.5. Mean density of unmarked Chinook salmon in different habitat types: MC (main channel), OC (off channel), WC (wetland channel). Fish were sampled in the Lower River Reach (LRR) study area (rkm 110-141) during 2009. The asterisk denotes the wetland stratum was not sampled.

D.2 Habitat Analysis Questions

What are the average percent cover and minimum and maximum elevations at which species *s* occurs at site *n*?

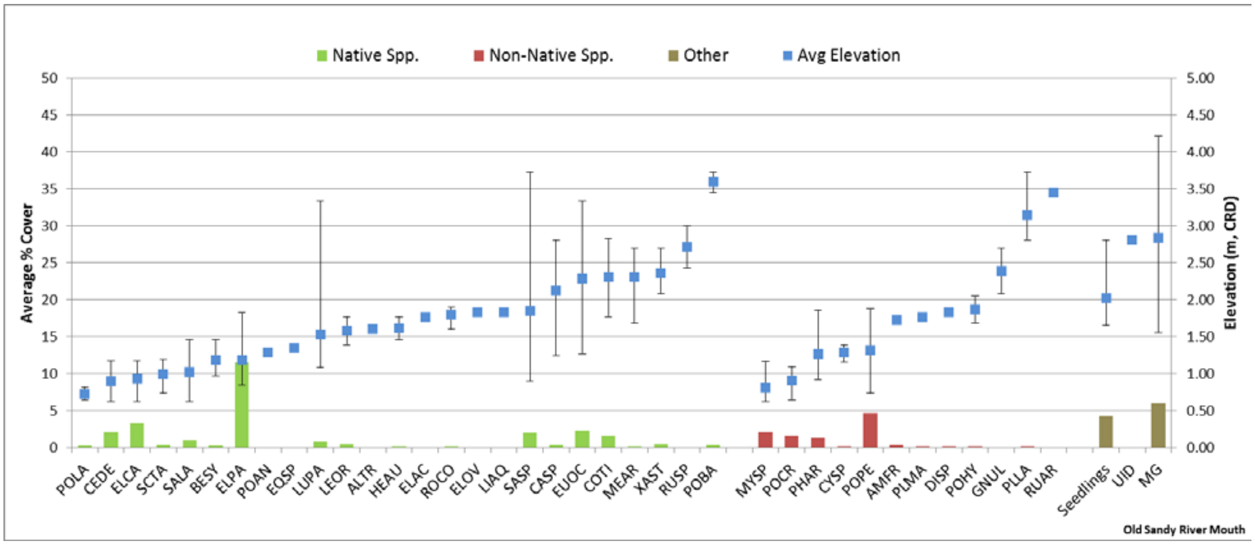

Figure D.6. Average percent cover of the vegetation species (bars) and the average elevation of the species (points) within the sample area of a study site. The minimum and maximum elevations are depicted by the error bars on the points.

How many plant species have been observed by river kilometer?

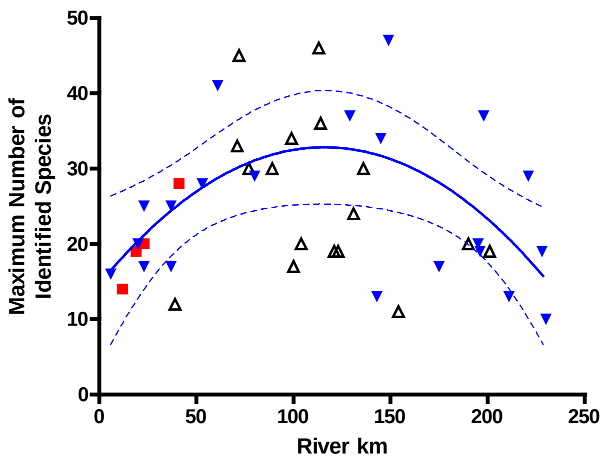

Figure D.7. Maximum number of species observed at reference marsh sites versus the distance of the sites from the river mouth. The mid-river section, associated with the peak of the quadratic curves, had significantly more identified species than the lower and upper portions of the river (Kruskal–Wallis; p = 0.008). The fitted curves and the confidence intervals were based on the least-disturbed marsh sites only (the blue triangles). Red squares represent previously diked sites and open triangles represent sites affected by dredge material placement.

What is the average elevation of the vegetation survey areas relative to river kilometer?

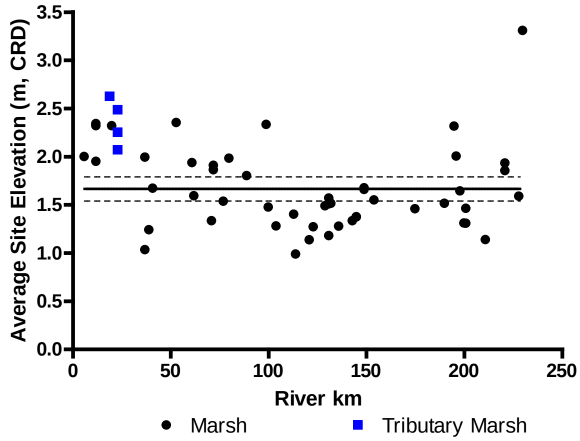

Figure D.8. Average site elevation (meters Columbia River Datum) of the vegetated sample area for emergent marshes along the estuarine gradient. The fitted linear regression and its 95% confidence interval are shown in black.

Distribution

PDF Copies		PDF Copies		
1	**Cynthia Studebaker** USACE Portland District P.O. Box 2946 Portland, OR 97204	6	**Local Distribution** Pacific Northwest National Laboratory	
			Borde AB	Mailstop SEQ
			Coleman AM	Mailstop K9-33
			Diefenderfer HL	Mailstop SEQ
			Johnson GE	Mailstop BPO
			Sather NK	Mailstop SEQ
			Serkowski JA	Mailstop K9-33

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

U.S. DEPARTMENT OF
ENERGY

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)
www.pnnl.gov