

QUALITY ASSURANCE FOR EMAP-WEST SURFACE WATER ASSESSMENTS

David V. Peck
U.S. EPA

National Health and Environmental Effects Research Laboratory
Western Ecology Division
Corvallis, OR



EMAP Assessment Workshop
January 2003



Where are We Coming From?

❖ EMAP-SW today:

- << 20 people to "do QA" for lots of different indicators/data components
 - ☞ QA is everyone's business!!
- QA with purpose:
 - ☞ Results provide value, rather than QA for QA's sake
- Stealthy (it's there but you don't notice)
- Assist vs. Audit
 - ☞ Help to head off problems rather than just finding them later

The Challenges of EMAP-West

❖ Complicated Survey Design

- Integrated survey components
- Design constraints
 - ☞ Partition groups
 - ☞ RF3 limitations

❖ Large spatial scale, with focus on populations, not sites

- Analytical error is a relatively small proportion of total variance
 - ☞ Tends to be where QA effort, \$\$\$ is traditionally focused
 - Can't ignore, but overkill doesn't help much
 - ☞ Can affect status estimates if large
 - Analytical error confounded with natural variation

The Challenges of EMAP-West

❖ Indicators

➤ Different types:

- ☞ Chemical: QA/QC toolbox is well-stocked
- ☞ Biological and Habitat:
 - More complex indicators
 - QA/QC tools less well-established, subject to limitations

➤ Response design- what is done at each sampling "point"

- ☞ Different for different indicators
- ☞ Integrated into a daily operational scenario that can be consistently implemented by a field crew at a lot of different stream types and still provide comparable data

The Challenges of EMAP-West

❖ The Reality of Logistics

➤ One shot sampling + Short Index Period + Lots of data=

☞ "Win or go home..."

- Do it right the first time

☞ "Is that your final answer?"

- People who know what happened may not be around

❖ "The Weakest Link"?"

➤ Everybody involved in acquiring the data is equally important

❖ Flexible consistency

The Long and Winding Road...



**Data
Generation**

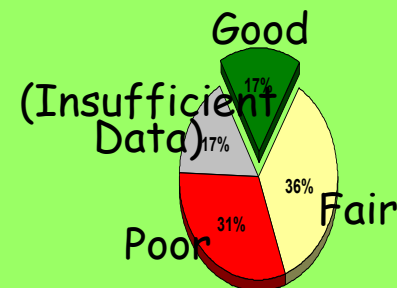
**Data Entry
&
Verification**

**Survey
Analysis**

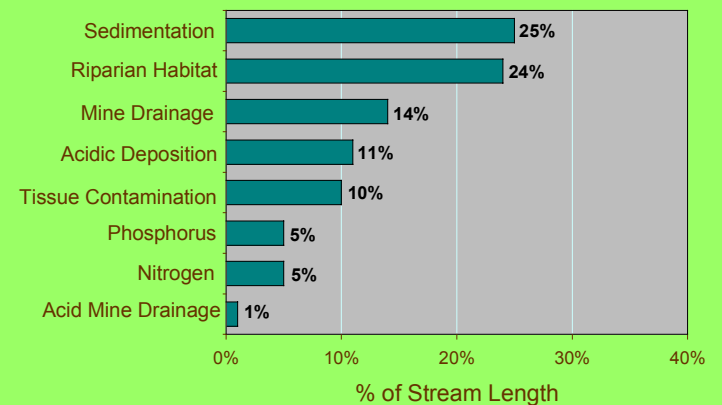
**Metric/Indicator
Development &
Evaluation**

Validation

Interpretation & Assessment



Proportion of Stream Length



EMAP-SW QA Factoids

❖ Documentation

➤ QA Project Plans

- ☞ EMAP-SW "umbrella" QA Plan
- ☞ Chemistry/Fish Tissue laboratory QA plan
- ☞ Benthic laboratory QA plan
- ☞ Periphyton laboratory QA plan

➤ Methods

- ☞ Field Manual
- ☞ Chemistry/Fish Tissue laboratory SOPs
- ☞ Others described to varying detail in QA plans

➤ Metadata

- ☞ Needs for internal analysis and assessments
- ☞ Needs for public distribution of data (STORET, EMAP)

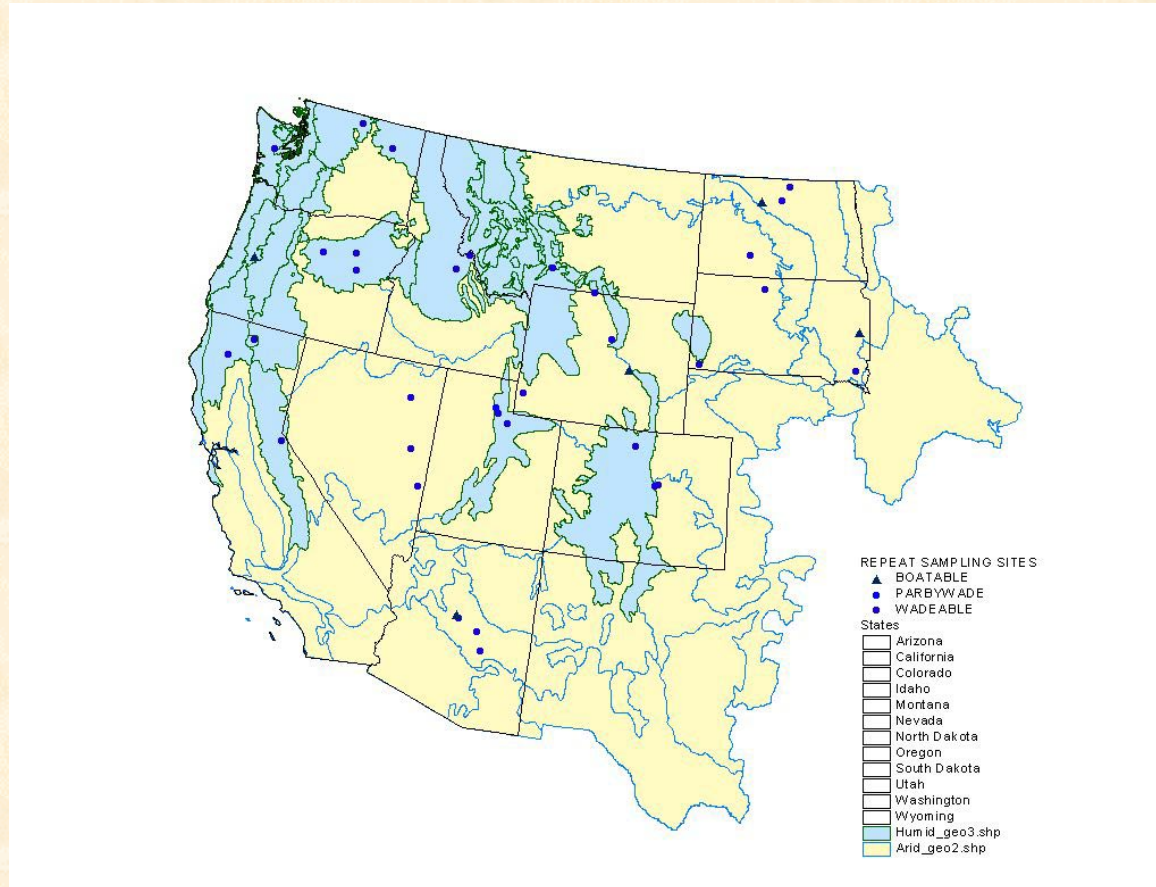
EMAP-SW QA Factoids



❖ General Field QA

- Site visits and debriefings
 - ☞ Every State has been visited at least once (?)
- Standardized, field-friendly data forms
- Good field people

Repeat Visit Sites



- ❖ Within and among-year visits provide data to characterize variance
 - Different components important in status vs. trend estimation

QA for Water Chemistry

❖ Laboratory

- Interlaboratory PE studies (NWRI, NIVA)
 - ☞ Range of samples, lots of labs
 - ☞ Consistently performed well
- Audits
 - ☞ NHEERL
 - ☞ Must be doing something right again...
- QA/QC data summaries

QA for Physical Habitat

- ❖ Once way behind, now probably most advanced in some QA aspects
- ❖ Automated QA checks combined with metric calculating programs

Some Real Data!!

❖ Invasive Plants

- Very difficult to train
- Variable levels of botanical expertise

☞ Need easily identifiable target species

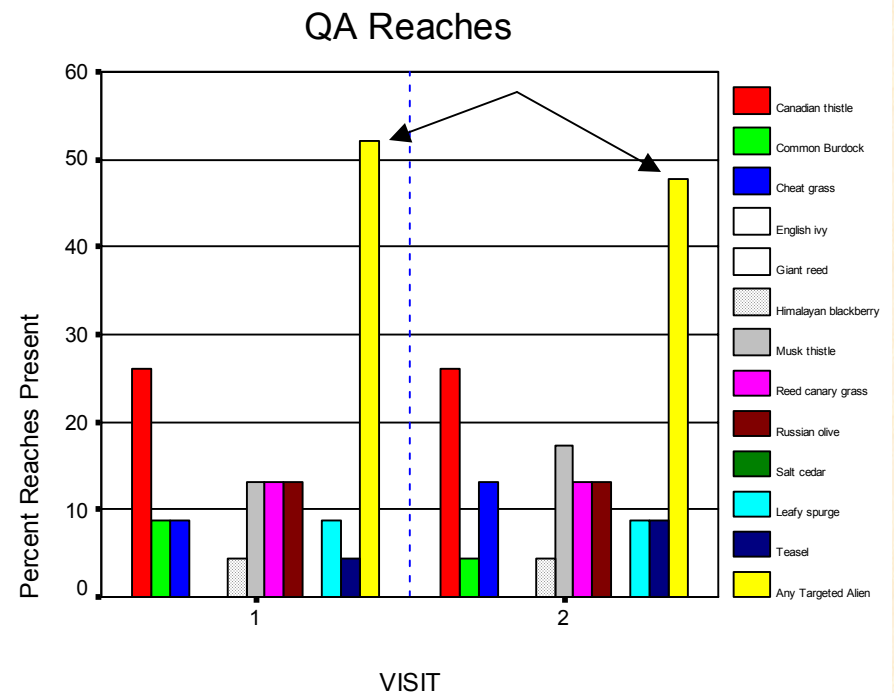
- Field time constraints

- Compare to expert crews in Oregon and Montana

☞ Experts list all species

- Sampling QA
- Ecological QA

☞ Results - very good for Sampling QA

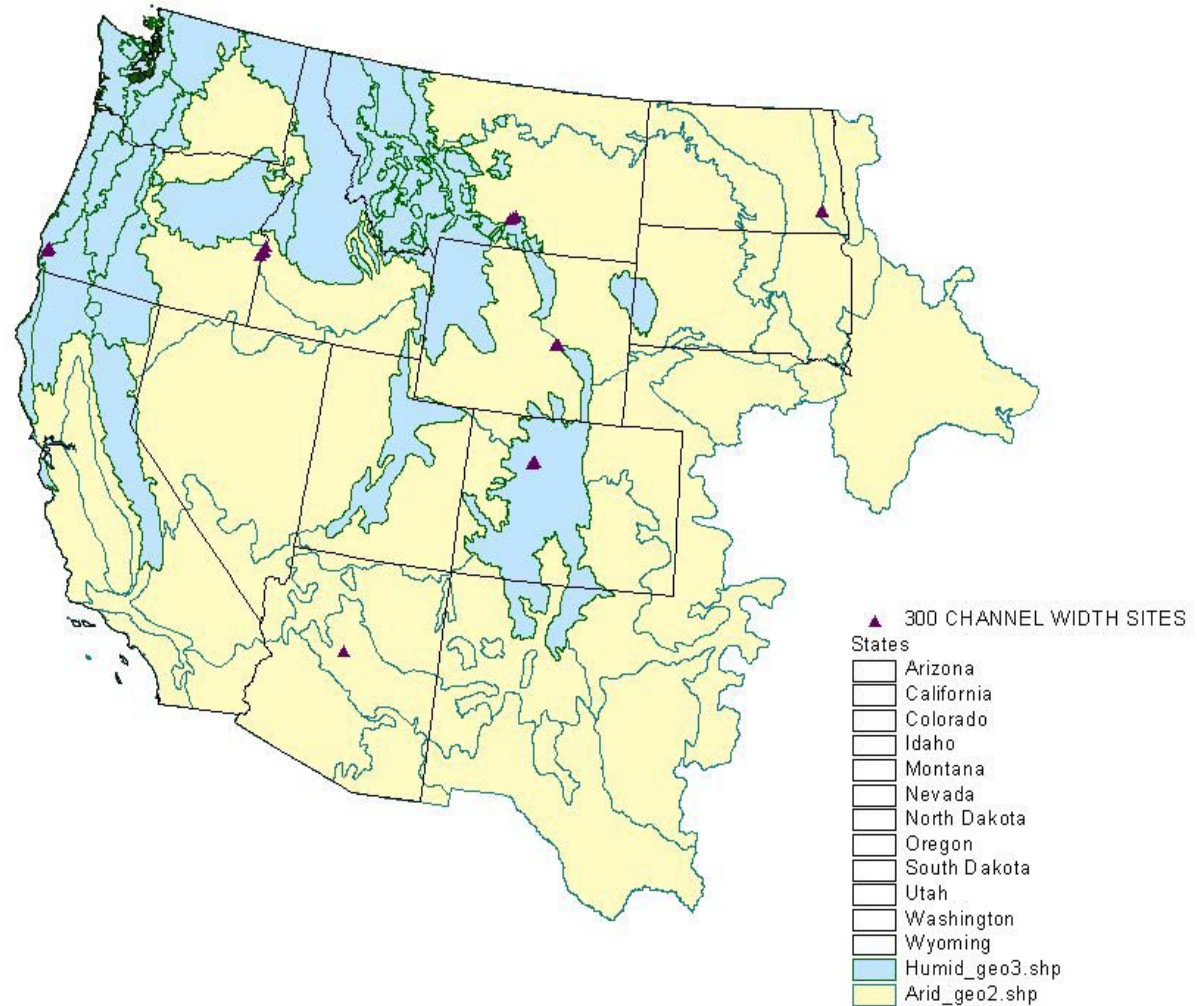


QA For Aquatic Vertebrates

- How long of a stream reach do I have to sample?
 - ☞ Field check on sufficiency using Jaccard index (Cao et al. 2001 CJFAS)
 - ☞ 300 channel width sites (non-wadeable)
- "If it isn't written down (or in the jar), it never happened..."
 - ☞ Hard to verify possible range extensions without vouchers
 - ☞ Hard to take good photos in the field
 - ☞ Hard to adjust counts based on ID confirmations if too few specimens
- Impact of permit restrictions???

300 Channel Width Sites

- Each raft crew does at least one per year
- Sites that are close together or that had long floats anyway



QA for Benthic Macroinvertebrates

- ❖ Standard field protocol ensures we collect sufficient material?
- ❖ Protect material from damage
 - Proper preservation
- ❖ **Keeping track of samples**
- ❖ Laboratory
 - Sorting completeness
 - Internal taxonomic accuracy
 - Complete re-sample of 5% (subsample, sort, ID)
 - QA/QC data summary reports

QA for Benthic Macroinvertebrates

❖ Repeat Visit Samples (subset):

- Second subsample, sort, ID from one field sample of revisit pair

- ☞ Compare accuracy, similarity, precision within sample to among visits

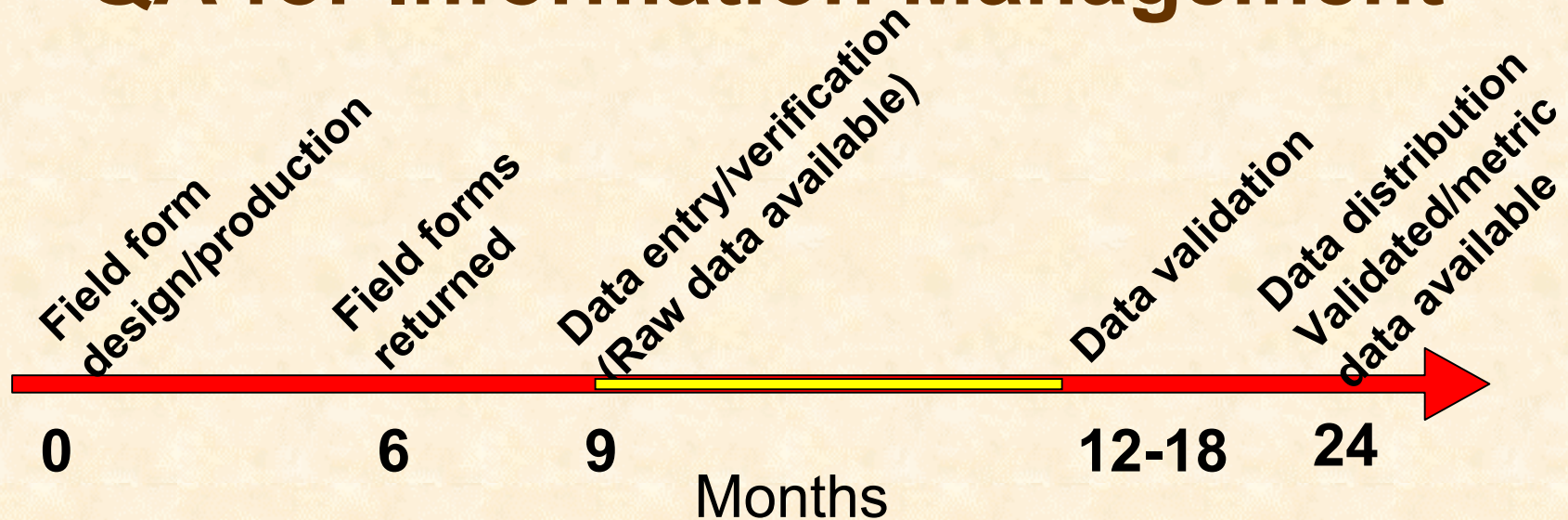
❖ EMAP vs. STAR grant (C. Hawkins)

- Targeted riffle samples
- Allow for comparison of EMAP data to independently sampled "reference sites"

QA for Periphyton

- ❖ Cooperative agreement with Portland State University to do taxonomy and data analysis
- ❖ EMAP vs. STAR grant (Utah State/Mich. State) protocol comparison
 - Extra periphyton ID sample from targeted habitat type
 - Collected in Nevada, S. Dakota in 2002
 - Will allow for comparison of EMAP with independently collected "reference sites"

QA for Information Management



❖ Entry/Verification/Validation

- IM staff interacting with indicator scientists, field coordinators
- The length of this time period is highly dependent on responses from outside labs, or other reviews (e.g. ID specimens).
 - ☞ In most cases we have met or exceeded our goals.
 - ☞ The big unknown is that yellow bar when we are waiting for labs or other data

Data Entry & Verification

❖ Field data forms

- Scan forms
- Software interpretation
 - ☞ Input restrictions
- Operator confirm/review
 - ☞ Visual, scripts flag suspect values
- Output file review
- "Good site"= paper to file in about 15-30 min.

❖ Laboratory data

- Convert to SAS-compatible structure if necessary
- Get data documentation from lab

Data Validation

- ❖ Structure checks
 - All variables are the expected formats and types:
- ❖ Missing value checks
 - Missing values are truly missing (not 0)
 - Deal with missing values if necessary
 - ☞ Substitute alternative but equivalent value
 - ☞ Compute directly from other variables
 - ☞ Estimate indirectly from other variables
- ❖ Illegal value checks (e.g., correct for "Inventive coding")

Data Validation

- ❖ Range checks

- ❖ Logic checks

- Internal consistency

- ☞ Chemistry: Ion balance, Conductivity, Analyte comparisons (e.g., pH/ANC/DIC)

- ☞ Habitat: Width vs. depth

- Biota: "Range extensions"

- ☞ Collected somewhere it is not supposed to occur

Data Validation

❖ Taxonomy, taxonomy, taxonomy

➤ Benthos (Periphyton)

☞ Identify "indistinct" taxa (database issues)

- "Genus sp. A" = "Genus sp. A" (???)
- "Genus spp." ≠ "Genus spp."

☞ Comparability between labs (CDFG, EcoAnalysts)

- Nomenclature and level
- May be some sample trading

➤ Vertebrates

☞ Different references = different names

➤ IM doing link to "standard" (ITIS) taxonomic database to improve consistency and comparability of nomenclature

QA for Indicator Development

❖ Metric/Indicator Development and Evaluation

➤ Entire process could be viewed as QA exercise

➤ Focus of this workshop

☞ Metric Screening

- Range, Variability, Natural Drivers, Redundancy, Responsiveness

☞ Indicator Evaluation

- Reference vs. Disturbance

QA for Survey Analysis

- ❖ Make sure we have enough sites sampled in each study component in each state
 - Heard where we are at this morning
- ❖ "No Site Left Behind..."
 - Need to account for non-sampled sites as well in order to determine weights accurately
 - ☞ "tweeners"- sites evaluated as target, but found to be non-target before sampling visit
 - ☞ Not always a form, easy to forget about

QA for Survey Analysis

Confirming site information

- Was it the right place? (lat/long check)
 - ☞ Provide both DMS and decimal degrees to eliminate conversion errors
- Was it really sampled/not sampled? (sample tracking)
- Which record to keep?
- Make coding consistent

QA Never Ends...

- ❖ Walk in each other's shoes...
- ❖ Start preliminary analyses early with available data
 - Familiar with structure and format of various data files
 - Proper merging...
 - Which sites/records to retain
 - Data nuances that analysis technique requires
- ❖ Even "validated" data aren't perfect
 - Different types of analyses reveal different aspects of data
 - Different people can spot different issues with data