# Shipboard and Sampling Design and Data Quality Control System

Paul Kostovick

Supervisory IT Specialist, Ecosystems Surveys Branch

paul.kostovick@noaa.gov

August 5, 2013

NOAA FISHERIES

Northeast Fisheries Science Center

# Presentation Content

- Overview of Data Collection System
  - Vessel Sensor Acquisition
  - Biological Data Collection
- Data Backup and Recovery
- Data Transfer – Ship-to-Shore
- Data Post-Processing
- Data Dissemination
- Strengths, Challenges, and Proposed Solutions

# Historical Data Collection Methods

- Paper logs 1963 – Fall 2000
  - Transcribed, coded, and sent for digital entry
  - Errors introduced via transcription and data entry
  - Assessment-ready data ~ 4 months

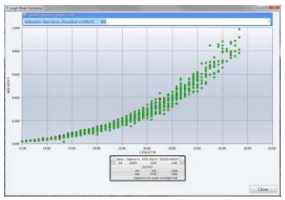
# Historical Data Collection Methods

- Digital data acquisition
  - Fisheries Scientific Computing System (FSCS)
  - Version 1 – Spring 2001 – Spring 2011
    - File-based configuration and data storage
    - Some real-time validation
    - Assessment- ready data ~ 1 – 2 months
  - Version 2 – Fall 2011 – Present
    - Database-centric configuration and data storage
    - Greater flexibility in implementing real-time validation
    - Assessment-ready data ~ 1 month

# Data Collection System Overview

## Vessel Sensor Acquisition - SCS

## Biological Data Acquisition - FSCS



FSCS Database

External Applications

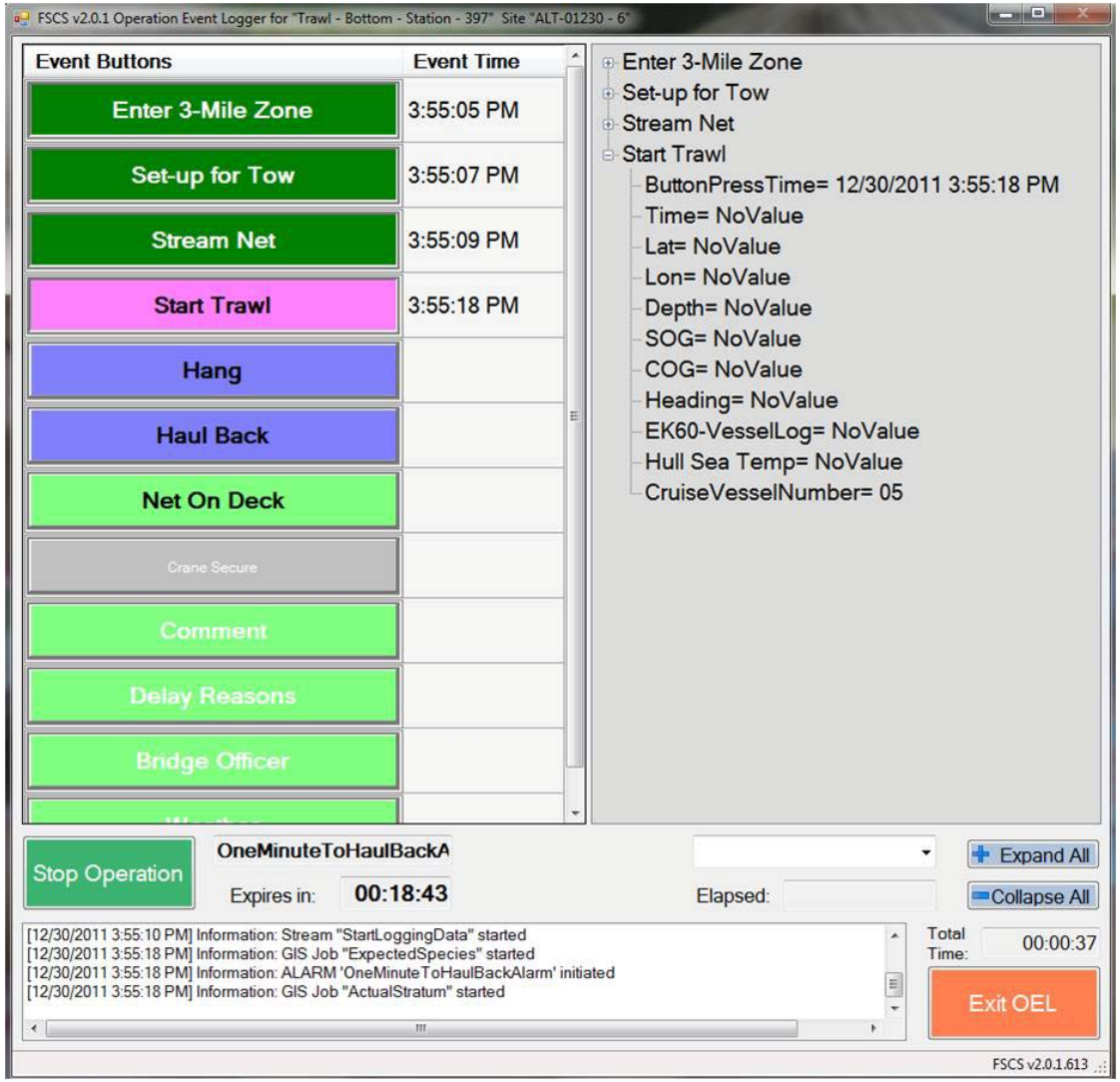
- Configuration
- Validation
- Data Storage
- Adaptive Sampling

# Vessel Sensor Acquisition

FRV Henry B. Bigelow

- Acoustics Doppler Current Profiler (ADCP)
- EK60 Depth -18, 38, and 120 kHz
- ES60 Depth 50kHz
- Redundant GPS Units - MX 420 (2) and POSMV
- POSMV Vessel Heave, Pitch, and Roll
- Port and Starboard Trawl Winch Lineout, Tension, and Speed
- Port and Starboard Trawl Block Lineout and Tension
- Scanmar Trawl Wing spread, Door spread, Height, Net Opening, and Depth
- Young Windbird – Relative and True Wind Speed and Direction

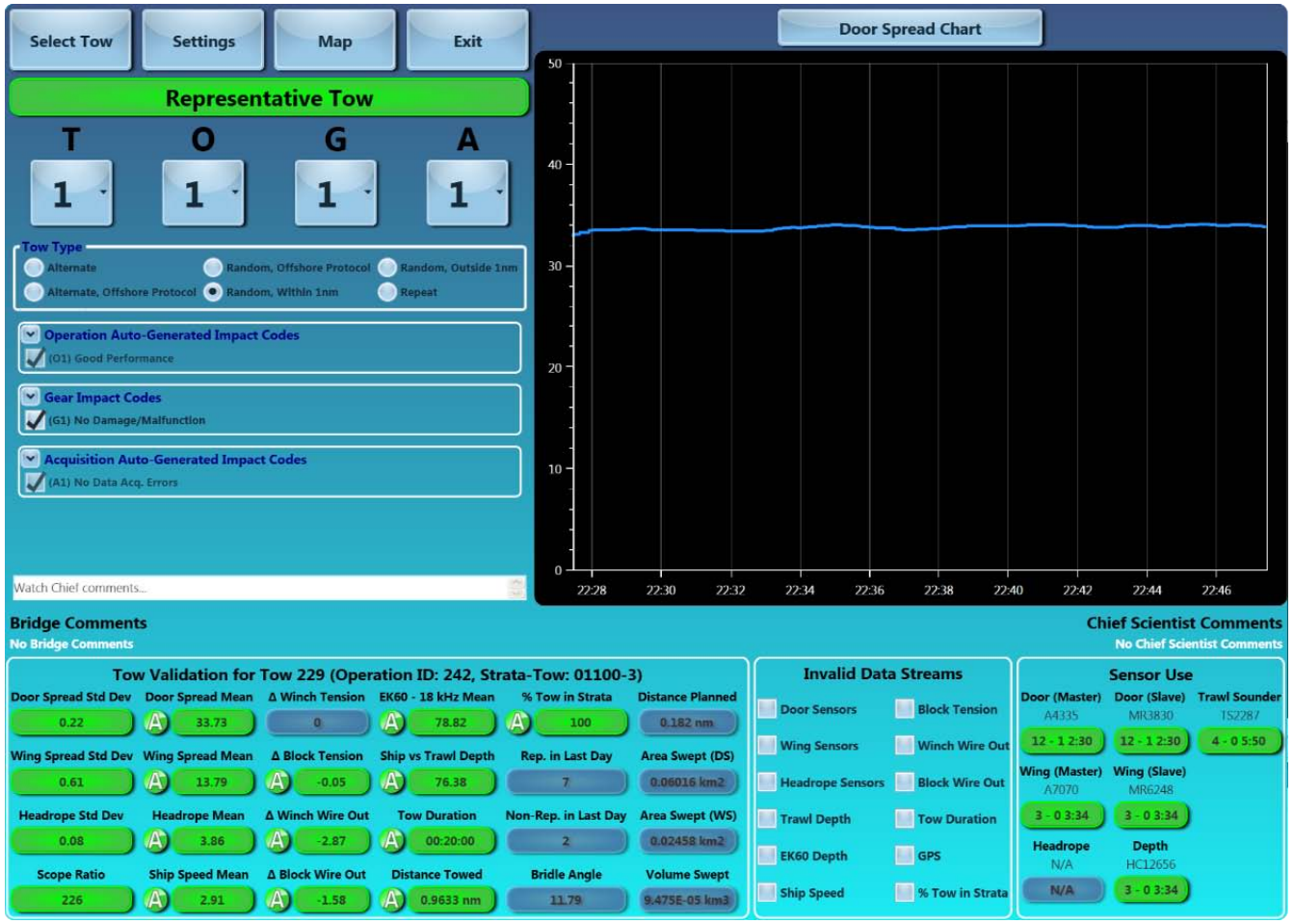
# Operation Event Logger – Sensor Data



# Tow Evaluation

- Standardized method of determining tow validity
- Promotes consistency of trawl geometry and towing procedures
- Utilizes a combination of trawl mensuration systems and vessel sensors
- Tolerance limits and optimal values were calculated from calibration experiments
- Tow evaluation software quickly summarizes data relevant to the evaluation of a standard tow

# Tow Evaluation



# Tow Evaluation – Relevant Data

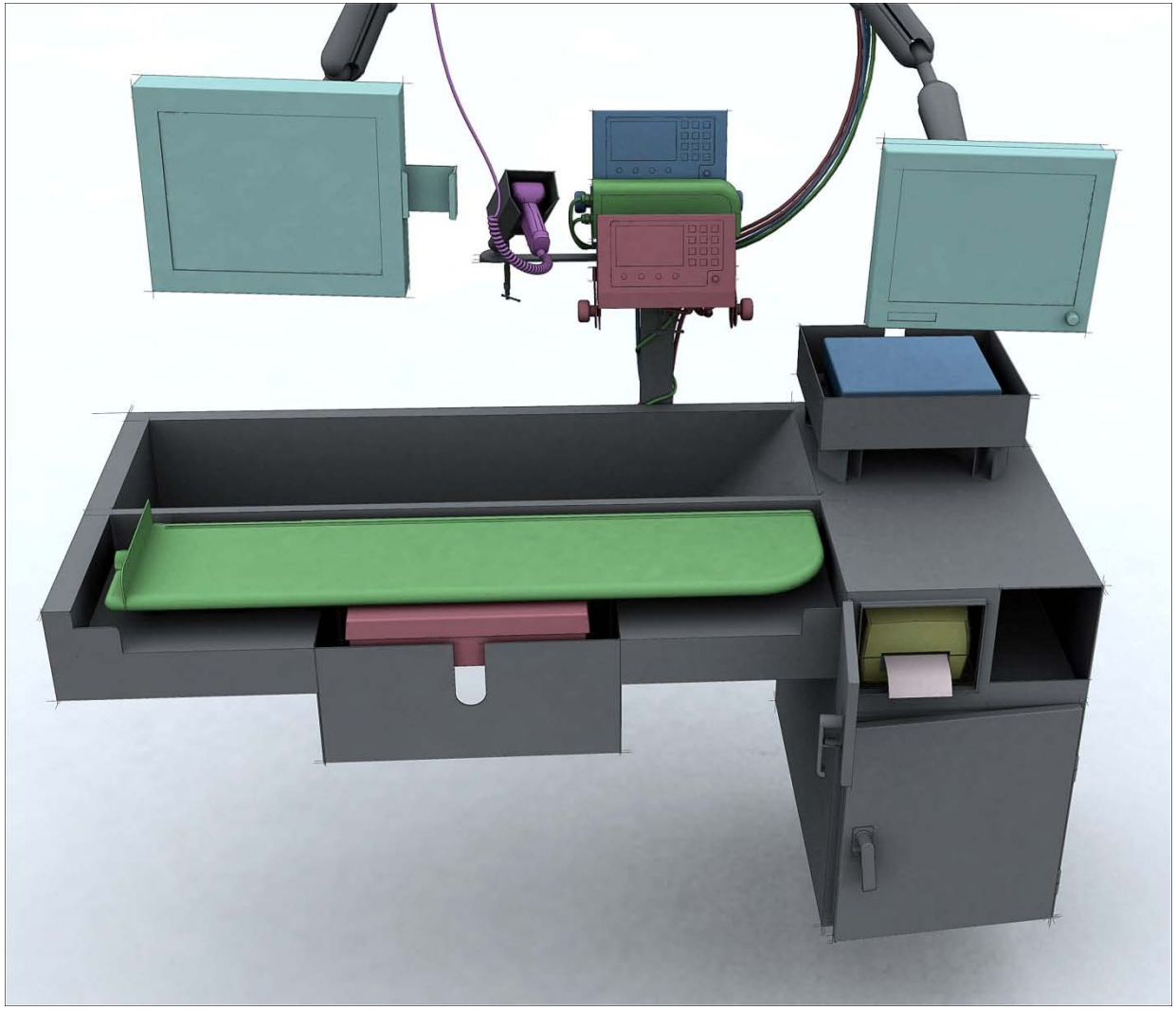
| | Primary Sensor | Data Filtered |
|---|---|---|
| **Trawl Mensuration Data** | | |
| Door Spread | Scanmar Distance | ≤0 & >50m |
| Wing Spread | Scanmar Distance | ≤0 & >30m |
| Headrope Height | Scanmar Trawl Sounder | ≤0 & >10m |
| Trawl Depth | Scanmar Depth | ≤0 & >1000m |
| **Vessel Tow Data** | | |
| Speed Over Ground | MX420 | <0 & >20kts |
| Tow Duration | Derived | n/a |
| Depth | EK60-18kHz | ≤0 & >1000m |
| Tow Location | POS MV | Not null |
| Tow Distance | Derived | n/a |
| **Winch Data - Port and Starboard** | | |
| Winch Wire Out | Rapp Hydema Autotrawl System | ≤0 & >2000m |
| Winch Tension | Rapp Hydema Autotrawl System | ≤0 & >50t |
| Block Wire Out | Block Counter | ≤0 & >2000m |
| Block Tension | Block Load Cell | ≤0 & >50t |

# Biological Data Processing

# Biological Data Collection

- Sampling Design
  - Total biomass – obtained from aggregate species weight
  - Length frequency for all vertebrate and some invertebrate species
  - Age sampling for managed species
  - Feeding ecology
  - Special sampling for internal and external stakeholders

# Biological Data Acquisition Tools - Hardware Suite

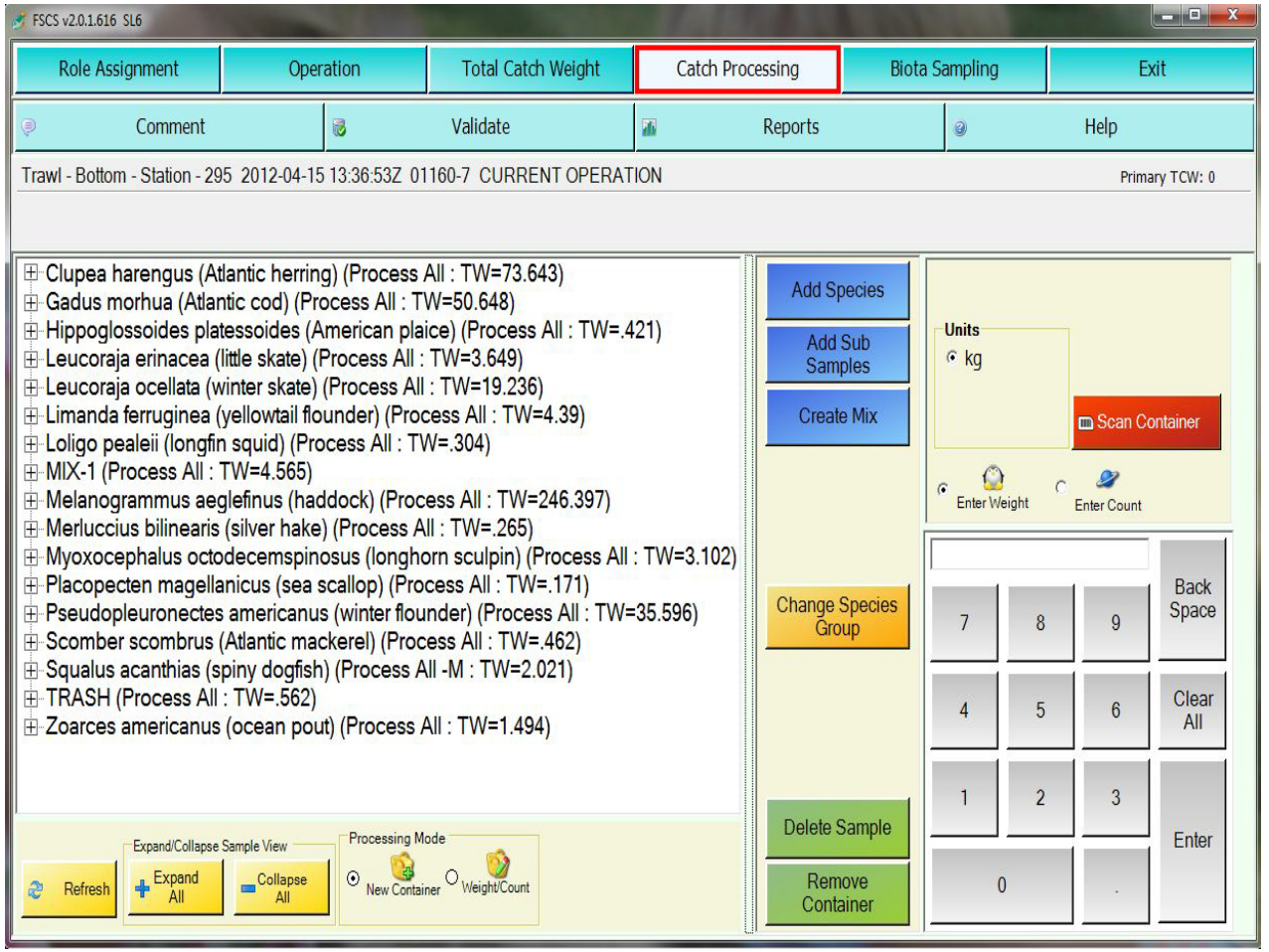


# Biological Data Collection – Sorting the Catch

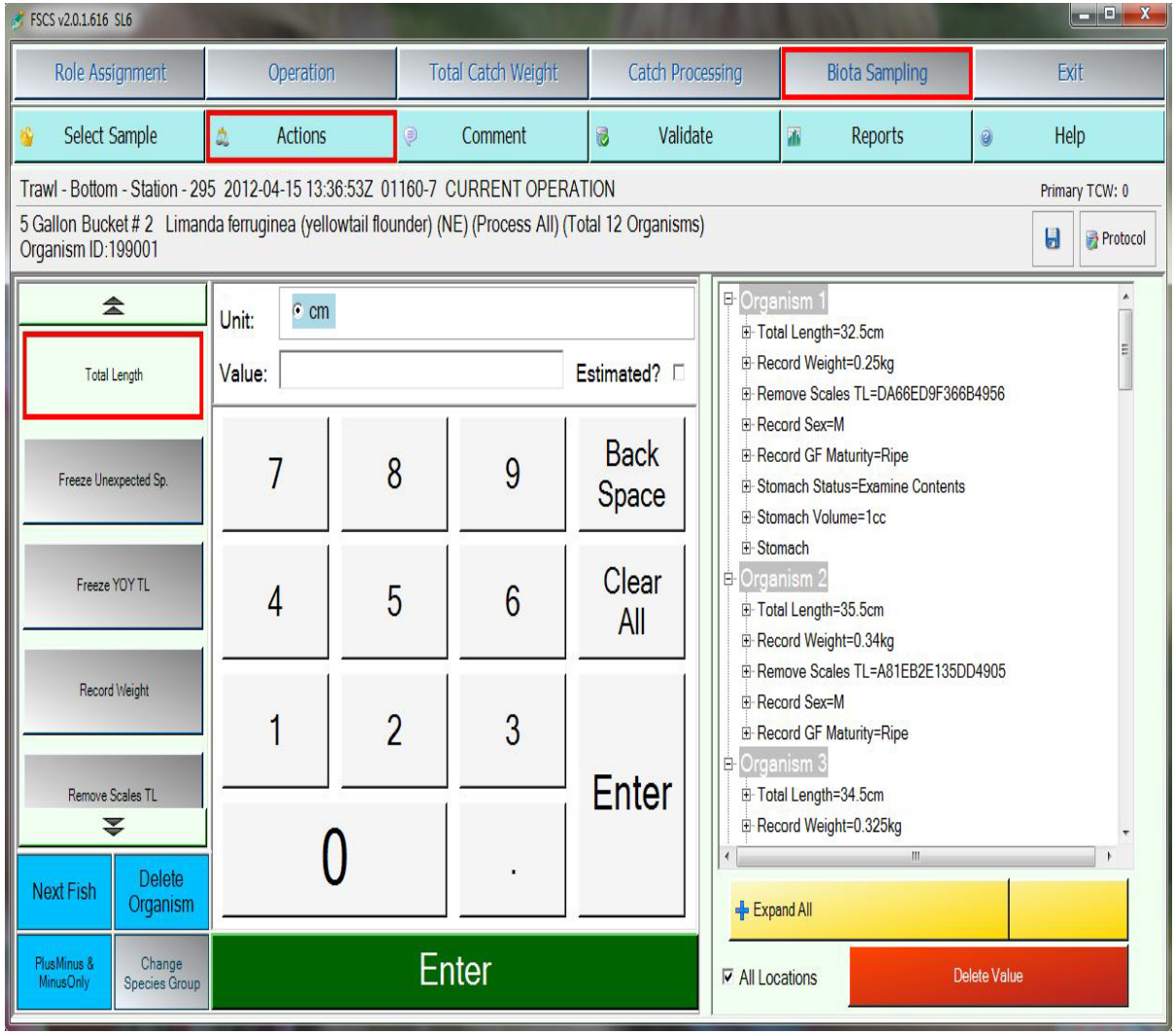
# Biological Data Collection – Sub-sampling

- Discard by weight
- By length
- Size anomaly
- Discard by count
- Mix – discard by weight
- Mix – discard by count
- Remaining weight

# Catch Processing – Aggregate Species Weight



# Biota Processing – Individual Fish Parameters



# Age-based Sampling

- Organism Parameters Collected for Age Analysis
  - Length
  - Weight
  - Sex
  - Stage of gonad maturation
  - Removal of age structure for laboratory processing

# Age-based Sampling Collection Methods

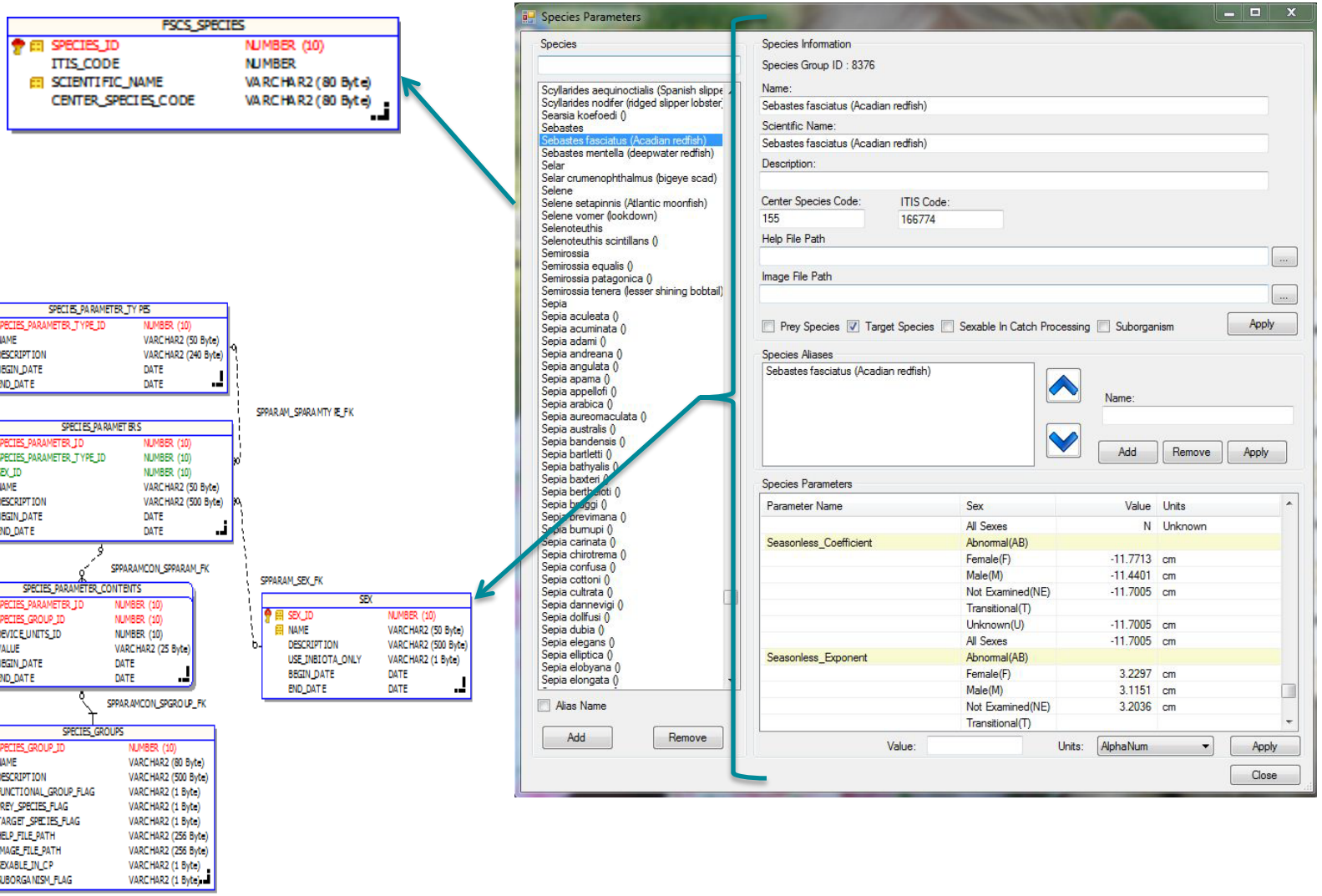
- Analysts provide species tables indicating the target number of samples to be taken by length group for each stock area
  - Small Fish Fewer Ages - Large Fish Many Ages
    - Different Sampling Targets for Different Sized Fish
- Sampling regime repeated at each survey site

| Common Name(Scientific Name) | Range (CM) | Number of Samples/Bin | Scales | Otoliths | Freeze |
|---|---|---|---|---|---|
| Cod, Atlantic (*Gadus morhua*) | 0 – 10<br>11 – 39<br>40+ | 1 per 1<br>1 per 1<br>ALL | | <br>X<br>X | X |

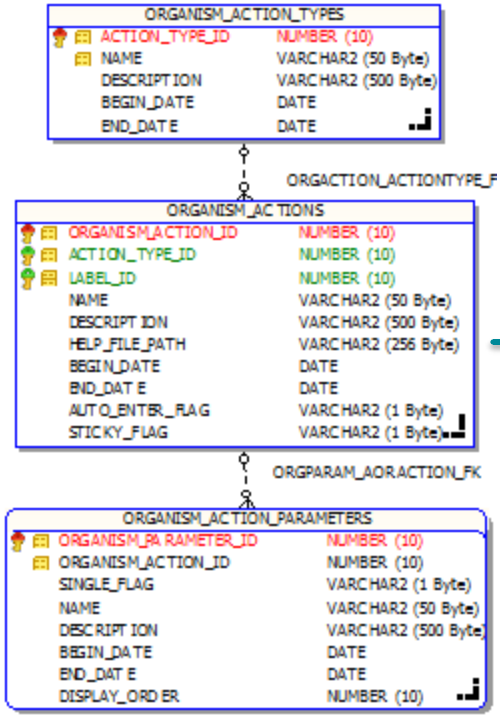
# FSCS 2.0 Benefits

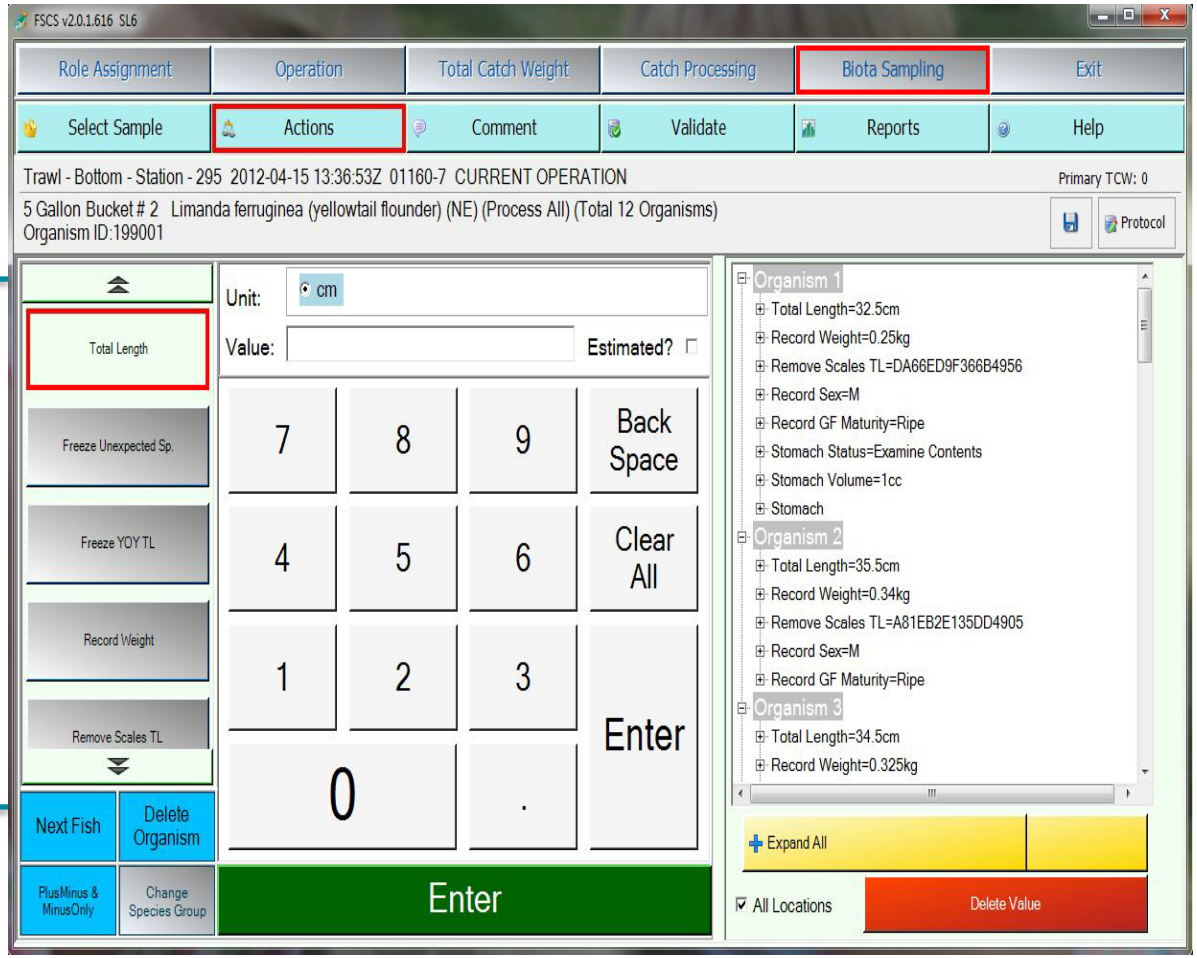
- Flexible survey design
- Better data quality – real-time Q/A
- Increased sampling efficiencies

# Flexible Survey Design - Species



# Flexible Survey Design – Fish Collection Parameters

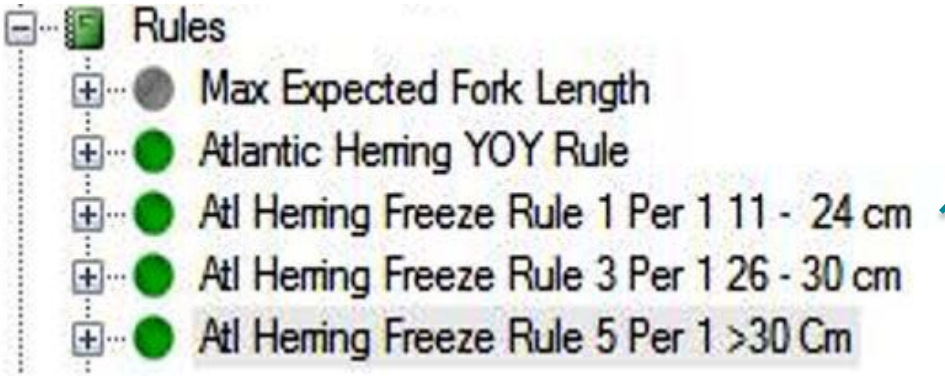


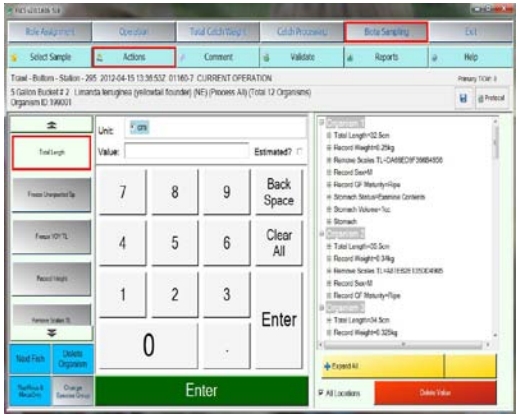



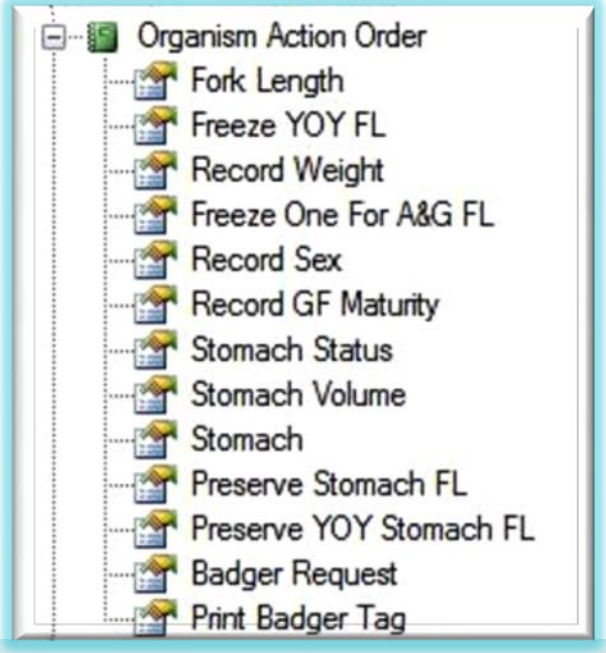
# Flexible Survey Design – Full Customizable Protocol

## Species Protocols – Atlantic Herring Example

### Rules turn on or off organism actions

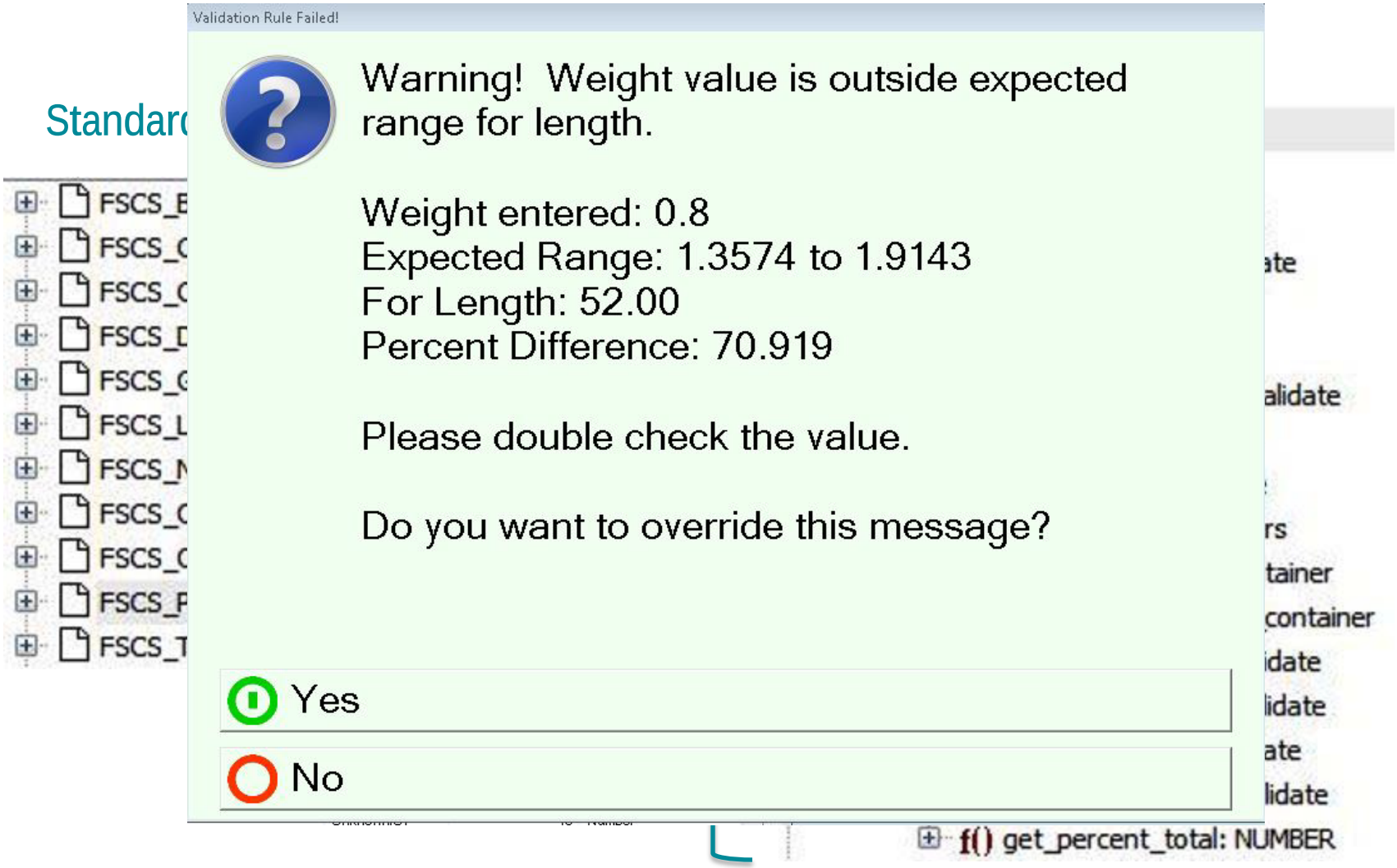


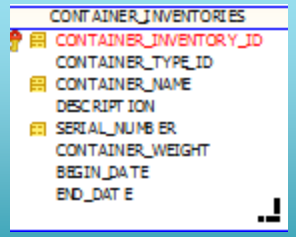

- Rules Prevent Missed Sampling – Human Error
  - Database tells user when to collect parameter
  - Database keeps track of samples taken

# Flexible Survey Design – Custom Validation



# Real Time QA – Container Tracking

Step 1. Build Container Inventory



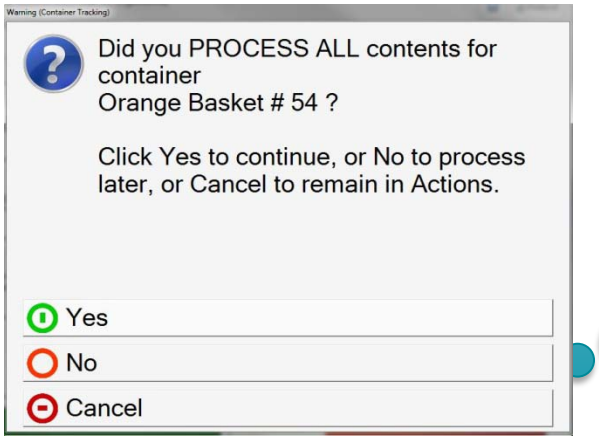

Step 2. Assign Species to Container

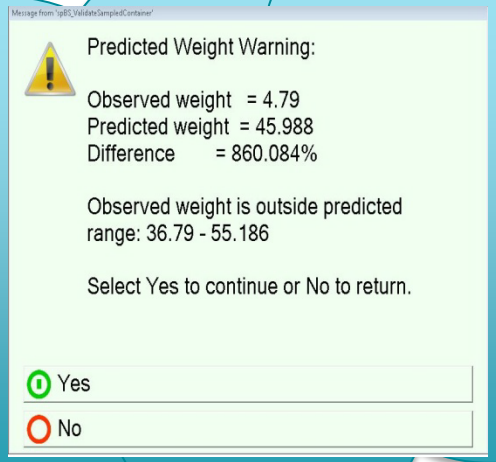
Scomber scombrus (Atlantic mackerel) (Process All : TW=.212)
5 Gallon Bucket # 23 Wt=0.212kg

# Container Tracking

## Step 3. Scan Container to Sample

## Step 4. Complete Container Sampling

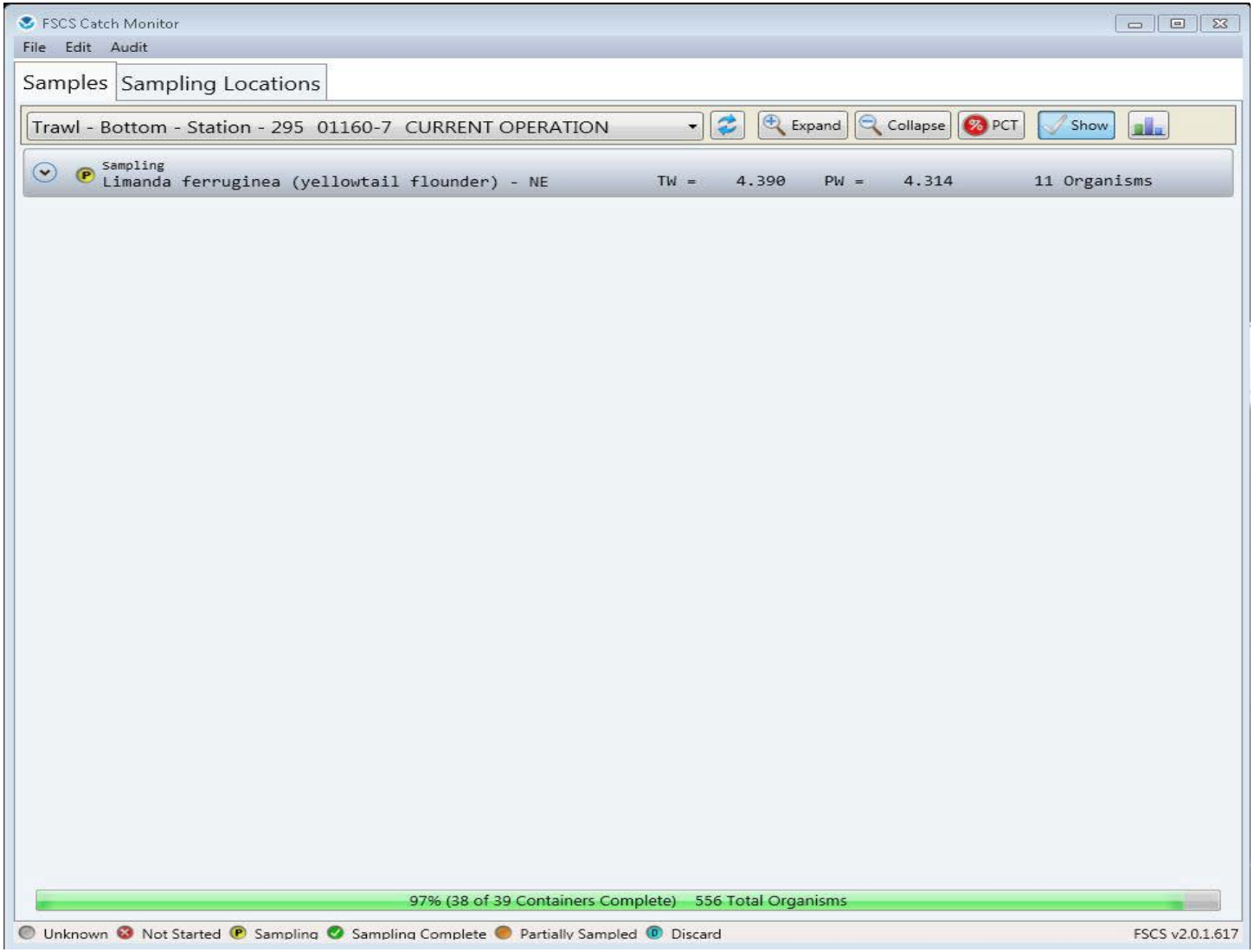


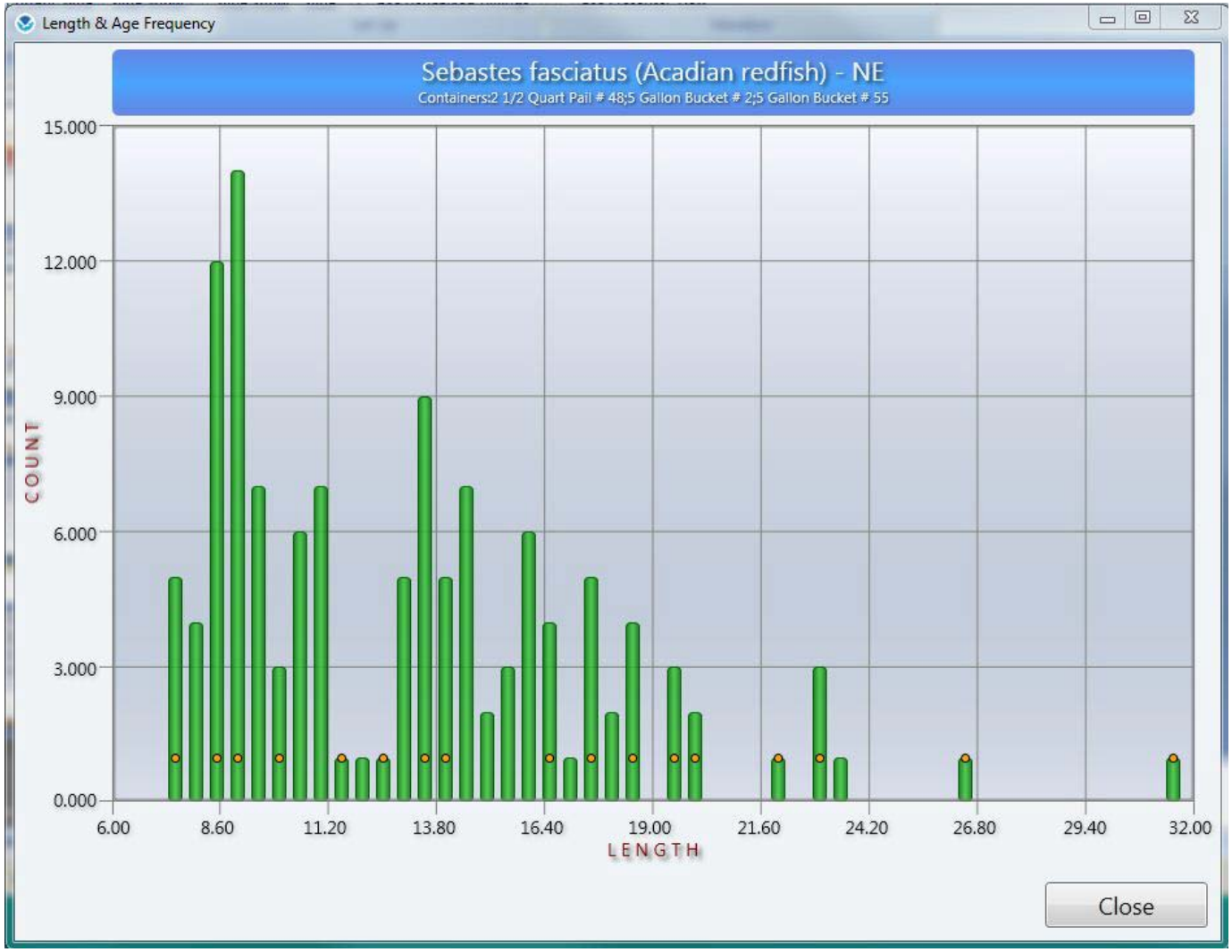
# Better Data Quality – Container Tracking

- Elimination of sampled containers without observed weights
- Selecting an incorrect species significantly reduced
- Observed vs. predicted container weight comparison

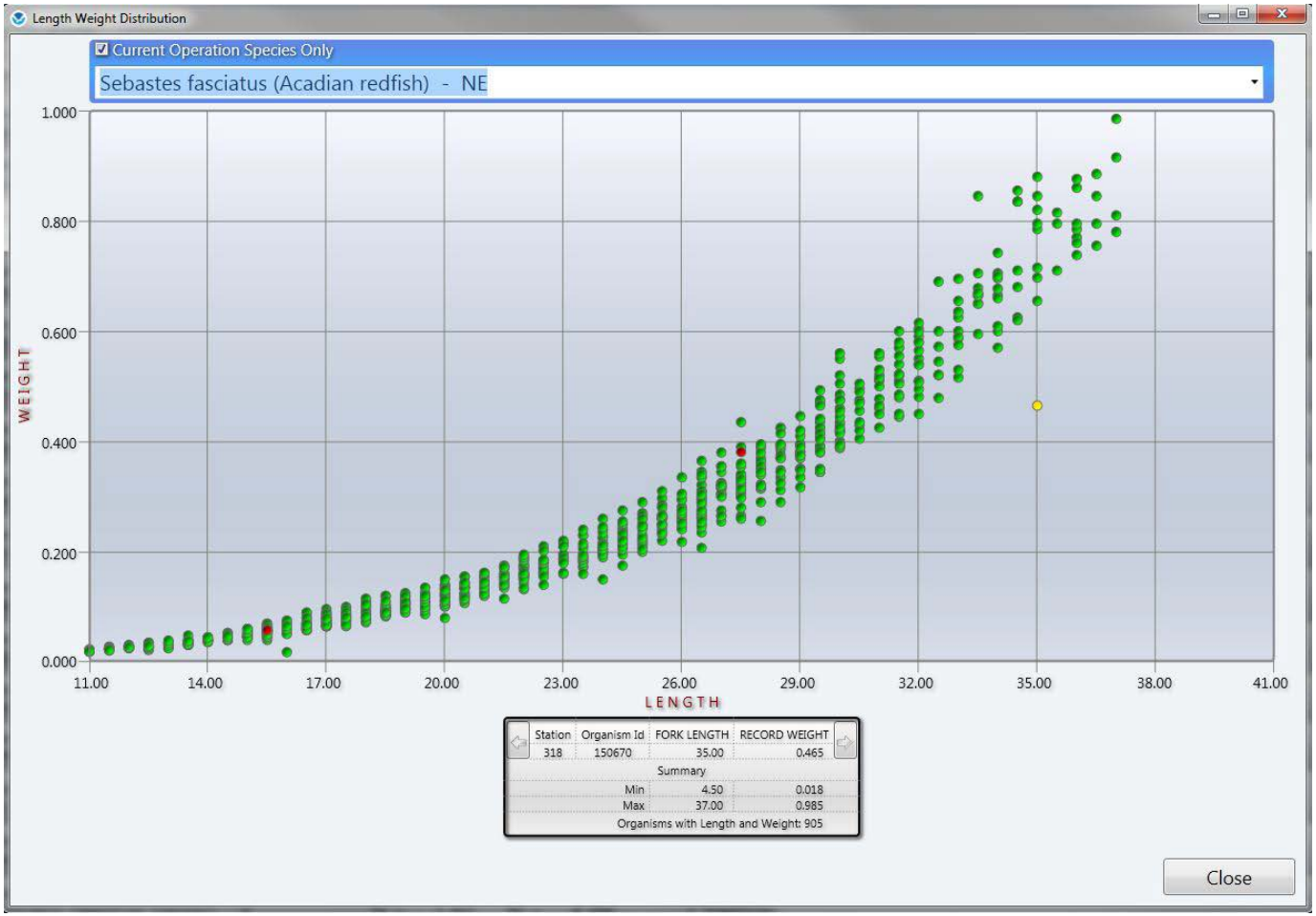
# Better Data Quality – FSCS Monitor



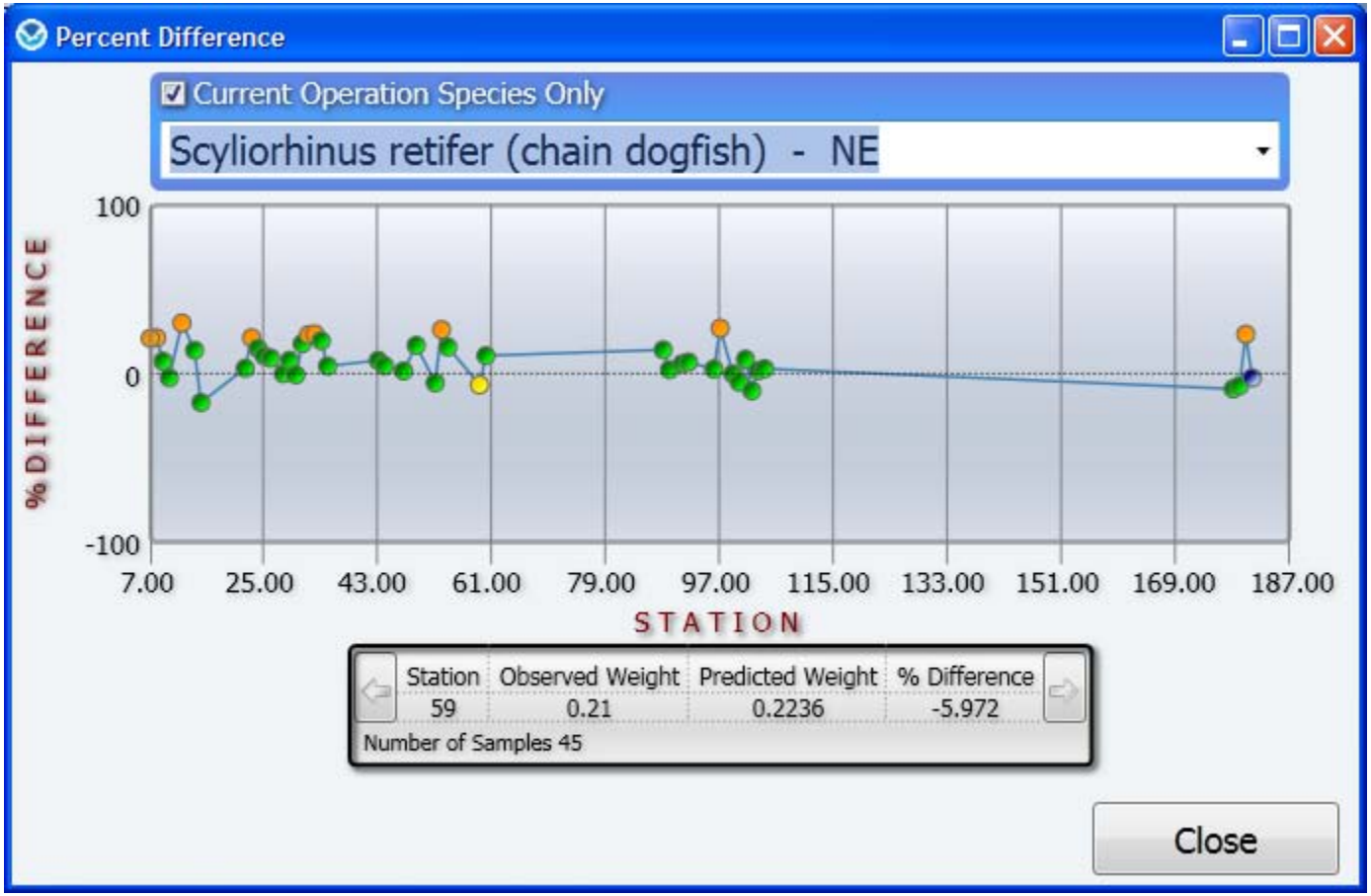
# Better Data Quality – FSCS Monitor



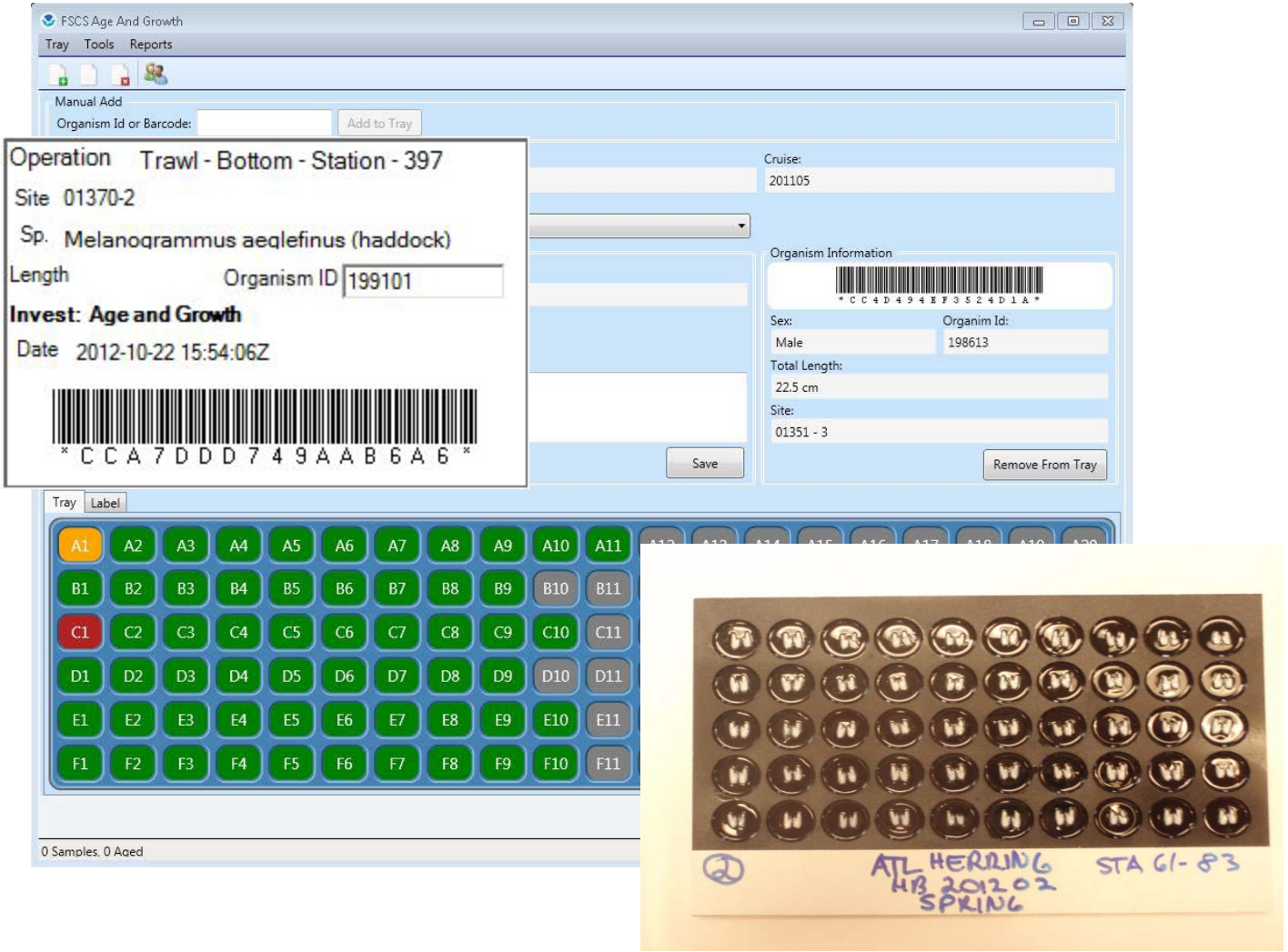
# Better Data Quality – FSCS Monitor



# Better Data Quality – FSCS Monitor



# Better Data Quality – Age and Growth



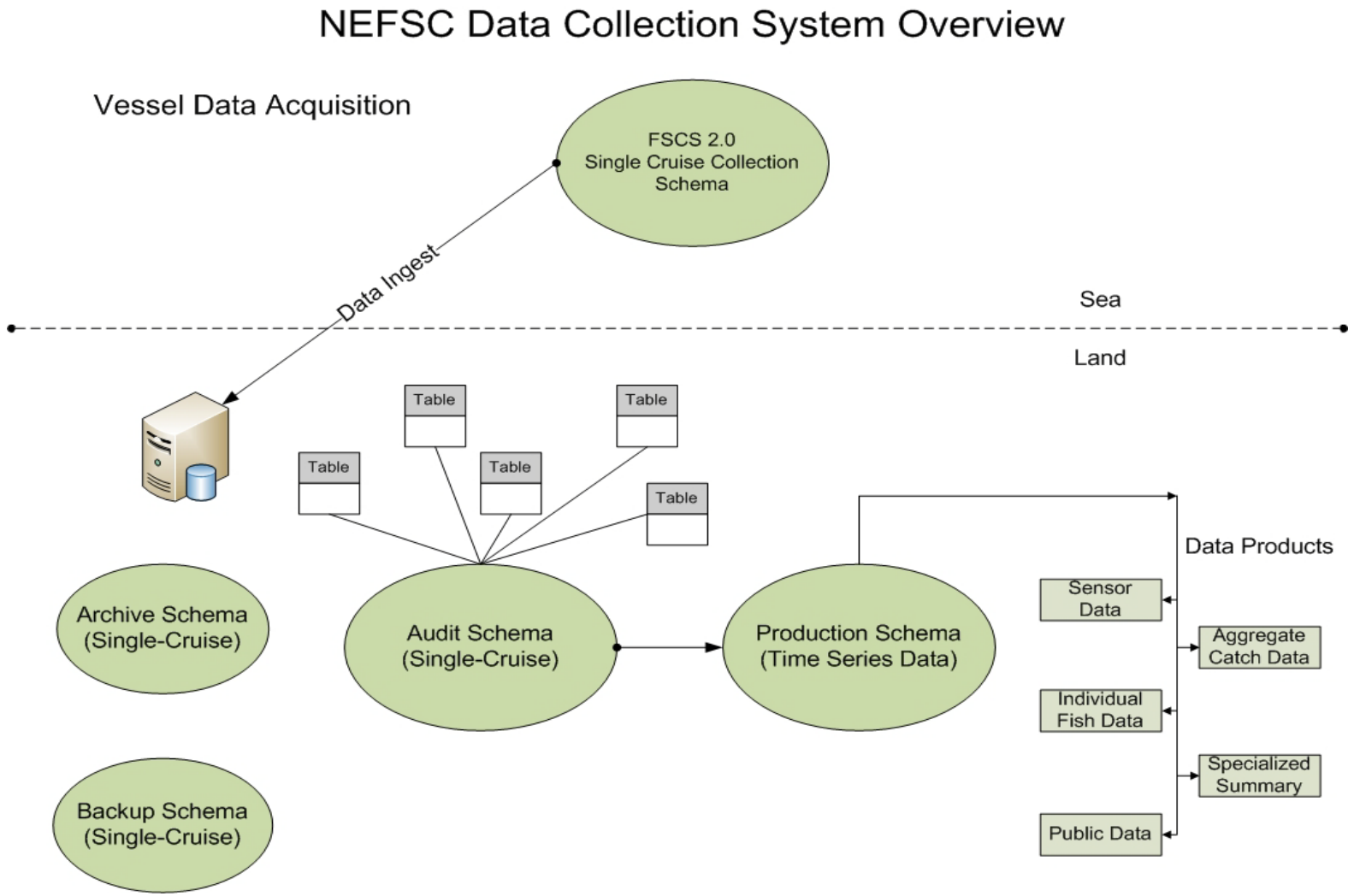
# Increased Sampling Efficiencies

- Rule-based biota sampling
  - Fewer samples while meeting sampling targets
  - Reduces post-cruise processing
    - Age and growth
  - Quicker availability of “clean” data
    - Complete database export available upon demand
    - 10 – 20 days quicker than past surveys
    - Full audited survey (at-sea) achievable goal

# Data Backup and Recovery

- Data storage – Dell PowerEdge R710 server
  - RAID 5 disk array
- Backup server – Dell PowerEdge R710
  - Full and schema backup
- Oracle backup/recovery through RMAN
  - Transaction logging

# Production Database Architecture



# Data Post-Processing

- Archive schema maintained for raw collected data
- Station audit verifying tow evaluation selection criteria
- Biological statistics comparing:
  - Relative difference between overall predicted vs. observed weights
  - Individual length vs. weight
- Errors not verifiable by algorithms addressed
  - Man-made errors
- Journaling implemented for all changes

# Data Dissemination

- Managed species data released to stock assessment scientists
- All other species provided secondarily
- Cruise results
  - Complements sailing orders cruise summary
- Resource Surveys Report
  - Stakeholder report focused on commercially important species
- Special samples and associated data
- Data requests

# Strengths

- Tow evaluation software
  - Quantitative method for standardizing tows
  - Higher confidence in evaluating tows
- FSCS 2.0 data acquisition system
  - Better data quality
  - Intuitive user interface
  - Flexible survey design features
  - Quicker turnaround of survey data
- Communication and feedback from assessment users ensures appropriate biological sampling schemes

# Challenges

- Documentation
  - Historical and current meta-data
  - Standardized operations manuals
  - Up-to-date production database manuals
- Standardized training programs
- Long-term data management plan
- Public data dissemination products

# Proposed Solutions

- Increase IT staff in support of modernization of collection, storage, and dissemination systems
  - Build efficient long-term data storage structures and applications easing usability in stock assessments
  - Establish modern standardized documented quality assurance system
  - Build web-based tools to allow access to public data in order to decrease staff time spent on data requests

# Proposed Solutions

- Continue developing real-time at-sea data validations further reducing data turnaround
- Integrate collection systems, where possible, with outside stakeholders in order to collaborate and share resources (e.g., NEAMAP partners)