

Data, Data and More Data: Managing and Making Sense of Data from PNW Salmon and Habitat Monitoring Programs

Carol Volk

Integrated Status and Effectiveness Monitoring Program



Volk Consulting: Contractor to BPA
Carol.volk@noaa.gov

5/5/2009



- What is data?
- Where is all the data?
- What do we do with it?
- Is it useful?

ISEMP

Restoration Tracking Database

- Should we do something with it?
- Steps forward: making old and new data useful

What is data?

Data refers to facts usually collected as the result of experience, observation or experiment.

Data may consist of numbers, words, or images, particularly as measurements or observations of a set of variables.

Data are often viewed as a lowest level of abstraction from which information and knowledge are derived.

~Wikipedia

What kinds of data are out there?

Project Types

- Monitoring data
- Restoration data

Project sub types:

- Status and trend*
- effectiveness
- implementation
- project information

What kinds of data are out there?

Project Types

- Monitoring data
- Restoration data

Data generators:

- Multiple agencies

Project scale:

- watershed (graduate projects)
- state-wide assessments
- multi-basins

Project sub types:

- Status and trend*
- effectiveness
- implementation
- project information

Project designs:

- BACI
- GRTS randomized
- non-random

What kinds of data are out there?

Project Types

- Monitoring data
- Restoration data

Project sub types:

- Status and trend*
- effectiveness
- implementation
- project information

Data generators:

- Multiple agencies

Project scale:

- watershed (graduate projects)
- state-wide assessments
- multi-basins

Project designs:

- BACI
- GRTS randomized
- non-random

Data types:

- spatial
- tabular

Data formats:

- projections, datums
- text files, Excel, databases

What kinds of data are out there?

Project Types

- Monitoring data
- Restoration data

Project sub types:

- Status and trend*
- effectiveness
- implementation
- project information

Survey types:

- Fish
- Habitat
- Water quality
- Structure

Data generators:

- Multiple agencies

Project scale:

- watershed (graduate projects)
- state-wide assessments
- multi-basins

Project designs:

- BACI
- GRTS randomized
- non-random

Timeframe:

- multi-year projects
- single-year projects
- periodic assessments

Data types:

- spatial
- tabular

Data formats:

- projections, datums
- text files, Excel, databases

Collection methods:

- spawning surveys
- smolt trapping
- habitat evaluations
- multiple water quality parameters

How much data is out there?

3 data generators

2 project types

3 survey types

2 project designs

10 years

2 data types

5 data formats

5 collection methods

10 basins in Washington state

How much data is out there?

3 data generators

2 project types

3 survey types

2 project designs

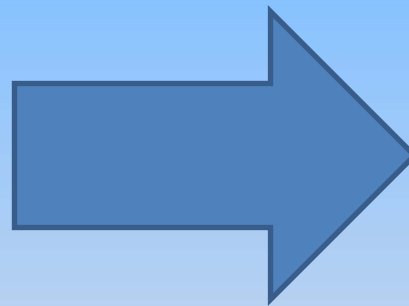
10 years

2 data types

5 data formats

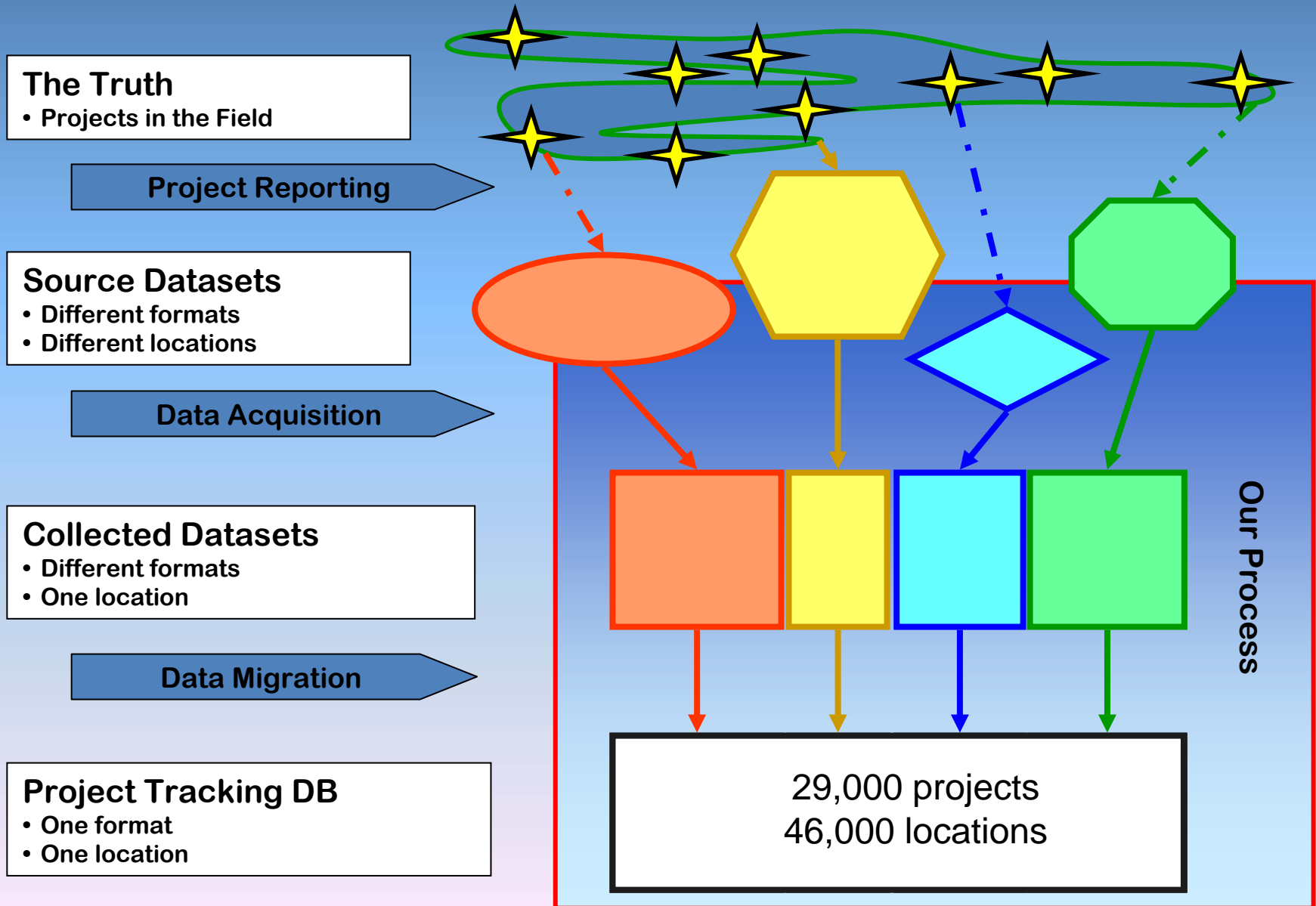
5 collection methods

10 basins in Washington state



180,000
files of data
in Washington
state

HOW: The Data Acquisition Process



Where is all of the data?

- Regional databases
 - EPA STORET
 - DART (Data Access Real Time)
 - USGS (Real time and historic)
 - Metadata portals (REO, UW)

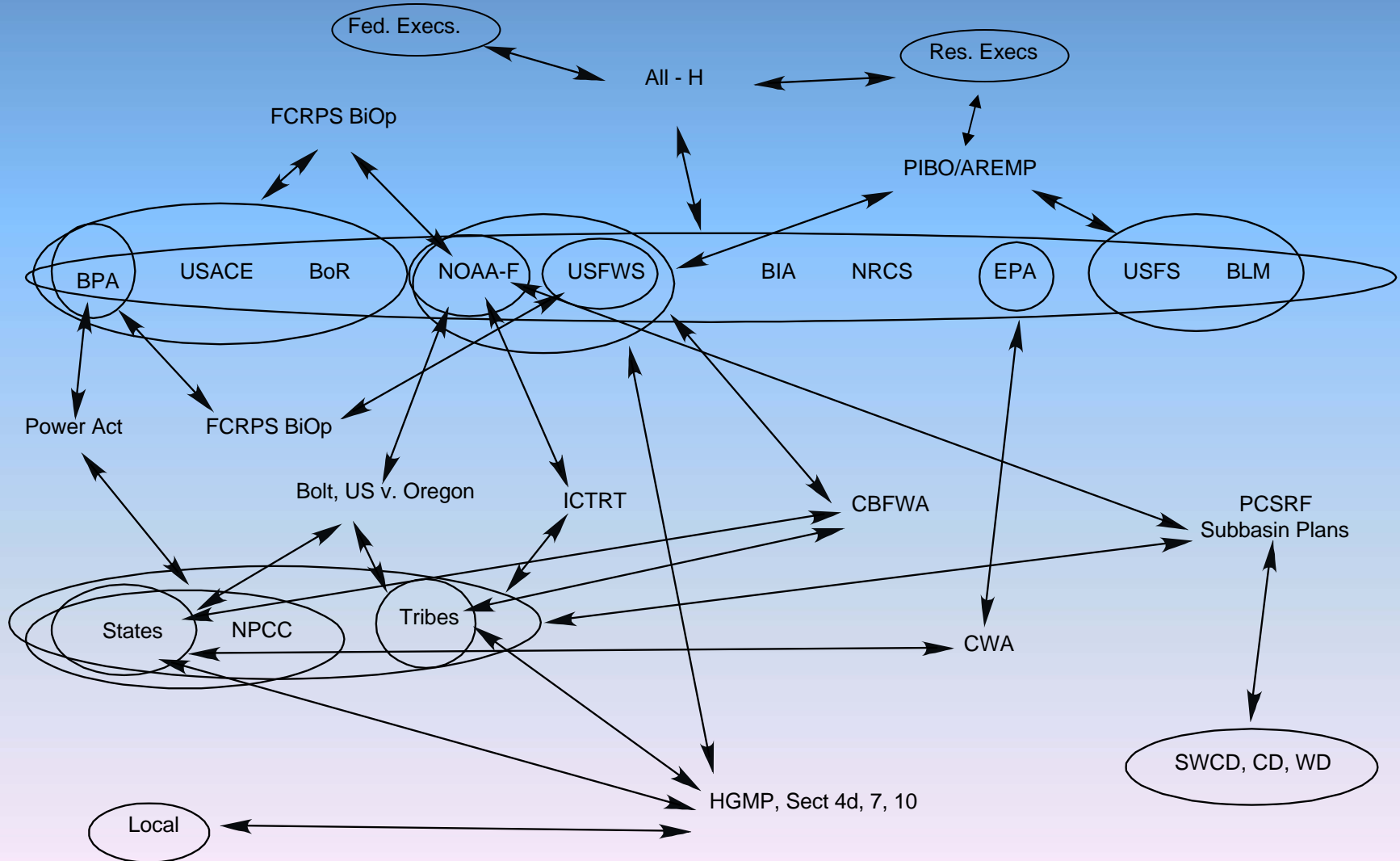
Where is all of the data?

- Regional databases
 - EPA STORET
 - DART (Data Access Real Time)
 - USGS (Real time and historic)
 - Metadata portals (REO, UW)
- Literature
 - Journal articles
 - Agency white papers

Where is all of the data?

- Regional databases
 - EPA STORET
 - DART (Data Access Real Time)
 - USGS (Real time and historic)
 - Metadata portals (REO, UW)
- Literature
 - Journal articles
 - Agency white papers
- “File cabinets”
 - **Possession of the collector**

Coordinating Monitoring, Evaluation and Recovery Planning in the Columbia River Basin



What do we do with data?

Processing

- Collect and compile
- Deal with formatting
- Deal with data definitions and crosswalking
- Migrate data into desired format
- Summarize
- Analyze
- Provide instruction

- What is data?
- Where is all the data?
- What do we do with it?
- Is it useful?

ISEMP

Restoration Tracking Database

- Should we do something with it?
- Steps forward: making old and new data useful

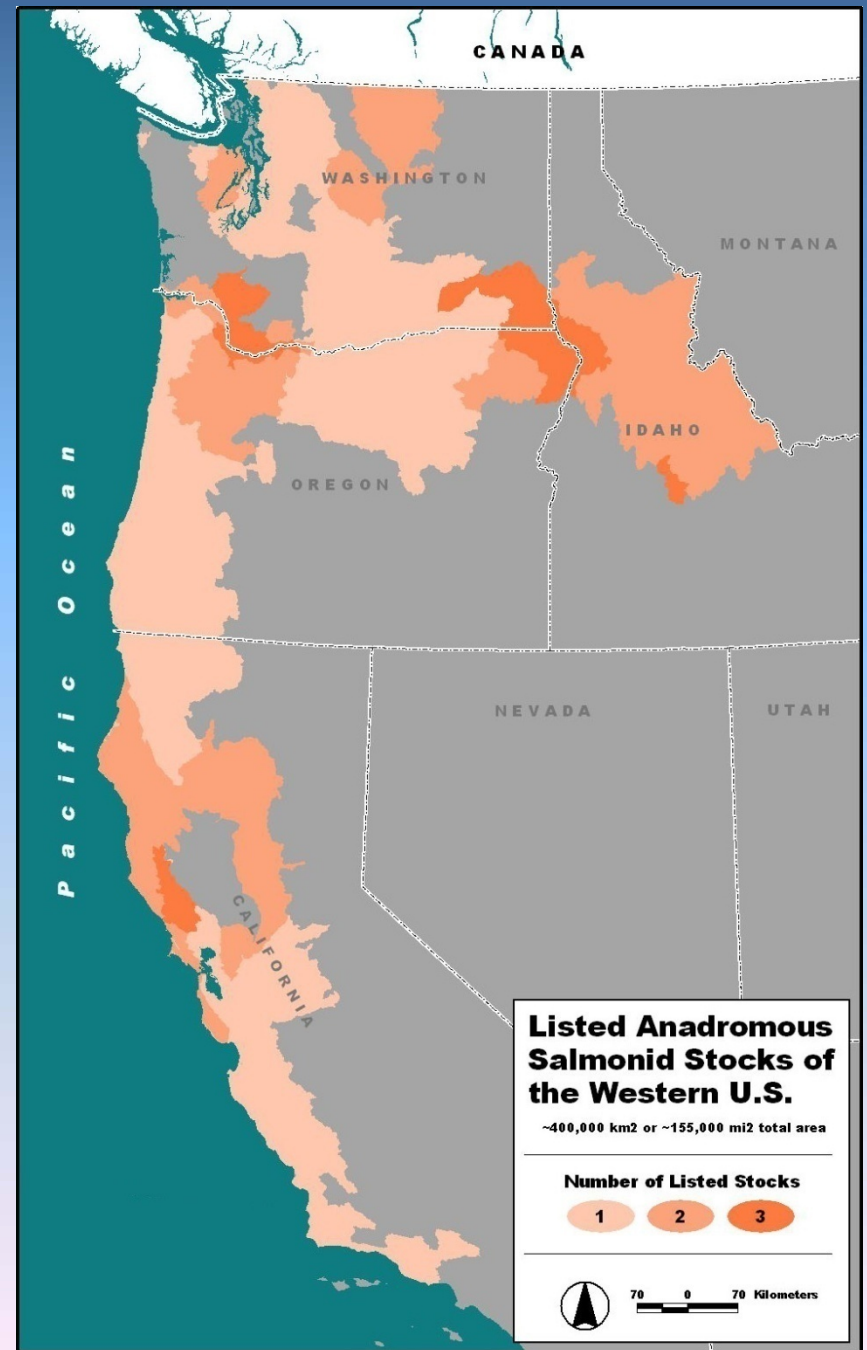
Is this data useful?

Examples:

1. ISEMP John Day water quality project
2. Data dictionaries and the restoration project tracking database
3. Project coordination: Restoration and project placement in MFJD basin

ESA listings of Anadromous Salmonid ESUs

- Recovery and Viability Criteria being developed
- Recovery Plans being developed
- Implementation of Recovery Actions underway in many areas
- 5 Year Status Review cycle initiated



Integrated Status and Effectiveness Monitoring Program ISEMP

www.nwfsc.noaa.gov/isemp

Project goal is to develop a process for creating and evaluating monitoring programs for listed salmon

Developing:

- experimental designs for restoration projects
- experimental designs for status and trend monitoring
- metrics and indices
- plans for supplementing existing monitoring programs
- designs for a variety of life stages (smolt, juveniles, adults)
- designs impacted by variety of habitat variables (reach to landscape)

-DATA MANAGEMENT

ISEMP objective #1:

Assess monitoring data in pilot basins

- Figure out what data is available in pilot basins
- Compile data
- Summarize data
- Evaluate data gaps
- Offer direction to monitoring programs

Compilation of Historic Data in the John Day Basin

Data collection in 2005-2006:

~80 hours in 3 month period

Processing data:

~1300 hours over 12 months 2006 (formatting)

~2000 hours in 2007-2008 (metadata updates)

Analyze data:

-ONGOING

What we found...

- Large quantity of historic data
- Multiple data types
 - water quality
 - habitat
 - fish abundance
- No consistent data format
- No central data storage within agencies
- Agencies struggling with multiple reporting responsibilities

30,000 Stream Temperature Records

Then:

-15 collecting agencies/groups

-20 years of data

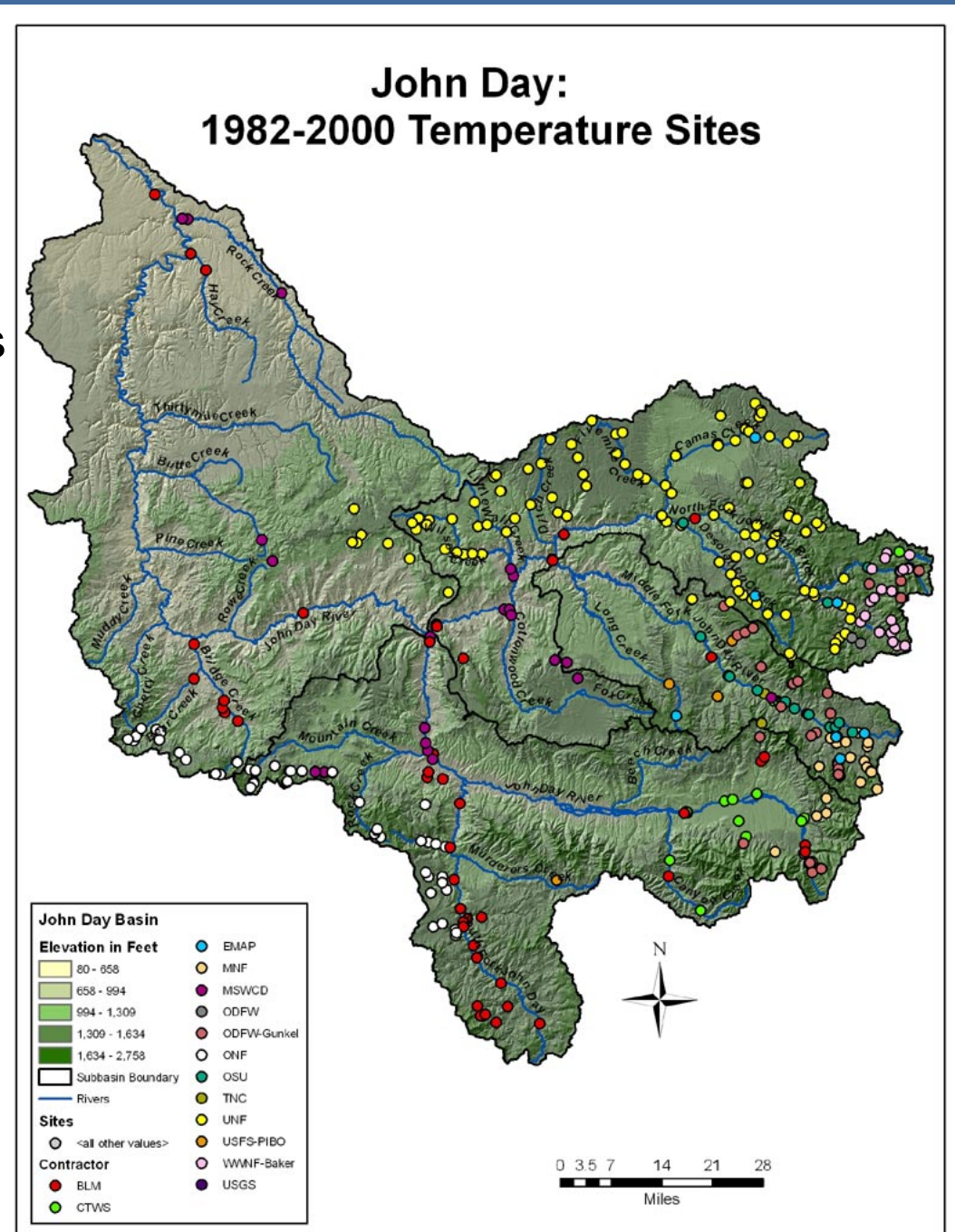
-300 locations

-0 coordination and standardization

Now:

-1 Data Dictionary

-1 Water quality database



Limiting Factor for any M&E Program

'Messy'
field data



Regional
data analysis

- large volume
- multiple types
- multiple sources
- multiple formats

- centralized source
- standardized format
- metadata
- processing tool
- summary analysis
- distribution

ISEMP Data Management Strategy

- **Data Storage and Analysis**
 - RME Status and Trend Monitoring Database
 - Metadata tracking
 - Protocol manager tool
 - Data dictionary
- **Data Processing and Reporting**
 - Database templates
- **Data Distribution**
 - Salmon Data Management website
 - ISEMP website
 - Geo-databases

Why a data dictionary for salmon habitat restoration?

- Scale of salmon recovery- larger than the analysis scale the data were created for
- Need to combine current and historic data



Necessitates the creation of tools to handle disparate data

Example #2: Creating a data dictionary for habitat restoration data:

Bottom up approach :

1. Look at the universe of available project data

Sources (federal, state, private etc.)

Project Attributes (time, contact, title, type, description)

Location Attributes (state, county, lat/long, stream)

2. Decide what information is important to the project purpose

Use the data to formulate future restoration and monitoring plans

Who, how, what, where, when, why

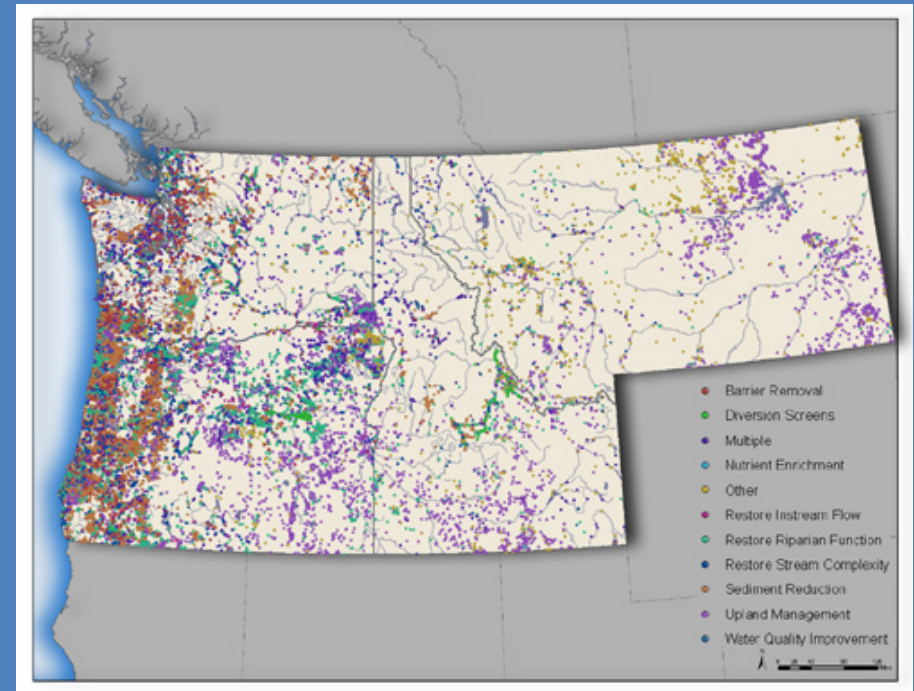
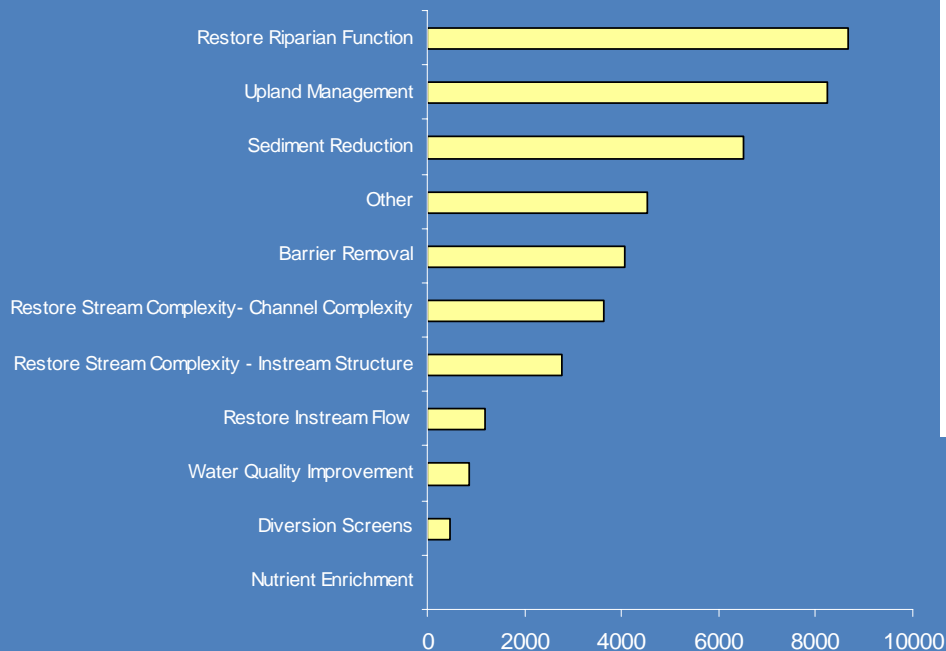
3. Define that information and resolve it into a common language using crosswalks

Need for effectiveness monitoring

Project type and location standardization

Pacific Northwest Salmon Habitat Project Database

- 26 data sources
- 29,000 projects at 47,000 locations in WA, OR, ID, MT



<http://webapps.nwfsc.noaa.gov/pnshp>



Number of projects by Source

OWRI	7,428
REO	6,065
BLM	4,740
SRFB	857
NRRSS	798
WDFW Fishway	620
WDFW WRIP	609
ASOTIN_NEW	543
GRMWP	531
Ducks Unlimited	529
OR Water Trust	492
WA DOE	433
Montana Water Center	311
IDFG	242
IDFG Screen Shop	220
StreamNet	193
John Day NMFS	185
Salmon NMFS	183
L. Col., Clearwater, John Day	152
CBFWA	151
NOAA Restoration Center	142
WDFW Culvert	130
IDAHO DEQ	129
WDOT	51
Washington Water Trust	20
Wenatchee NMFS	13

Welcome

Pacific Northwest Salmon Habitat Project Tracking

The forms on the PNSHP site use Oracle's JInitiator, which should be installed automatically when you access one of them for the first time. If you are experiencing problems, please contact us at 206-860-3433 or nwfsc.sdm@noaa.gov.

- Barrier Removal
- Diversion Screens
- Multiple
- Nutrient Enrichment
- Other
- Restore Instream Flow
- Restore Riparian Function
- Restore Stream Complexity
- Sediment Reduction
- Upland Management
- Water Quality Improvement

Project Count By Type

Sediment Reduction	6,584
Restore Riparian Function	6,166
Upland Management	5,914
Barrier Removal	3,811
Restore Stream Complexity - Instream Structure	2,736
Other	2,307
Restore Stream Complexity - Channel Complexity	1,830
Restore Instream Flow	958
Water Quality Improvement	801
Diversion Screens	442
Nutrient Enrichment	45

[PNSHP Document List](#)



[US Department of Commerce](#) | [NOAA](#) | [NMFS](#) | [NWFSC](#)
 Northwest Fisheries Science Center
 NOAA/NMFS/NWC
 2725 Montlake Blvd. E
 Seattle, WA 98112
 Phone: (206) 860-3433
[Privacy Policy](#) | [Disclaimer](#)
[Login](#)

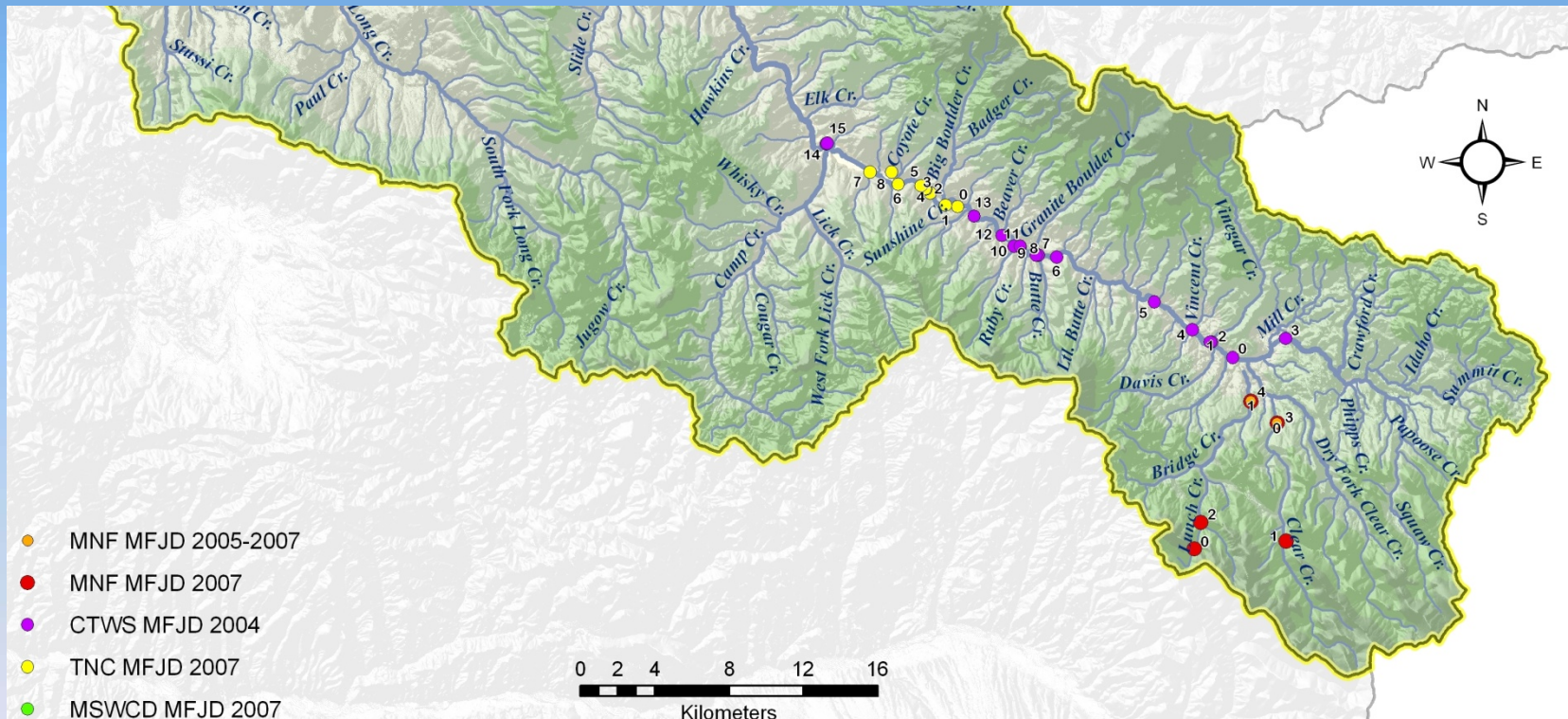
25,500+ Projects

40,000+ Locations

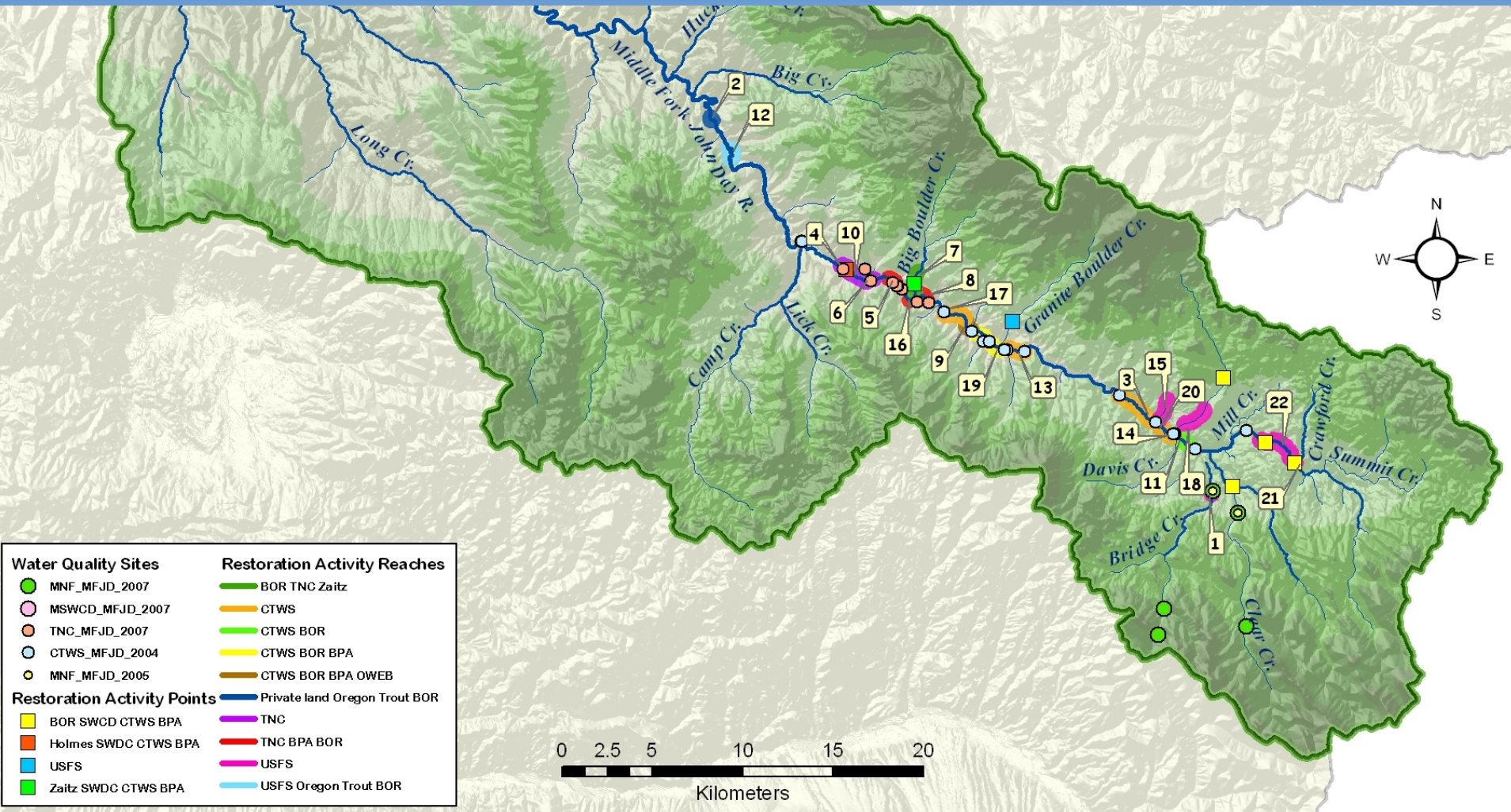
This Project was driven by the need for monitoring but very little monitoring data is available for restoration projects

Develop new tools to describe monitoring activities using a standardized set of terms

IMW restoration and temperature logger site coordination



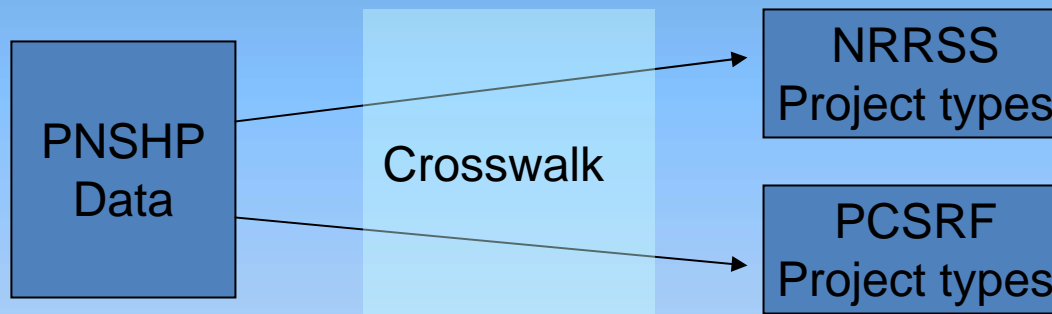
Restoration and logger sites



Should we use all of this data?

- What information is available about the data?
 - Can you translate it?
 - What was the study design?
 - What was the purpose of the study?
- Is data dense enough for the question of interest?
- What is the quality of the data?
- How long will it take to assemble the data?
 - Data requests?
 - Formats?
- Can you trust it?

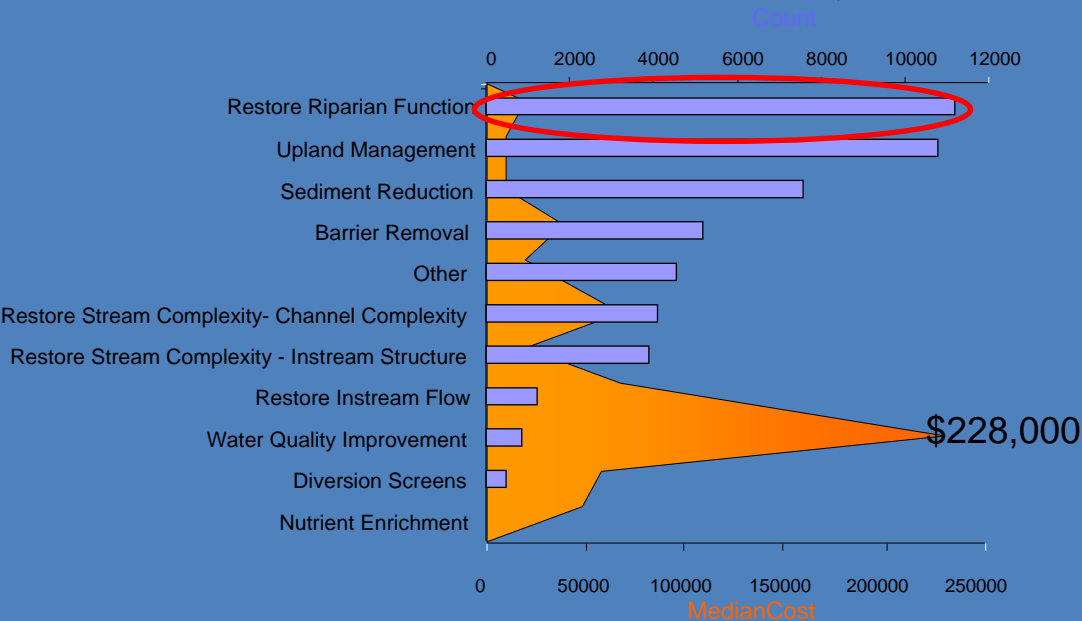
How do different data dictionaries affect project metrics (type, cost)?



Crosswalks allow us to relate one data dictionary to another

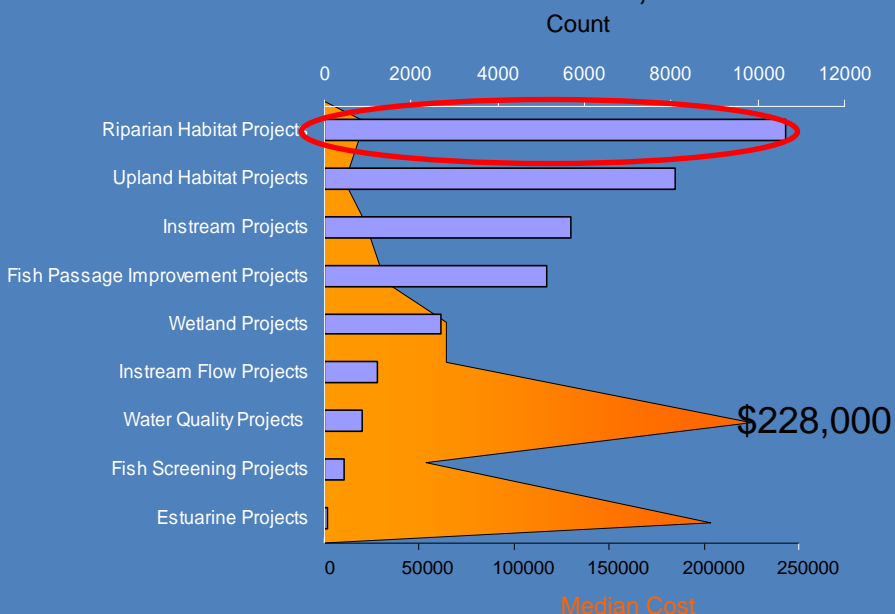
PNSHP Type	PNSHP Subtype	PNSHP Type Definition	NRRSS Type	NRRSS Type Definition
Sediment Reduction	All Subtypes	<i>Projects that diminish excessive sediment transport into streams</i>	Water Quality Management	Practices that protect existing water quality or change the chemical composition and/or suspended particulate load. Remediation of acid mine drainage falls into this category as does CSO separation. Excludes urban runoff quantity management (see Stormwater Management).
Water Quality Improvement	All Subtypes	<i>Projects that improve stream water quality parameters.</i>		

PNSHP 49,844

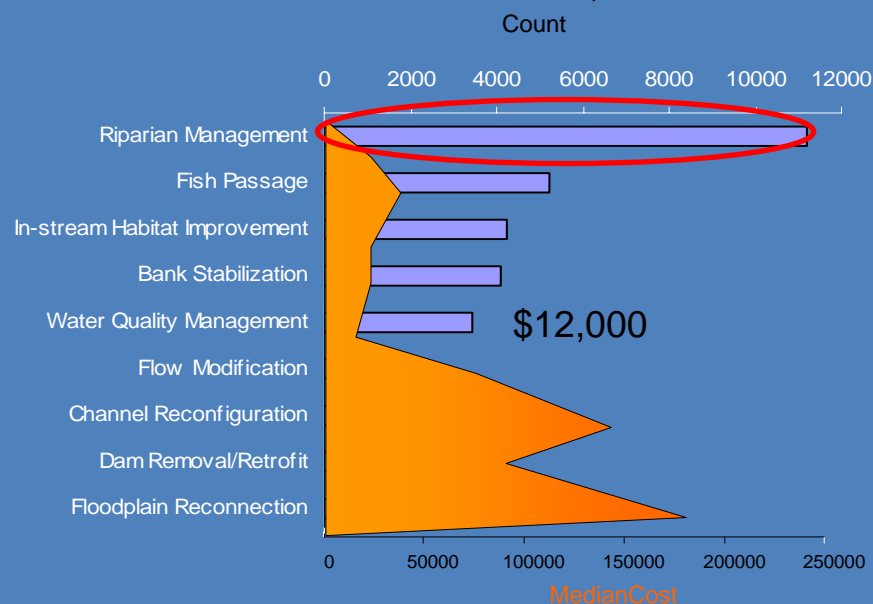


Does the choice of data dictionary change the results?

PCSRF 34,866

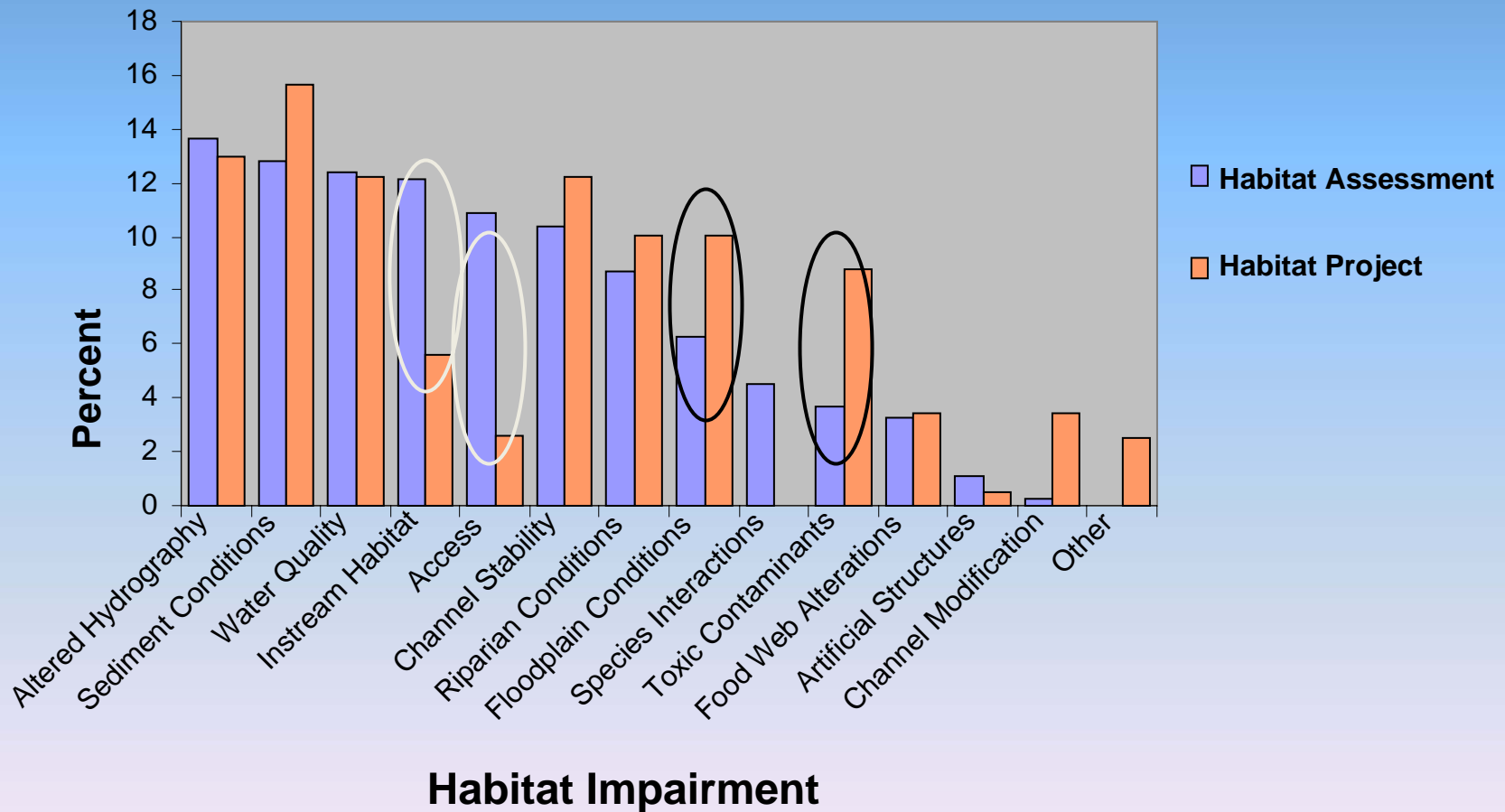


NRRSS 30,615

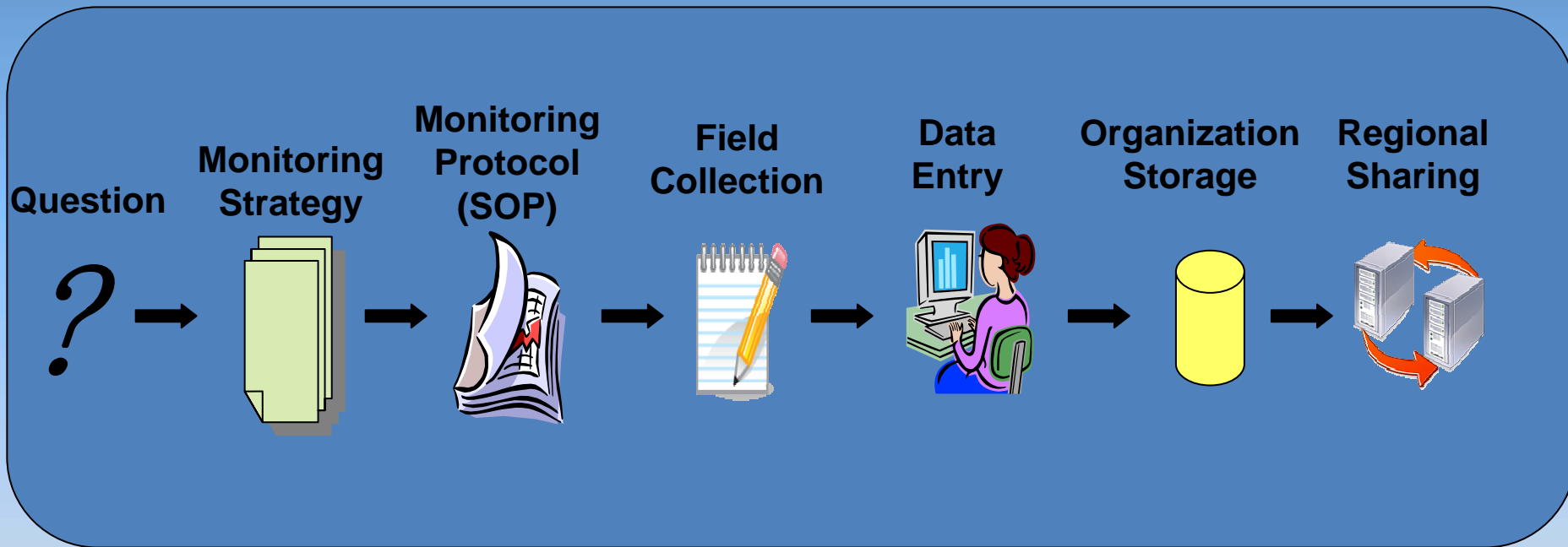


Level of detail of data dictionary should match the intended use

Do restoration projects address habitat need?

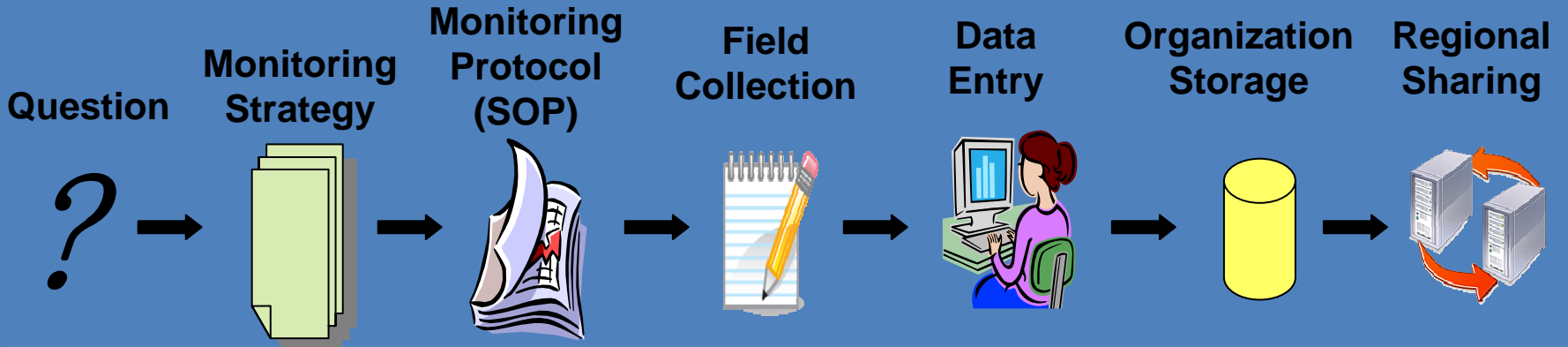


Work Flow

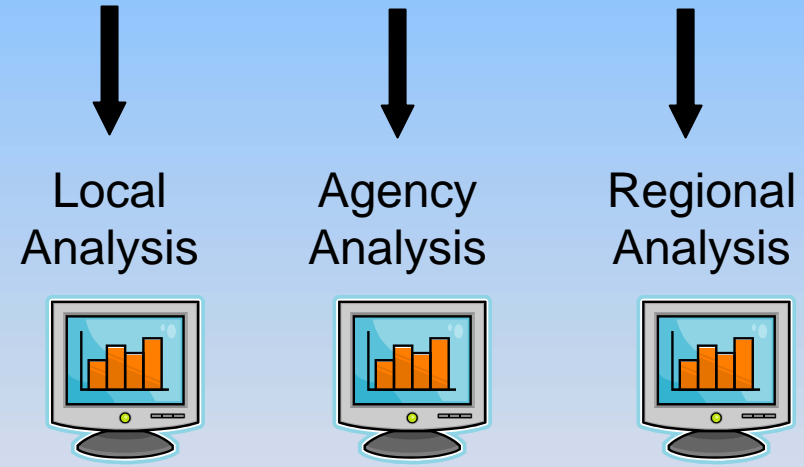


Organizations use workflows to produce products

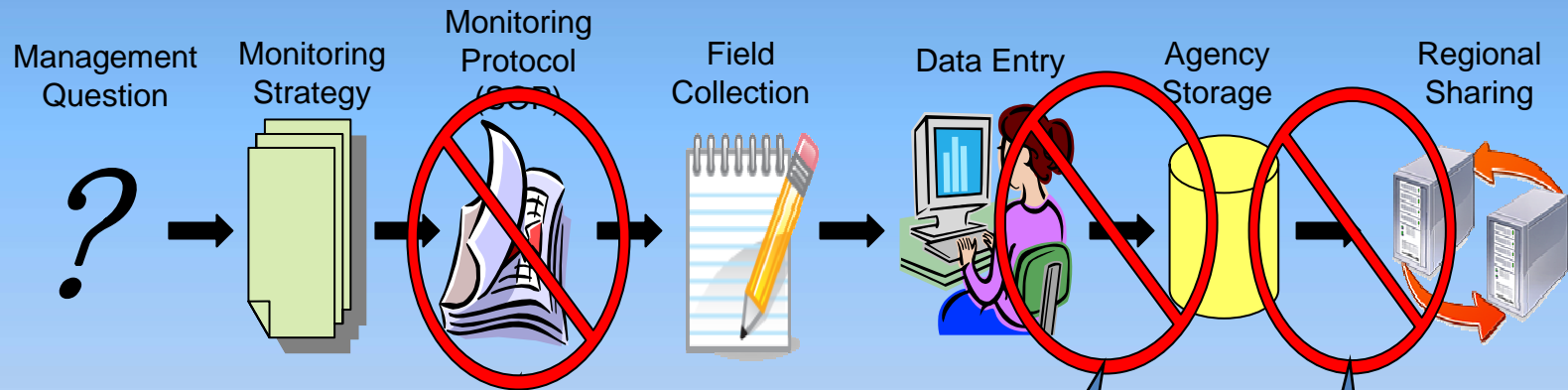
Work Flow



Products



Work Flow Gaps



Meta data does not exist in most cases.

Data is not entered into agency databases or standard formats and data is not validated during data entry. These issues impede the flow of data to agency databases and limit regional data sharing.

Regional schema or exchange formats have not been defined. Vocabulary is not consistent between organizations

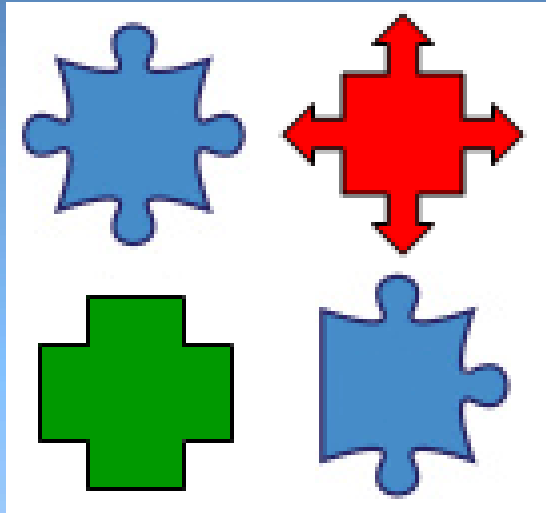
Road blocks

- Disparate data dictionaries
- Gaps data flow
- Objective issues:
 - Evolving questions and technology
 - Short term goals: get answers now
- Changing Personnel
 - Current structure relies heavily on “Legacy knowledge”
- Lack of training support/time

Making data more useful

- Why?
 - Easy to remember what you did
 - Increases chance that data will be used 'properly'
 - Greater good/greater success of field

Data Sharing



- Well documented
- Discoverable
- Accessible

- Inconsistent Formats
- Inconsistent Methodologies

Data Interoperability



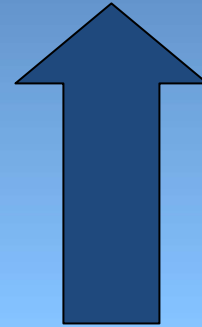
- Well documented
- Discoverable
- Accessible

- Consistent Formats
- Comparable Methodologies

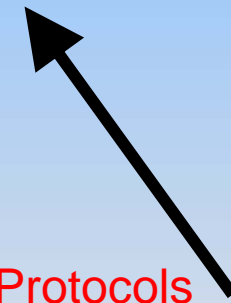
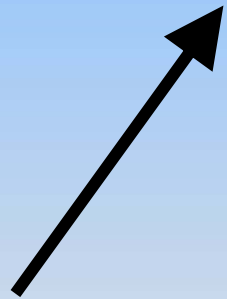
Regional Data Analysis

Standard
Analyses

Approved
Protocol



Central Repository



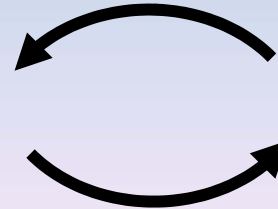
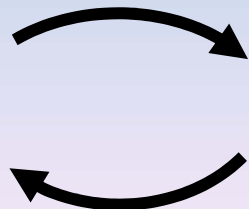
Validation

Protocols

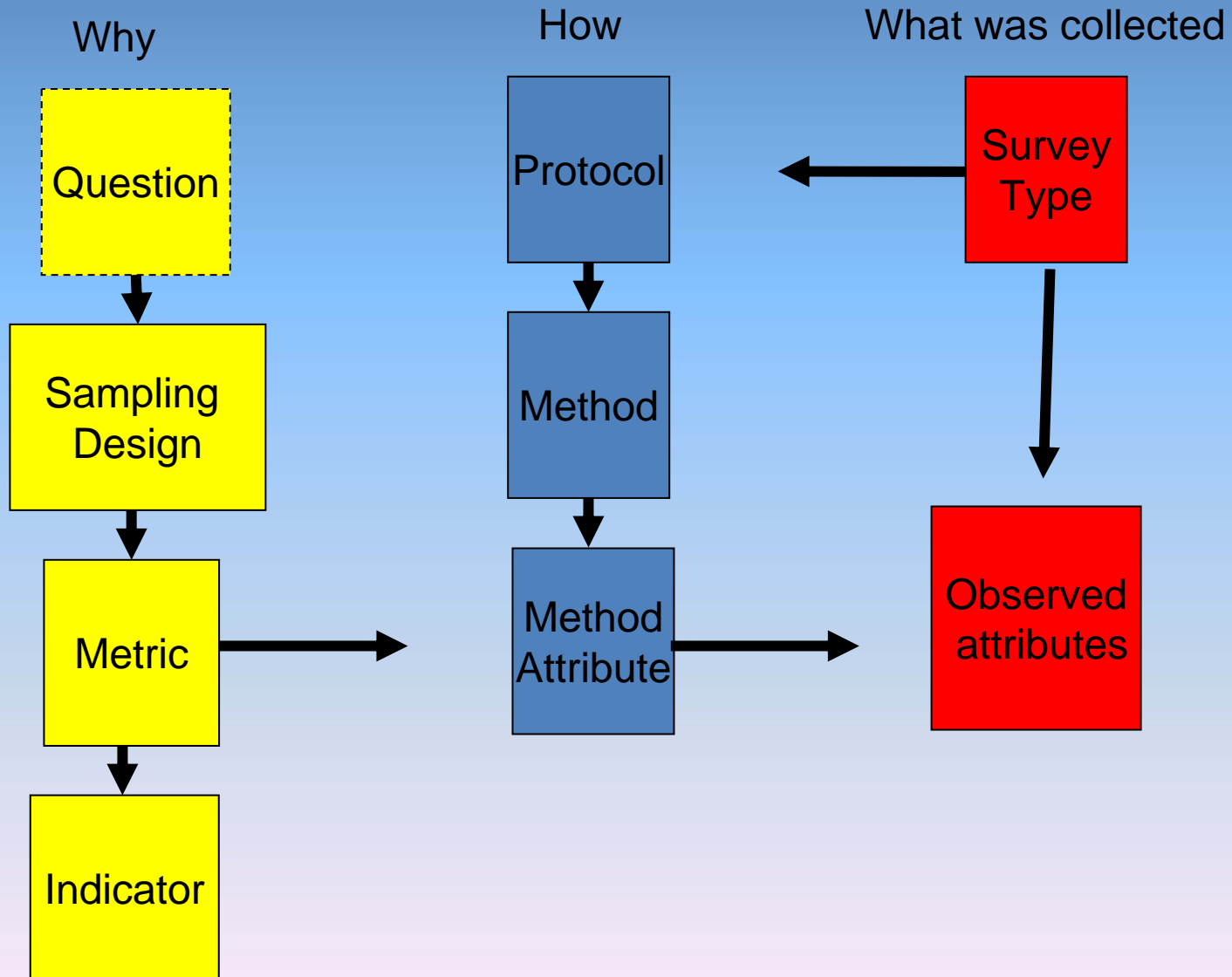
Database
Template

Data Collection

Protocol
Manager



What needs to be recorded and managed?



How do we do it?

- Metadata
 - Data dictionaries (PNAMP, NOAA-Fisheries)
 - Sampling designs
 - Protocols (References, Methodology)
 - Methods
 - Codes
- Managing observed tabular data
 - Normalized structure
 - Consistent
- Geodatabases (e.g. manages projections)

Metadata: Data Dictionaries can manage the community's data flow -- Protocols

Review Protocol

ProtocolName Washington State Department of Ecology Stream Habitat 2006

ShortName WSDOE 2006 **ParentName**

Contact Glenn Meritt

Title

Reference Meritt, G. 2006.

Description This protocol is intended to measure biological and physical/environmental indicators of stream habitat following the recommendations of the Upper Columbia Monitoring Strategy

ProtocolVersion 3

ProtocolProduction Development

ProtocolDate 1/12/2007

Record Created 1/12/2007 10:53:00 AM rentmeesterst

Record Updated

Methods

	MethodName	Order	Shape	UnitSystem	Version	Reference
▶	Survey Event	1	Reach	Metric	3	Jeremy Moberg. 2
	Macroinvertebrate Sampling	2	Quadrat	Metric	3	Jeremy Moberg. 2
	Slope and Bearing Measurements	3	Line-transect	Metric	3	Jeremy Moberg. 2
	Cross-sectional Transects	4	Line-transect	Metric	3	Jeremy Moberg. 2
	Riparian Assessment	5	Point	Metric	3	Jeremy Moberg. 2
	Fish Cover	6	Channel segment	Metric	3	Jeremy Moberg. 2
	Human Influences	7	Channel segment	Metric	3	Jeremy Moberg. 2
	Longitudinal Profile	8	Point	Metric	3	Jeremy Moberg. 2
	Large Woody Debris Assessment	9	Channel segment	Metric	3	Jeremy Moberg. 2

Record: 1 of 9

Record: 1 of 8

Metadata: Data Dictionaries can manage the community's data flow -- Methods

Review Method

MethodName **Order** **Shape**

Reference **UnitSystem** **ProtocolName**

Description **Date** **Version**

Record Created
1/12/2007 11:47:00 AM
Record Updated

Attributes

Order	TableName	AttributeName	Alias	Description
1	tbl_Station	StationName		Natural primary key for tbl_Station. Created by concatenating [Dce
2	tbl_Station	StationType		Type of station (e.g. Transect, Thalweg, intermediate)
3	tbl_Station	TransectName		Foreign key from tbl_Transect
4	tbl_Station	TransectDirection		Direction to transect (upstream or downstream)
5	tbl_Station	BankReference		When defining left and right bank, are you looking upstream or dow
6	tbl_Station	Station	BF_PROP	Name of station as assigned by researcher
7	tbl_Station	DistanceFromTransect		Distance to transect
8	tbl_Station	DistanceFromLeftBank		Distance from left bank
9	tbl_Station	BarPresent		Is a mid-channel bar present?
10	tbl_Station	Dry		Is the channel dry?
11	tbl_Station	StationWettedDepth	WET_DEPTH	Wetted depth at station
12	tbl_Station	StationEmbeddedness	EMBED	Embeddedness at station
13	tbl_Station	StationSubstrateType	SIZE_CLS	Substrate type or size class observed at station
14	tbl_Station	StationNotes	Comments_08	Notes about this station
15	tbl_Station	StationMeasurementNotes		Notes about your station measurement
5	tbl_Transect	LeftBankfulHeight	LB_BH	Left Bank Bankfull Height
6	tbl_Transect	RightBankfulHeight	RB_BH	Right Bank Bankfull Height
7	tbl_Transect	LeftBankStability	Stability	Stability of left bank
8	tbl_Transect	RightBankStability	Stability	Stability of right bank
9	tbl_Transect	LeftBankUnderCut	Undercut	Length of undercut on left bank
10	tbl_Transect	RightBankUnderCut	Undercut	Length of undercut on right bank

Record: of 26

Metadata: Data Dictionaries can manage the community's data flow -- Attributes

Review Attribute

TableName
AttributeOrder
ProtocolName Washington State Department of Ecology Stream Habitat 2007

AttributeName
MethodName

AttributeDescription

DataType
ExpectedLowerRange
AttributeAlias

AttributeUnits
ExpectedUpperRange
AttributeDefaultValue

AttributeDecimals
AttributeMinimum
AttributeAllowNull

AttributeLength
AttributeMaximum
AttributeVisible

AttributeNotes

Record Created 1/12/2007 4:10:10 PM

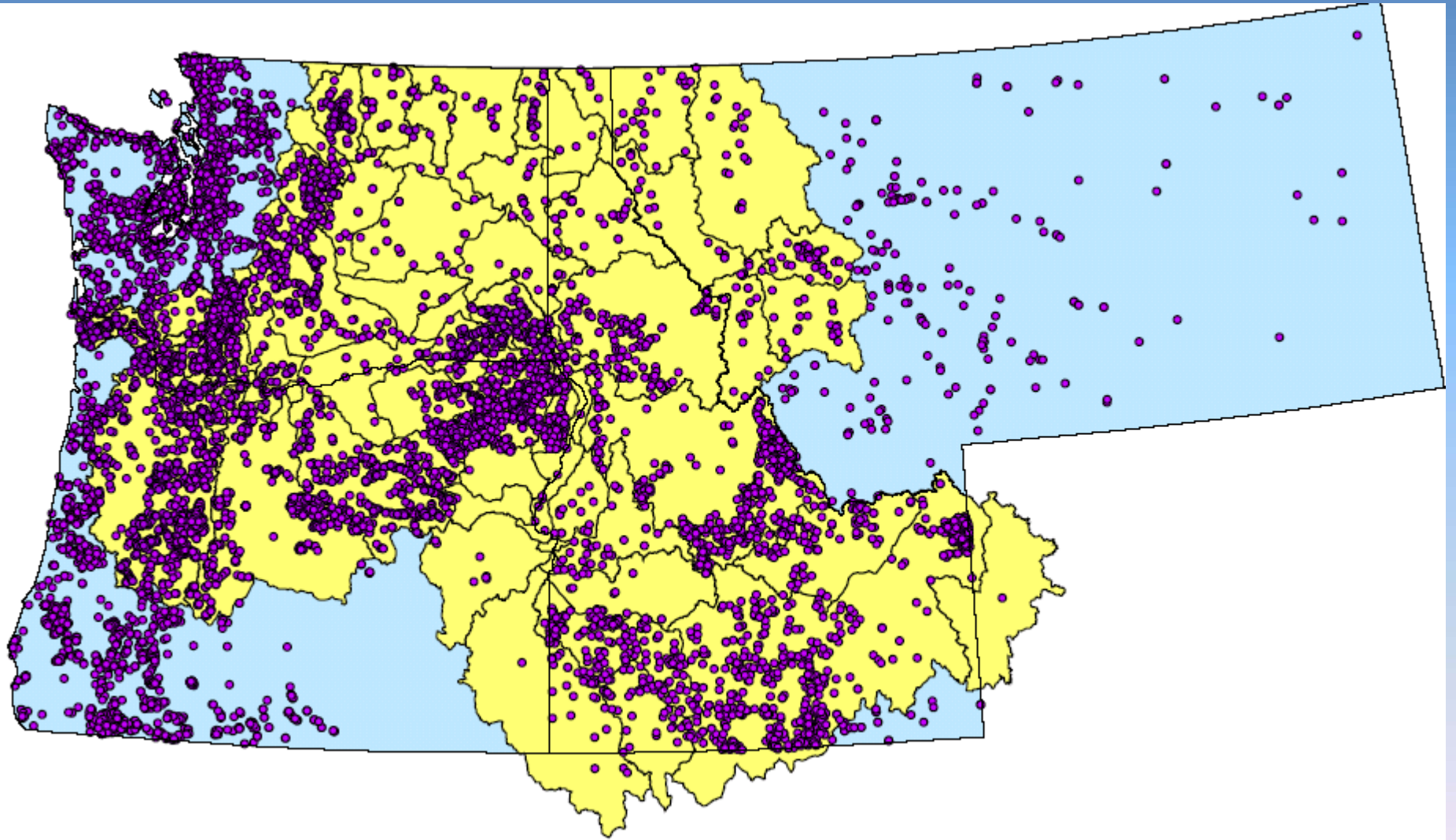
Record Updated

Code Values

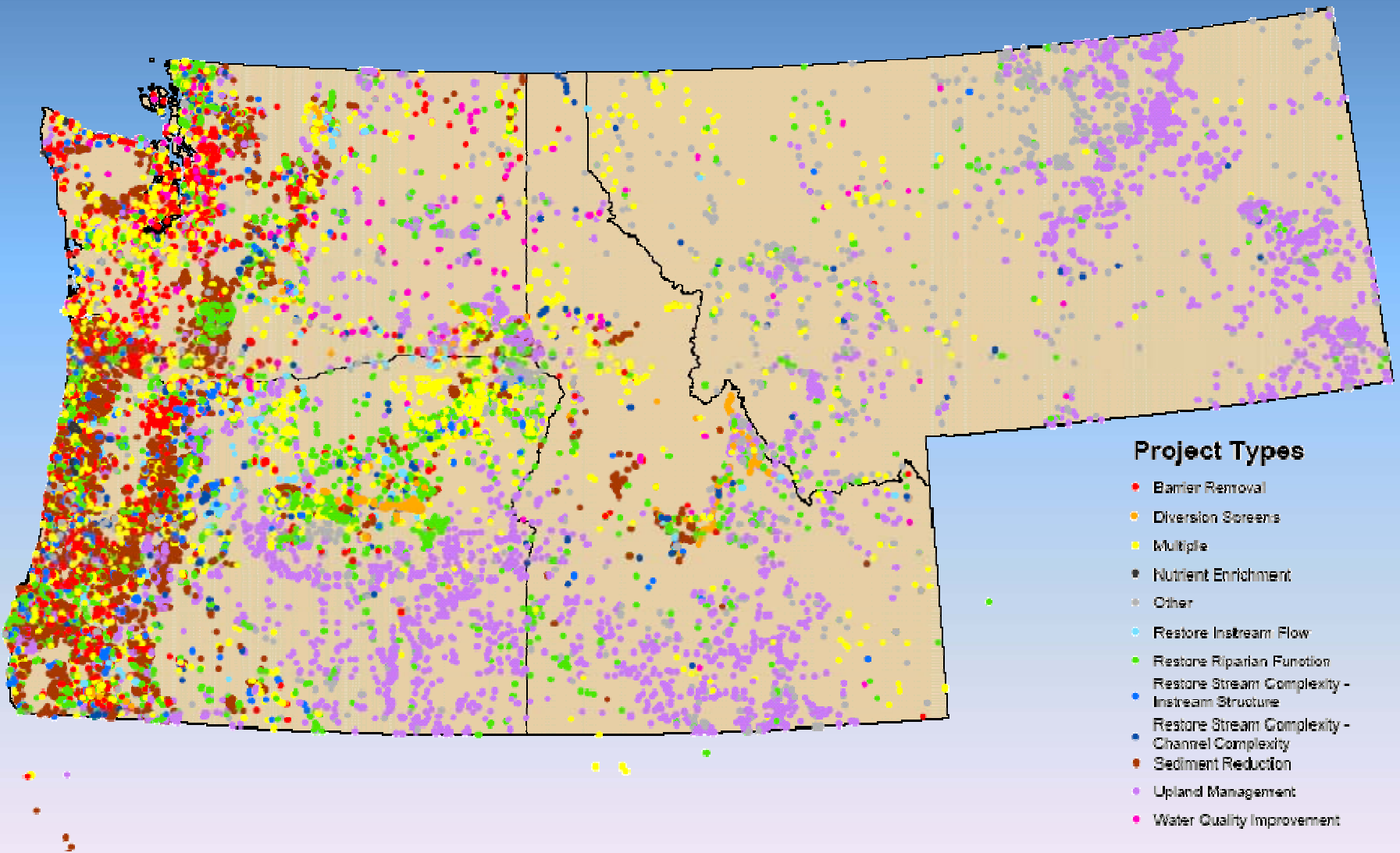
Order	Code Value	DomainCode	Description
5	Cobbles	CB	> 64 to 250 mm; Tennis ball to Basketball size
9	Fines	FN	< 0.06 mm; Silt, clay, muck and not gritty between fingers
6	Course Gravel	GC	> 16 to 64 mm; Marble to tennis ball size
7	Fine Gravel	GF	> 2 to 16 mm; Ladybug to marble size
3	Hardpan	HP	> 4000mm; Firm, consolidated fine substrate
11	Other	OT	Concrete, metal, tires, car bodies, etc
2	Rough Bedrock	RR	> 4000mm; Rough surface rock bigger than a car
1	Smooth Bedrock	RS	> 4000mm; Smooth surface rock bigger than a car
8	Sand	SA	> 0.06 to 2 mm; Smaller than ladybug size, but visible as particles and gritty between fingers

Record: of 1 (Filtered)

25,000 Habitat Restoration Projects Across the Pacific Northwest



25,000 Habitat Restoration Projects Across the Pacific Northwest with a Data Dictionary

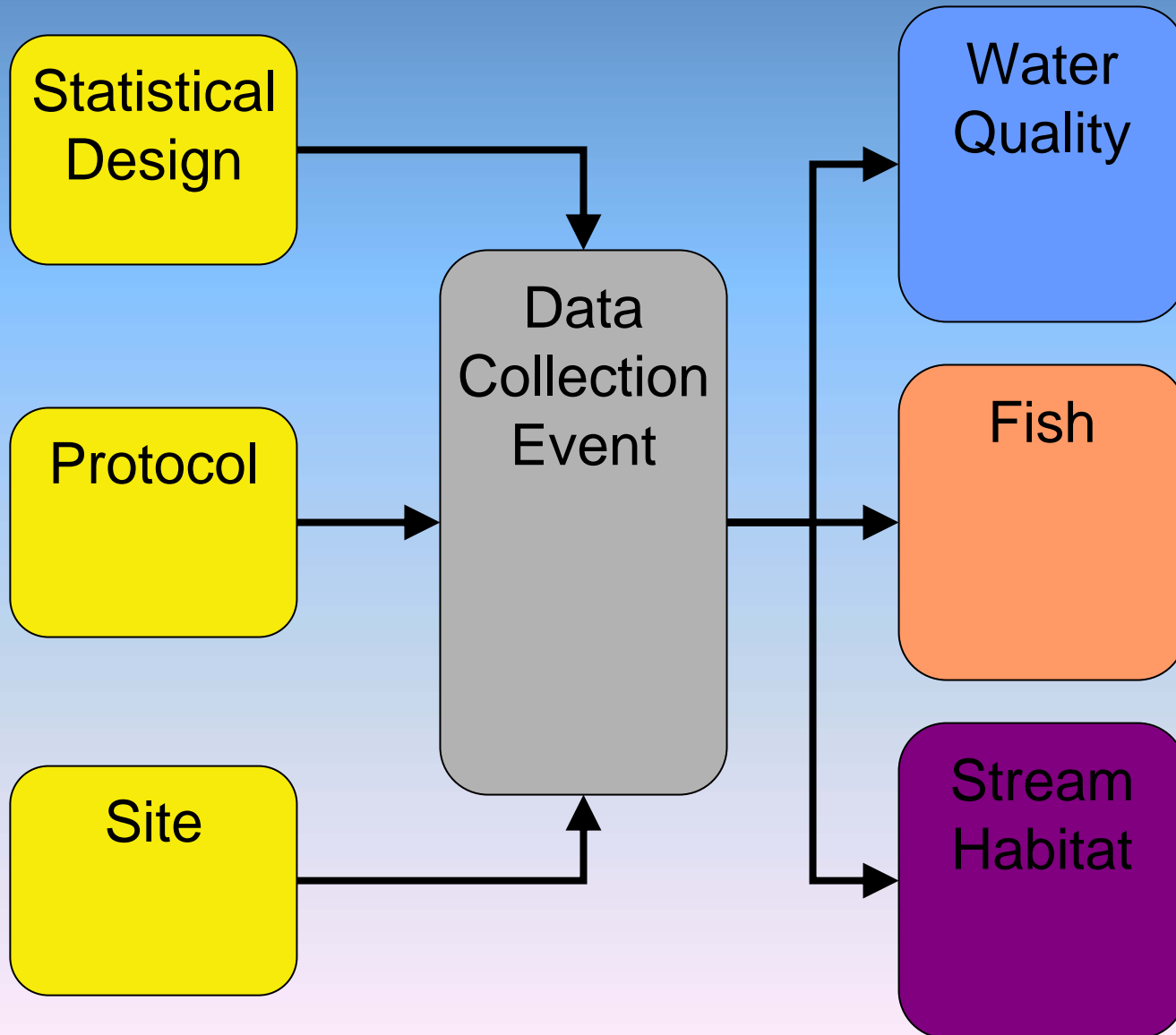


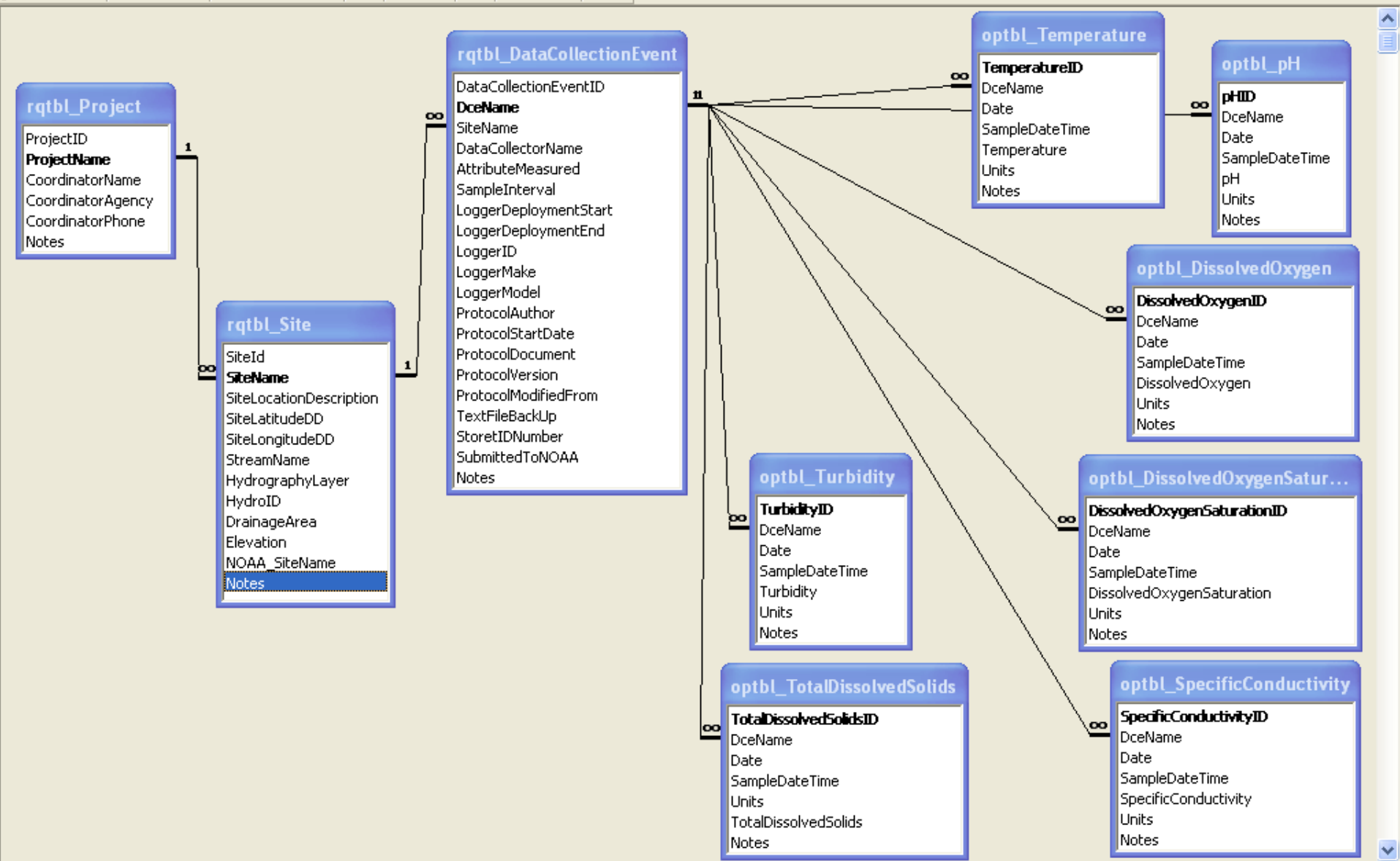
The Salmon World with and without Data Dictionaries

- Examples of the utility/power of going through the data dictionary process
- Doesn't solve everything, but helps go a long way with the data -> information transition

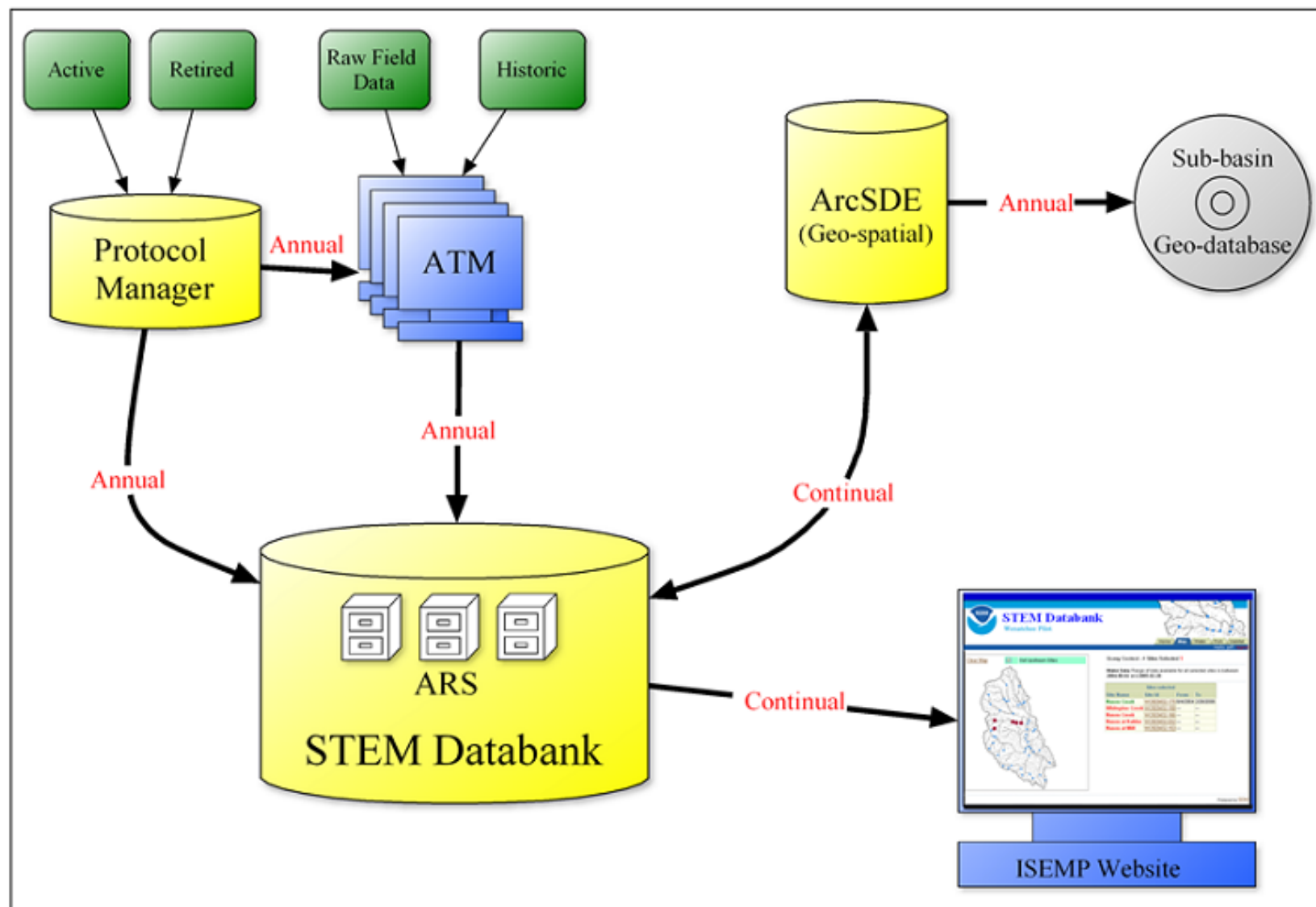
ISEMP's Aquatic Resources Schema

Normalizing observation data storage





ISEMP Data Management System



Legend:

- Raw Data (Green box)
- User Interface (Blue box)
- Database Container (Yellow box)
- DVD (Grey box)

Observation Data Summary

John Day: thousands of Excel files ~ 10 MS Access databases

Wenatchee: 4 years of monitoring data into 7 databases
(one per survey type)

Entiat: 100 potential restoration sites managed in one gdb

Spend our time improving metadata rather than reformatting datasets

Data management toolbox

1. Developed agency and regional tools for data processing:
Aquatic Resources Schema and template databases
2. Centralized data repository (STEM)
 - Data storage
 - Metadata tracking
 - Data analysis
3. Metadata tracking
 - Library of protocols, methods and attributes

Data Management Toolbox

4. Queries and standards for data quality
5. Summary and analysis tools
 - Library of analysis queries and definitions
6. Tools for data distribution
 - Web interface database
 - General project website
 - Geo-databases

In Summary

- Current methods of storing data is inefficient for regional analyses and conclusions
 - Data flows
 - Data storage tools
- Current datasets can be useful but takes time
- Start small.

Organize once and analyze a thousand times.

Acknowledgements

ISEMP Team (Chris Jordan, Pamela Nelle, Nick Bouwes, Jody White, Chris Beasley)

- Especially Steve Rentmeester

Upper Columbia River Data Steward (James White)

Katie Barnas, NWFSC