

FIS Integrated Reporting Research & Design Project
Regional Fishery Dependent Data Collection

February 29, 2016

initiated by:

Mark Brady
Information Architect
National Marine Fisheries Service

prepared by:

Bryan Stevenson



www.electricedgesystems.com
www.fisheryfacts.com

Table of Contents

1	Background.....	4
1.1	Excerpt from NMFS Project Proposal.....	4
1.1.1	Background	4
1.1.2	History	5
1.1.3	Overview.....	7
1.1.4	Objectives	7
2	High Level Approach and Findings	7
2.1	Initial Regional Visits – Avoiding Assumptions about the Basics	8
2.2	State and Federal Trips & Harvest.....	8
2.3	Overview of Current Regional Data Flow Details	9
2.3.1	Generic Flow of the Data Sets	10
2.3.2	Description of Data Collected on Each Trip	12
2.3.2.1	Logbooks.....	12
2.3.2.2	Dealer Purchases.....	12
2.3.2.3	Observer Reports	13
2.3.2.4	VMS Data.....	13
2.4	Current “Matching” System Architecture Common to All Regions	14
2.4.1	How Matching Typically Occurs	14
2.4.2	Current “Matching” Needs	16
2.4.3	VMS: Why it’s different and how it could be used	17
3	National Conceptual Proposed Solution: A Trip Registry.....	18
3.1	Integrated Reporting vs. Logical Equivalent	19
3.2	Overview of Proposed Solution	21
3.3	Sample Usage for Clarity	22
3.4	Benefits of a Logical Equivalent Trip Registry	23
3.5	Detailed Look at Trip Registry Functionality	24
3.5.1	Overview.....	24
3.5.1.1	General Data Structure Details to be Aware of	27
3.5.1.2	Preserving Data History	27
3.5.1.3	Lookup Tables	28
3.5.2	NMFS Loading of Key Data to Avoid Errors	29
3.5.2.1	Sample NMFS Loaded Tables.....	29
3.5.3	Secure Login	37
3.5.4	User Account Management	38
3.5.5	Trip Scheduling	41
3.5.6	Submitting Trip End Notifications.....	46
3.5.6.1	Alternate Approaches to Determining the End of a Trip.....	47
3.5.7	Trip Schedule Reporting	48
3.5.7.1	Extending this Functionality – Single Trip Report.....	50
3.5.8	Communication and Handling of Trip ID and VRN.....	51
3.6	Impacts to Consider	53
3.7	Methods of Unique Identifier Handling for Paper Reporting	56
3.8	General Steps to a True Integrated Reporting System.....	57
4	Southeast Region.....	58
4.1	Regional Overview	58
4.2	Current Data Flows.....	59
4.3	Implementation Details and Steps Specific to the Region	65
5	West Coast Region	65
5.1	Current Data Flows.....	65
5.2	Implementation Details and Steps Specific to the Region	70

6	Alaska Region.....	70
6.1	Current Data Flows.....	70
6.2	Implementation Details and Steps Specific to the Region	75
7	Pacific Islands Region	75
7.1	Current Data Flows.....	75
7.2	Implementation Details and Steps Specific to the Region	80
8	Greater Atlantic Region.....	80
9	Error Reduction by Design	80
9.1	Avoiding and Limiting Human Input Whenever Possible	81
9.1.1	Form Controls other than Text Fields	81
9.1.1.1	Select Boxes (single and multi-select).....	81
9.1.1.2	Checkboxes.....	82
9.1.1.3	Radio Buttons	82
9.1.1.4	Date pickers.....	82
9.1.1.5	Sliders & Range Selectors	82
9.1.1.6	Auto Complete.....	83
9.2	Quality Assurance Other than via Form Controls	83
9.2.1	Defining Business Rules	83
9.2.1.1	Data Driven Business Rules for Free Form Text Entry Scenarios	84
9.2.1.2	Other uses of Data Driven Business Rules.....	85
9.2.1.3	Automated Data Quality Checks and Outlier Reports	86
9.2.2	Avoid Overloading the Meaning of Data.....	86
9.2.3	Errors – How to Present and Manage Them	87
10	Device and Communications Considerations	88
10.1	Mobile Technology for Unique Identifier Handling.....	88
10.2	Mobile Technology for Data Collection.....	89
10.3	Data Transmission from Sea.....	89
10.4	VMS Alternative for Positional Data	90
11	Conclusion.....	90
12	Appendix A – Terminology & Definitions.....	91
13	Appendix B - Regional Data Flow Diagram File Access	92

1 Background

Understanding the context in which this project was undertaken is key to understanding the details provided in this report. For this reason an excerpt from the project proposal has been placed here instead of in an appendix. The report itself starts in section 2.

1.1 *Excerpt from NMFS Project Proposal*

1.1.1 Background

The NOAA Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, the Marine Mammal Protection Act, the Endangered Species Act, and other legislation to conserve and manage our nation's living marine resources. Resource management decisions are based on data collected about fishing trips, the resulting catch, and bycatch. A considerable amount of this data is reported incorrectly.

Fishing trips are reported on by multiple systems, including a vessel system, a seafood dealer system, a NOAA onboard observer system, and a vessel monitoring system (VMS). The primary cause of reporting errors is the disjoint nature of these systems. Attempting to link a single trip across all systems is complicated by the fact that the very data used to link may be incorrectly reported.

In addition, some reporting systems are still paper based. Of those that are electronic, not all take full advantage of electronic data collection capabilities. Exploitation of such capabilities will drastically reduce data collection errors.

An *Integrated Reporting System* is one that addresses these issues. In particular, it has the following properties:

- Reporting from all sources, for a single fishing trip, is done on a single electronic report.
- The same trip ID coding system is used in all subsystems and a single trip has a single trip ID in all subsystems.
- Each data item is collected from its single most reliable source, rather than being collected from multiple sources for the sake of comparison.
- Error prevention (QA) supersedes error correction (QC). Errors that remain after QA are collected and directed to the appropriate QC team for correction.
- The design is based on the fact that human data entry is the primary source of errors. Therefore:
 - Eliminate human data entry with machine generated data wherever possible.
 - Where human data entry cannot be avoided, constrain choices in the data entry interface.
 - Check validity of human data entry on input and provide immediate meaningful feedback.
 - Facilitate the co-reporting of subsystems. For example, both the vessel and dealer will report on their mutual sale. The system should facilitate error correction around the time of sale.
 - Flag or quarantine remaining errors.
- The system design is the simplest of all alternatives that satisfy the functional requirements.

- The system is all electronic.
- The system utilizes existing infrastructure where appropriate to save costs.

A transition to integrated reporting can be a major undertaking. Therefore, it is more feasible to develop a system based on existing subsystems, which means sacrificing property 1 of an Integrated Reporting System while preserving the functionality of an integrated system. In order to accomplish this we need to add a few properties and define an *Integrated Reporting System –Logical Equivalent*.

Integrated reporting system – logical equivalent additional property:

Trip IDs and important trip ID corollaries are machine generated in the appropriate subsystem and are automatically communicated to other subsystems that will utilize them. See Table 1.

Table 1: Trip Identifiers

Identifier	Generating Subsystem	Notes
Trip ID	VMS	This is the primary trip identifier. A web based proxy may stand in for VMS for vessels without VMS.
Vessel ID	Vessel reporting subsystem	The vessel ID may be embedded in the vessel reporting subsystem and can be provided automatically with any vessel communication.
Dealer ID	Dealer reporting subsystem	The dealer ID may be embedded in the dealer reporting subsystem and can be provided automatically with any dealer communication.
Trip Start and End Dates	VMS or Vessel subsystem	...but not both subsystems

1.1.2 History

NMFS fishing trip reporting systems have evolved over the years. Modules have been added as needed with little regard as to how those modules worked together. Furthermore, the development of fishing trip reporting systems has been based on a number of assumptions, which turned out to be false.

Assumption 1: Fishers and seafood dealers will almost always enter data correctly. In truth, data entry errors are very common. The reasons for this are numerous. They include: typographical errors, apathy, and intentional misreporting.

Assumption 2: Fishers and seafood dealers have unlimited time and attention for reporting. This is more of a hidden assumption than a conscious one. It is obviously false. When selecting reporting items, regulatory bodies think of all those things that seem essential. Every division has a list of desired items. However, reporting takes time. The reporter must find the regulatory reporting instructions, read them, and remember them. They must understand how to properly operate their reporting software and hardware. Then they must take the time to make the necessary observations and report them. They must do all this while carrying out their normal operations. A heavy reporting burden gives many reporters a sense of futility, resulting in inaccurate reports.

Assumption 3: The use of redundant reporting, from distinct data streams is helpful in error checking and enforcement. Data streams originate from entities such as vessels, seafood dealers, NMFS observers, and VMS systems.

Assumption 3 certainly appears to be a good idea on the surface. However, there are a number of problems associated with redundant reporting. Firstly, it contributes considerably to reporting overload as described in Assumption 2. Secondly, many of the inconsistencies observed go unused by law enforcement. Thirdly, when there are inconsistencies, determining which entry is correct is labor intensive or infeasible. Sometimes all entries are incorrect.

A specific type of redundant reporting involves the reporting of weight by species from both the vessel and the dealer. One supposed advantage of this redundancy is that it will prevent collusion between vessel and dealer. However, weight values will almost always differ because the vessel is making a rough estimate by volume and the catch is often unsorted, whereas the dealer is sorting the catch and utilizing a scale. Such differences do not indicate collusion. If a vessel and dealer *do* wish to collude, they can do so by verbal agreement. There are more effective means for prevention of intentional underreporting.

The real reason that we need weight reporting by both the vessel and dealer is that only the vessel has information regarding which fish are associated with a particular gear and area. Only the dealer has accurate weights. When properly used, such reporting is not actually redundant.

Assumption 4: Reporters will be able to coordinate their efforts without assistance. It seems reasonable to assume that if a Vessel X declares that they will sell to Dealer A that they will indeed sell to Dealer A. We now know that Vessel X reporting a sale to Dealer A and then actually selling to Dealer B is one of the most common reporting errors. This generates two orphans, or non-matching reports, one from X and one from B. If there is a quality control group for the region they will consider two possible causes. One is that there were two trips, each lacking one of the two required reports. The other possibility is that there was only one trip but one of the reports was in error. Determining which of these actually occurred is labor intensive or infeasible.

Assumption 5: Matching trips across distinct data streams is straight forward. Trip matching is important because analysts almost always use data at the trip level from multiple data streams. Analyses that depend on a single data stream are usually a shortcut or a last resort.

Assumption 5 is derived from Assumption 1. Matching is based on certain data fields and if those data fields are correct and properly designed then matching is indeed straight forward. The fields used for matching are the trip ID, trip start and end dates, vessel ID, and dealer ID. This data is often erroneous, making Assumption 5 false.

Compounding the problem is the fact that critical data fields are sometimes not properly designed. Inexplicably, separate streams sometimes utilize distinct trip ID codes, making them useless for matching. In other cases, an identifier is used to designate a sub-trip ID in one stream and a full trip in other streams.

Origins of Integrated Reporting

In spite of these false assumptions, a handful of analysts across NMFS have long recognized the need for integrated reporting. In 1999, this need was translated into a system design by Aris Corporation. The system was for the Northwest Region and was called the Electronic Fish Catch Logbook. The Electronic Fish Catch Logbook had many of the characteristics of an Integrated Reporting System. The Electronic Fish Catch Logbook was never put into production for reasons unrelated to its feasibility or merit.

1.1.3 Overview

This project is crucial to the success of electronic reporting efforts that are starting throughout NFMS because the potential of electronic reporting systems can only be realized when they are integrated.

1.1.4 Objectives

The objective of this project is to produce a design of an Integrated Reporting system. The system will provide more accurate and timely data than that provided by matching of trips after trip completion. The design shall be based on integration principles and interviews with reporting experts in each of the regions. The design will provide specific adaptations for each region and will build upon existing systems.

The resulting design will be provided to the regions for use in incremental modernization of their reporting systems.

2 High Level Approach and Findings

First we'll start with some definitions/terminology that will be used throughout this report – please review Appendix A before proceeding or you run the risk of misinterpreting what is being discussed.

As the focus of the project was to look for ways to alter NMFS regional systems to allow for Integrated Reporting, here are some things that Integrated Reporting is and is not (including scope of what is being proposed in this report).

Integrated Reporting is:

- The collection of fishery dependent data streams tagged with a unique identifier for the trip it was collected for
- Intended to be implemented at the regional level and not as part of a national centralized system

Integrated Reporting is NOT:

- Just the storage of fishery dependent data using a unique trip identifier that relates to the trip the data was collected for. Although this is a result of Integrated Reporting, by itself it is not. The data must be collected as stated above.
- In any way the same sort of thing as Fisheries One Stop Shop (FOSS) or the National Permit System (NPS).
- Related to FOSS or NPS in any way. FOSS and NPS are the result of national development efforts and are used on a national level.

There were multiple details asked to be investigated in this project with the key item being the current efforts to collect and “match” fishery dependent data in each region. The other details speak to good system design principles including error reducing protocols (including screen mockups to demonstrate), communications options, and input devices.

Commercial trips in their current form were the focus. This means no pilot projects, research trips, or recreational/charter trips were considered.

The investigation started with a visit to each region to assess the current state of affairs.

2.1 Initial Regional Visits – Avoiding Assumptions about the Basics

All regions were visited for a face-to-face meeting with people designated for these introductory meetings (generally experts in one or more data sets of interest and/or management). The meetings served as a formal start to the analysis efforts in each region and were designed to pull out the broad details about the fisheries that are being managed and how data is collected in those fisheries. This was to avoid assumptions and proved a very useful step as there were indeed many instances where they turned out not to be true.

2.2 State and Federal Trips & Harvest

It is important to understand that data needed to produce stock assessments encompasses two main sources of information about how much seafood was harvested over a time period.

Our definitions of a “state” and a “federal” trip are as follows:

1. **State Trip** – a trip that typically requires a state permit (or does not require a federal permit) and occurs inshore (typically 0-3 miles offshore but not always). These trips may catch federally managed fish and there are often special reporting requirements of state dealers to record additional information about these species (or report them separately) when they purchase them.
 - a. More often than not the only source of federally managed species from states comes from state dealer purchases. State trips do not have observers, do not require VMS, and often do not involve logbooks.

2. **Federal Trip** – A trip that typically requires a federal permit and is targeting one or more federally managed species.
 - a. Federal trips can require an observer(s) to be onboard, may require VMS usage, and often require logbooks.

NMFS typically has a fair amount more information about commercial trips than the states do, but the states often have information NMFS needs for the bigger picture of a stock assessment (at the core it is about how much of each species were harvested – including discards). NMFS works in partnership with the states through the FINs and counsels to try and keep data collection of this important data as consistent as possible.

The main challenge that NMFS faces is relying on the states for data about federally managed species harvested on state licensed trips while not having full control of what and how states collect that data. In some regions this is very challenging. It seems to be more challenging as the number of states in the region increases. This makes sense given that it is often more difficult to find consensus among larger groups because of a greater chance of having competing needs among its members.

2.3 Overview of Current Regional Data Flow Details

The information asked for focused on how the data in each data set are collected for the federally managed species in each region. The details focused on the data collection steps starting from the beginning of the trip and concluding with the data being “finalized” or otherwise considered official (all QC efforts performed and any issues resolved). The timing between each step was determined (often in the form of a range of time once getting to QC effort portion of the flows) as well as the broad strokes of what occurs at each step or what factors may effect that step and the amount of time it takes to complete. Details about data storage were also asked for and generally include the database type, version, schema (if applicable), and common name as well as the server name, and server physical location.

Once this core information was determined about each fishery, each region was asked which of these data sets (for each fishery) they need to know are from the same trip (i.e. matched) and if that match was stored once known. In addition, the regions were asked if they have formal applications designed to match the various data sets back to the trip or if they used a process as opposed to an application.

There are differing paths that the same type of data takes in each region. Although these details are very important and speak to the reality of the implementation of the solution proposed in this document, the issues seen in terms of matching the data back to the trip are conceptually the same. For this reason, the overarching conceptual issues, proposed solution, and system design issues to consider will be discussed at a national level with further regional specific portions added for each region where assessment was possible.

As an aside, this core information turned out to be much more difficult to obtain from the regions than assumed meaning that some finer grained system details could not be looked into. In the end this turned out not to be a significant issue because of the nature of the system issues discovered. The category of issues can be discussed at a conceptual level making these finer details less important in terms of this analysis and proposed solution.

There were some communication challenges that arose after the initial visit to the Greater Atlantic that made a full analysis (compared to other regions) not feasible at this time. The good news is that it is known the region is already undertaking an effort to integrate fishery dependent data systems in a way that is in line with the ideal solution being put forward in this report.

The current state regional data flows for each of the four data sets of interest are in each region specific section below.

2.3.1 Generic Flow of the Data Sets

Logbook, dealer purchase, and observer data all follow essentially the same steps from collection to data finalization. VMS data can also follow this generic flow for the non-GPS track data that it is sometimes used to submit. The GPS track flow of VMS data is identical in each region other than some regional regulations that dictate whether the VMS unit needs to be on 24/7 or not. Observer data does tend to have extra and more involved QC steps than the other sets, but that is because of the volume and scientific nature of the data they collect.

The following diagram shows a very simplified look at typical collection of fishery dependent data in all regions:

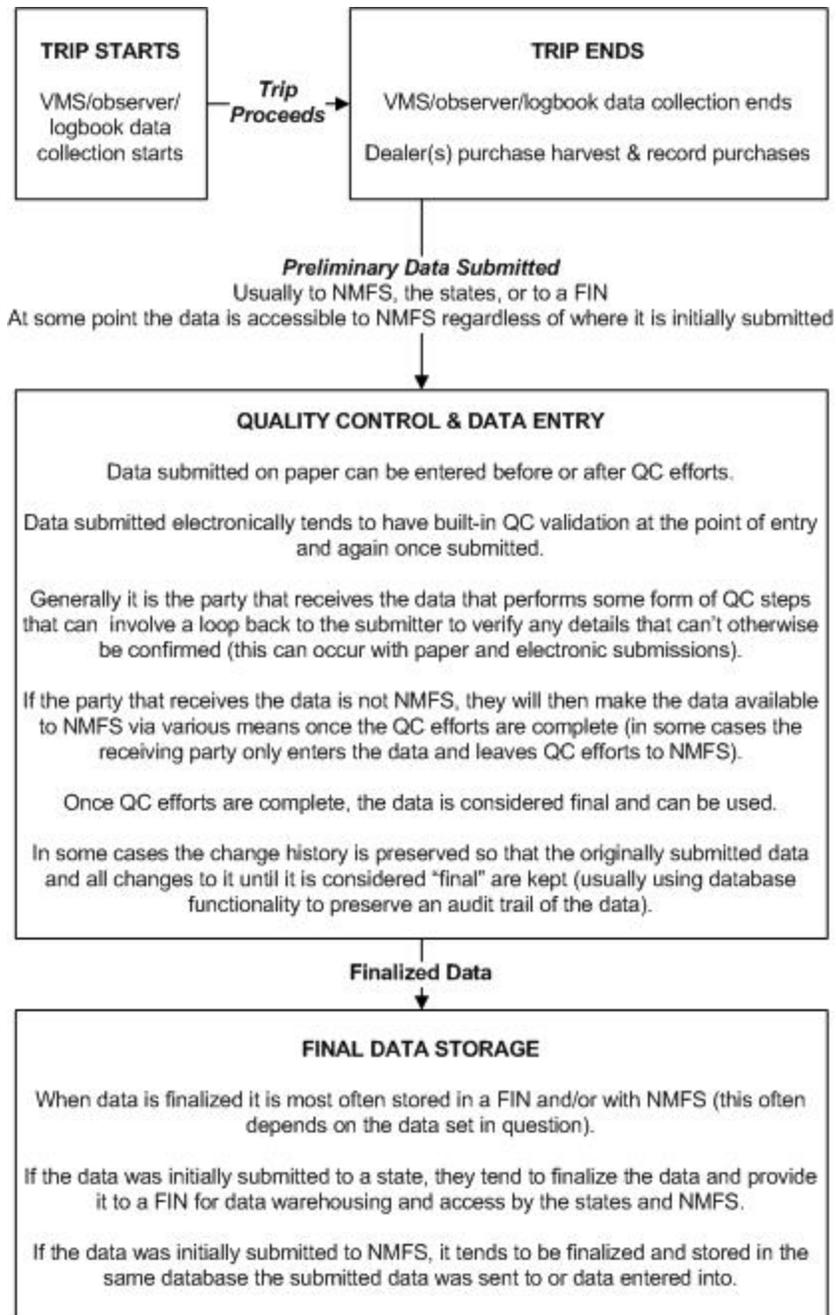


Figure 1: Generic Flow of Regional Data Collection

Below are some additional information about the generic flow of logbook, dealer purchase, observer, and VMS data:

1. Data is collected on paper or electronically
2. When all data has been collected, it is submitted either directly after completion or in batches for all trips dealt with at the end of every week/month/etc.

3. Data entry (if collected on paper) and QC steps can occur at the same time or one before the other (either being first/last).
4. QC steps are usually performed on data collected on paper or electronically. When collected electronically it tends to be validated as it is entered and then re-validated after it is submitted. Data submitted on paper can only be QC'd after it has been submitted and not at the source.
5. If QC steps find issues that cannot be derived/confirmed/explained, then the submitted data is generally sent back to the submitter and they are asked for clarification and to re-submit when done.
6. Once QC passes with no issues, the data is considered final or official and is stored either with NMFS or one of the FINs. Each state generally keeps a copy of any data they have collected and/or received, but once finalized that data goes to a FIN as the FINs are used as data warehouses that consolidate fisheries data. If other agencies such as NMFS need state data, they almost always get it from a FIN as opposed to the individual states in the region.

2.3.2 Description of Data Collected on Each Trip

As mentioned, there was not time to delve into full lists of data elements collected for each kind of commercial trip in each region. These details do not alter the proposed solution put forward. This is because the issue is really about how the data sets are linked and not about the data contained in the sets. So instead these details will be discussed in broad terms here as a general reference.

2.3.2.1 Logbooks

Logbook data varies from fishery to fishery and from region to region, but all share some core details. In addition to the core details, there can be details about target species, region specific regulatory information, as well as trip expense details asked for.

Core logbook data typically collected include:

1. VRN
2. Operator (name and/or ID # of some kind)
3. Sail date/time
4. End date/time (not always)
5. License/permit number(s)
6. Gear type
7. Gear measurements or other details
8. Harvest species
9. Harvest area (often broad and sometimes finer grained)
10. Harvest quantity
11. Harvest units
12. Harvest catch disposition (kept, discarded, etc.)

2.3.2.2 Dealer Purchases

Dealer purchase (aka fish tickets or landings) reports vary from region to region, but tend to be the same or very similar between fisheries in the same region. As with logbooks,

there is a fairly standard core set of details collected (especially within the data of interest to NMFS), but can differ usually from state to state and between state permitted and federal permitted dealers (if federal permitted dealers exist in the region).

Core dealer data typically collected include:

- Dealer permit number
- VRN the purchase was made from
- Date/time of purchase
- Species purchased (often broken down by grade and form strata)
- Weight of each species purchased
- Amount paid for each species (often broken down by grade and form strata)

2.3.2.3 Observer Reports

Observer reports tend to have 2 distinct parts to them – biological samples and non-biological details (often used to cross-check the logbook for the trip) about the harvest. This report focuses on the latter as it speaks to catch accounting (discards as well as the area(s) the harvest was caught). As with other data sets, there are fairly standard core details collected in most regions and fisheries that have observers with some regional or fishery variances. The non-biological harvest data is typically used in catch accounting.

Core non-biological observer data typically collected include:

1. VRN
2. Start/end date/time of the trip
3. Area species were caught in
4. Weight of species caught in each area (sometime a count instead or both)
5. Species may be broken down further by size of grade
6. Harvest disposition

2.3.2.4 VMS Data

When most people think about VMS data they think about a GPS track with direction and speed information. That is by far the main kind of VMS data collected and its collection and usage are virtually identical in all regions and fisheries that require VMS. The other type of VMS data is data that is sent via forms provided by VMS vendors or data generated in other third-party applications and sent using the VMS communication infrastructure (usually by using the e-mail capabilities most VMS vendors provide). This data tends to lean towards hauls or trip declarations, but there are some other kinds of data transmitted as well.

Core GPS track VMS data include:

- VRN
- Position
- Direction
- Speed
- Date/time

Core data often sent through VMS for hails/trip declarations include:

- VRN
- Departure port (usually at start of trip)
- Landing port (usually when harvesting has concluded and returning to shore)
- Trip start date/time (usually at start of trip)
- Trip end date/time (usually when harvesting has concluded and returning to shore)

2.4 Current “Matching” System Architecture Common to All Regions

It became very apparent that each region is using similar system architecture in terms of matching one or more of the four fishery dependent data sets to each other for the same commercial trip. Although the details and steps may vary, conceptually they are the same regardless of using a formal application all the way down to basic ad hoc methods of attempting to match data based on key details (like vessel identifiers and trip sail dates).

In broad terms all regions collect trip level fishery dependent data in the same way. For any one set, the data is collected in isolation of all other data sets. When being collected, there is no linkage back to the trip and this is the crux of the issue that will be discussed in this report.

To get an idea of how long it can take before any matching efforts can begin, here are the four data sets and how long post trip it can take before data is finalized and can be matched to the other data sets for a commercial trip (ranges provided span all regions):

VMS - Available in real-time and is ready for use right away

Observer - Can take from 90 days to a year to be finalized

Dealer Purchases - Can be finalized from 1 day to 3 months or more post trip

Logbooks - Can be finalized within 24 hours and up to a month or more

2.4.1 How Matching Typically Occurs

To determine which data from each set relates back to the same trip, “fuzzy matching” efforts are used. The matching effort generally uses data in common between sets being matched – such as a vessel identifier and the known sail date (and end date if collected) for a given trip and then attempts to find matches in the other data sets based on those details (if linking them is of interest in the region).

For example, these are the broad initial steps to attempt to match a fish ticket with the logbook for the trip the fish ticket is for:

1. First it is important to note that logbooks contain the VRN, the trip sail date, and often a trip end date. Fish tickets also contain a VRN and a purchase date. With this information an attempt at “fuzzy matching” can begin.

2. Using the VRN in the logbook, the fish ticket data is queried to find any matching fish tickets with the earliest purchase date on or after the sail date in the logbook (and before or on the end date if collected in the logbook - or a known subsequent trip sail date for the same vessel if available).
3. If the first attempt to find a match fails, then similar methods of zeroing in on the “most likely” match are used until either A) a match is found, or B) it is determined a match cannot be found (possibly due to a missing fish ticket or incorrect information in the logbook – or both). In the case of outcome B, further manual investigation is required (i.e. maybe to find missing reported data, double check source data quality, or other similar efforts). Each region has its unique data realities and thus has varying other ways to attempt matches.
4. In many regions these matching efforts may be run multiple times over time due to updates in source data to be matched and other reasons. These efforts take staff away from other pressing issues for no other reason than “because it’s the way it has been” and they have no current alternative.

Attempting to link data in this way amounts to systematized guesswork because the inputs used to link the data are questionable to start with (reasons put forward in the project proposal held true in each region). The term “guesswork” is not intended to mean there is no rigor attempted, but instead speaks to the accuracy (often wrong) of the data being used to find matches – if inputs can’t be trusted 100% of the time the process is built to fail by design. It’s almost like trying to solve a jigsaw puzzle when you can’t see the finished picture.

If a system were designed from scratch today with a requirement to link these data sets for every trip with certainty, the current idea of matching this data after the individual data sets are collected would never be considered. This is not a comment on the efforts of NMFS staff that have had to handle the real need to match this data with the resources available. These systems have grown organically over time with often unpredictable budget conditions, political interference, during crises, sometimes by non-programmers that needed to solve a distinct shortcoming of existing systems, and at least initially with very different goals (usually focused on each data set as an isolated silo). They have done and are doing what they can with what they have to work with.

Now that NMFS is making a concerted effort to transition to electronic reporting (and monitoring), it is an excellent time to re-assess how these data sets are related back to the trip and to each other. Most people working in fisheries will say that what often gets the attention is what is most urgent (“on fire”). This is a condition that does not often allow for a re-thinking of past and current efforts.

These are some fundamental questions that we asked during this project:

- ***Is the way that matching these data sets really working now and will it work in the future?***
 - The answer is no on both counts. Regardless of real or perceived high levels of accuracy, a data collection system should be designed with perfection in mind and not “good enough”. Just because it seems to work

today does not mean it is correct and does not mean it will still be acceptable in the future if the details that currently make it manageable change (like a key person holding it all together is overwhelmed or leaves).

- ***What impact on systems will the increasing move towards catch shares mean to the accuracy and timeliness requirements of data collection?***
 - The impact is clear. In all cases, catch shares based fisheries require more timely data. This means systems need to be more nimble and the data is monitored in near real-time. Really it is the fact that quota is money that drives these needs. Tracking it should be considered akin to a banking system and requires much more rigor than is typically found in current data systems.

- ***What does the increasing public interest in oceans preservation and sustainable seafood (i.e. traceability efforts and proof of sustainable fishing etc.) mean to NMFS data collection requirements going forward?***
 - The recent report from the Kingfisher Foundation¹ shows that public awareness of data issues is on the rise as is their desire for more access to fisheries data and information.¹

- ***What is the opportunity cost of staff time and focus (what else could they be doing) when trying to put the puzzle back together (current method of fuzzy matching)?***
 - The recent report from the Kingfisher Foundation¹ shows that managers and science staff are often left to deal with data related issues which take them away from their core duties which are performing scientific and managerial analyses.

¹Lowman & Wing, “Modernizing U.S. Fisheries Data & Information Systems”, Kingfisher Foundation, 2015

2.4.2 Current “Matching” Needs

Before delving into the solution, it is important for the reader to be aware that not all regions currently match all four data sets of interest.

In fact, none match and store a linkage to the trip for VMS GPS data. This is mostly because they don’t have copies of the data (except Alaska), they have no long term need to do so, and it is often felt the data is in no way associated with a trip (it’s a continuous log of data received from VMS units and the data contains no concept of a trip). The data is held in a national database and is generally only used to do ad hoc queries to look at or for specific events for a specific vessel (usually to do with signs of fishing activity near or in restricted areas, to prove a trip occurred that data is missing for, and similar uses).

In general, most regions are only matching two of the four data sets. These sets tend to be either dealer purchases and logbooks or dealer purchases and observer reports. This has been mentioned so the reader is aware that parts of the conceptual proposed solution across all regions will speak to data sets that the reader’s region may not yet be matching. This

has been said so it is clear that the author is aware of this and that the reader can simply read the sections that do not apply as a roadmap for the future as opposed to a prescription to aid current day processes.

It was found that when analysts have the ability to reliably perform analyses using all four streams; they do so habitually and to great advantage.

Anecdotally it has been mentioned that in those cases where cross stream analyses are not possible or convenient, they are sometimes thought of as not necessary. In reality, analytic power is being sacrificed.

2.4.3 VMS: Why it's different and how it could be used

It is important to note that VMS GPS track data seems to be the most ignored data set when in fact it may be able to aid in various circumstances.

Some examples are as follows:

- 1. Reducing the burden of reporting of positional data on fishers.** Positional data requested in most logbooks can be difficult for fishers to obtain while gear is being set/hailed and is known by NMFS data analysts to be of suspect quality. Rather than having the positional data recorded in the logs, an alternate approach would simply require times to be recorded when setting and hauling gear. If those times were analyzed in conjunction with the GPS track of the trip, then the locations of the setting/hauling of gear could be derived. This could be accomplished with a fairly straightforward mobile application that runs on a smartphone or tablet. Large buttons could be placed on the screen that can be pressed to indicate specific events (setting/hauling gear of a certain type as well as identifying a set number or other identifying information for the time recorded). The fishers could have multiple gear profiles saved in the application making it easy to place the custom buttons they need on the application screen for easy access while on deck involved in the harvesting activity. The data collected on mobile devices could easily be shared with a more robust data collection system (perhaps a full blown e-log on a PC/laptop) via Bluetooth in order to fill in the other details about each set (or whatever resolution is required in the regulations for the fishery).
- 2. Provide a simple cross check of reported dealer interactions.** By comparing the port(s) the vessel is known to have visited from the VMS GPS track data to the dealer(s) location the vessel reported it sold to could spot discrepancies in vessel/dealer interaction.
- 3. Provide the trip end date, time, and port.** If the regions can get a copy of the VMS GPS data, the data could be queried using the known departure date of a trip and the associated VRN to assess when that vessel returned to port after fishing activity took place – that is the trip end date/time. This removes the need for vessel operators to send trip end hail notifications for the purpose of knowing their landing port, date, and time. There may still be other needs to send trip end notifications, but those other reasons tend to be more about notifying someone on shore about the pending arrival of the vessel (usually a dealer). Those notifications could be looked at as being between business entities and not the business of NMFS – thus perhaps

NMFS does not need to be involved and can let those involved manage this sort of communication themselves as they see fit.

It should be pointed out that VMS data is different from other data sets as the architecture of the national VMS system does not lend itself to the direct injection of a trip ID as is desired with a logical equivalent model. If regions are able to keep copies of the national VMS GPS track data, they can then alter the data structure as they need to insert trip IDs for the purpose of linking the data to the trips it came from. There are alternatives to using VMS GPS track data put forward later in this report and those alternatives lend themselves to easier linkages between GPS track data and the trip that created the track.

VMS GPS data is just date/time stamped log of co-ordinates, position, and speed taken at regular intervals. The good news is that if a trip start and end date/time are known, then querying the VMS data to find all logged records that relate to a given trip should be very straightforward. This is one instance where matching by dates is acceptable. The issue though is that most regions do not have a copy of the VMS data to query – instead they can use the VTrack application to view the data stored in the national VMS database.

3 National Conceptual Proposed Solution: A Trip Registry

To reiterate, a single national system is not being proposed. However, the issue that exists in each region is the same, so the proposed solution is being discussed here at a conceptual level that can be applied to any region that chooses to do so.

As mentioned earlier, the crux of the issue with attempting to match this data back to the trip after the trip is over is that while data is being collected there is no linkage back to the trip that can be used. In fact in a large number of cases there is no knowledge of a trip occurring and it is only determined there was a trip when data from one or more of the data sets collected on the trip (i.e. logbook, dealer purchase, observer report) is submitted to NMFS (if at all and if not confused with another trip). This is why “fuzzy matching” has been in use – it was the only way to attempt to connect the data with the systems that were in operation.

There are exceptions where a trip is known about ahead of time. This occurs when trips are required to have an observer(s) onboard – these are known about ahead of time in order to schedule the observer(s). Even though they are known about ahead of time, no other fishery dependent data system is aware of those trips. The data about these scheduled trips is typically found in NMFS observer scheduling systems or in proprietary systems at observer contracting companies. Even if NMFS had access to the data from observer contracting companies, not all trips are observed – thus not ALL trips would be defined there.

The only way to consistently link all trip-level data back to the trip (the main objective of this report as laid out in the project proposal) with 100% accuracy is to use a unique identifier that is common to all submitted data sets.

The only way to have a common unique identifier to use with collected data for a trip is to define that trip ahead of time.

The proposed solution is outlined below and focuses on the conceptual side as opposed to delving into a full technical design (i.e. detailed blue prints to build a system). Pros and cons will be discussed, screen mockups provided, as well as system details to be considered (for the proposed solution as well as the impact to existing systems) when implementing the proposed solution. The proposed solution is in line with the “logical equivalent” as mentioned in the project proposal at the start of the document. That means a single pure integrated reporting system that collects and reports these four data sets will not be part of the proposed solution. Instead an approach that allows the independent silo sub-systems to collect a common unique trip identifier (trip ID) along with the data that is already collected by each sub-system is being proposed.

3.1 Integrated Reporting vs. Logical Equivalent

Before delving into the proposed solution details, below in Figure 2 is a comparison of process flows that shows a true integrated reporting system and a logical equivalent.

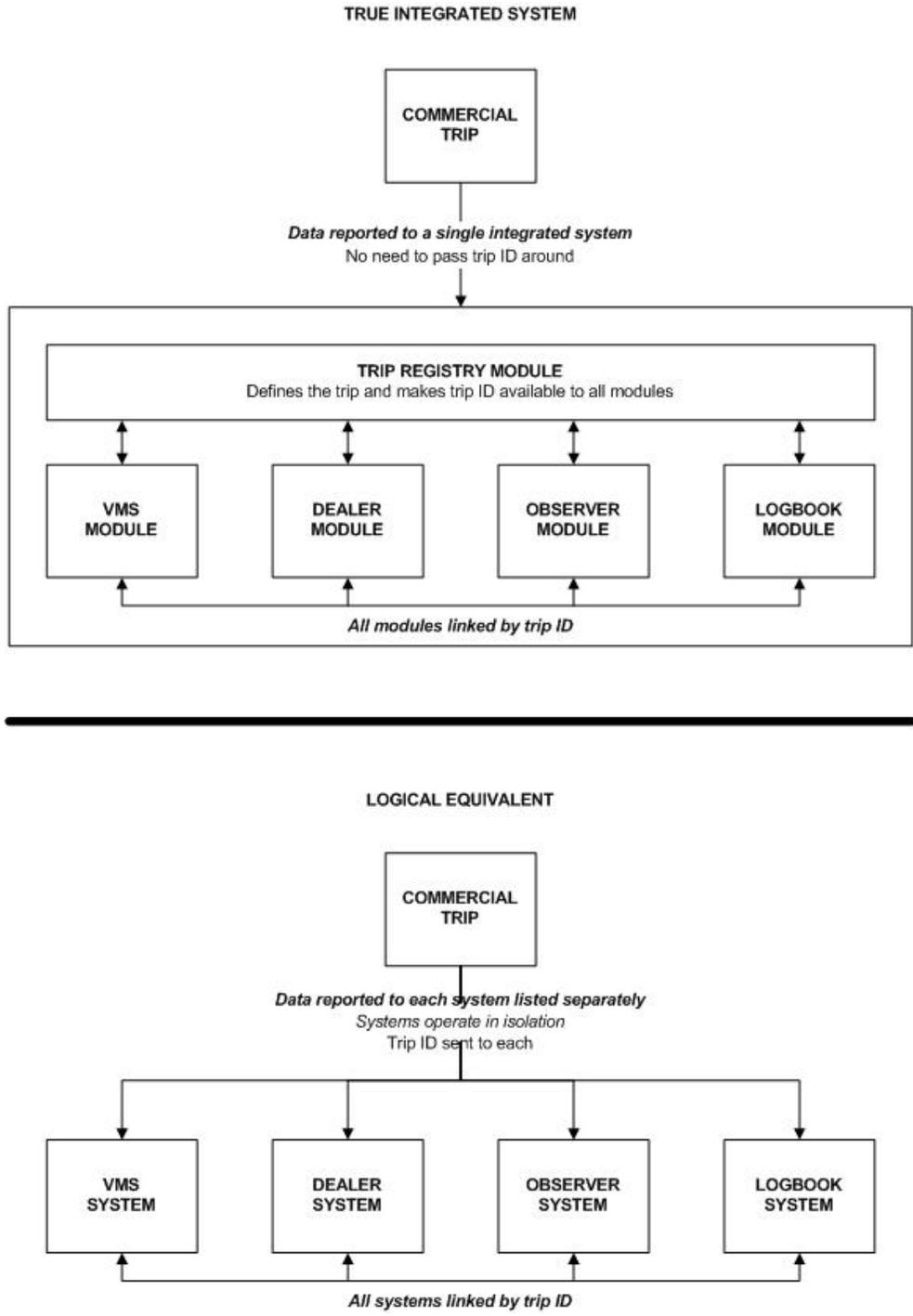


Figure 2: Pure Integrated Reporting System compared to the Logical Equivalent

The main takeaway from this comparison is that the true integrated system is a single system with one database that houses all the data for all four data collection modules and the trip registry where trips are defined. The logical equivalent attempts to “glue” together the data collection sub-systems, but makes the sharing of the trip ID among those systems more cumbersome than a true integrated system. Using the integrated model allows each module to communicate with the other modules very simply and without extra infrastructure (i.e. webservices, database links across schemas, and other non-desirable and error prone approaches). This allows information to be known about the data collection status on a trip with ease and lends itself to application logic to control aspects of data management that the logical equivalent cannot – most notably the ease of controlling unique identifiers like the trip ID.

The following scenario helps explain the difference in the real world. If all the data and functionality for data collection is housed in a single system, it becomes very easy to have an e-dealer interface that is aware of all active commercial trips because all the data is in one system. With the logical equivalent it would be necessary to build functionality into the e-dealer system for it to be able to query another system that may house the definition of trips. This may sound like a subtle difference, but complexity can increase drastically when separate systems are in use as is the case with the logical equivalent. With a logical equivalent you can only go so far before a true integrated system should be considered as opposed to pushing the logical equivalent past what it was intended to do.

This is not to say that the logical equivalent is not without merit. If done correctly it is still a long way ahead of how fishery data management is currently handled in the U.S. and can be used as an excellent stepping stone to move towards a pure integrated system in the future. The reality is that there is a lot of knowledge, time, and money invested in the individual systems and disrupting all of that to jump to a pure integrated system would be unwise. By using the logical equivalent model, areas where interoperability between silo systems get driven into the light and can be used to drive requirements of a future true integrated reporting system. This will help to ensure that the integrated system is built correctly and can ultimately replace the existing silo systems in their current form

3.2 Overview of Proposed Solution

The ideal solution would involve having every commercial trip registered in a trip registry a reasonable amount of time before the trip is to depart. This would allow for the trip to be defined along with a unique identifier for it (trip ID) which could then be accessed by the sub-systems that require it so that the data collected in those sub-systems could include the trip ID which would then get reported along with the data those sub-systems already collect. If this is accomplished there will be practically no matching needed in the future and the instances where there is some effort required would be much easier to resolve – ***this means staff can spend far more time using data to manage the resource instead of managing data issues.***

Overall, this solution is best implemented with e-reporting solutions in place as it will allow for the handling of the trip ID to be done automatically without human involvement (and the known errors humans make).

That said, there are options that can be used to address the handling of the trip ID in each sub-system being integrated. These options will still suffer from varying degrees of what

“Assumption 1” states in section 1.1.2 – humans can’t always enter data accurately. These human errors can be mitigated. When dealing with these errors, there will be far more good matches and thus what is left is a smaller volume of possibilities – *a smaller haystack in which to find the needle.*

3.3 Sample Usage for Clarity

Before delving into the mechanics of this solution, let’s start with an example showing how a dealer purchase can be linked back to the trip as the dealer report is filled in and submitted.

To keep this example simple - assume e-reporting is in use for dealer reporting and a trip for the sample vessel has been scheduled in the trip registry.

The broad steps involved in a perfect transaction would be as follows (handling imperfect transactions will be covered later). Please note that finer details about this flow are provided later:

1. A vessel owner registers an upcoming trip in the trip registry
 - a. When doing so the vessel’s VRN would be recorded against the trip as well as the departure date of the trip and other details (itemized later). Note that the VRN would be provided by NMFS and not hand entered by the vessel owner in order to ensure no typos of this important piece of data. Details provided as to how this could occur given later.
2. At some point before leaving, the vessel operator uses a mobile app to acquire the trip ID for the scheduled trip.
3. At the dock before leaving on the trip, the vessel operator scans an RFID or barcode attached to the hull of the vessel (which provides the VRN) using a mobile app. The app will check the scanned VRN to ensure it is the vessel associated with the trip ID the vessel operator has already acquired. If it is not correct the operator can correct it prior to departure.
4. If an observer will be on the trip, they can acquire the trip ID from the vessel operator using a mobile application when they board the vessel.
5. The trip departs as planned and the vessel operator and possibly an observer transmit their acquired trip IDs to their respective e-reporting applications (logbook and observer). Harvesting occurs and fish are caught.
6. The vessel submits a trip end notification (if truly required) and returns to port to sell the harvest
7. If the vessel operator did not verify the vessel as they should have before leaving on the trip, then they will have to do it before the mobile application they use will transmit the trip ID to the dealer. If the vessel comparison has been completed, the dealer would receive the vessel’s trip ID electronically from the vessel operator. Once the trip ID has been acquired, the dealer would enter the purchase details into the e-reporting interface they use. This trip and vessel verification process would

stop a potentially large error from occurring in the field as opposed to catching the issue after the fact as is now the case. Using mobile devices and Bluetooth and NFC (near field communication) are advantageous in these interactions because they are short range technologies and ensure the vessel operator and dealer as well as vessel operator and vessel are in close proximity – further helping to reduce errors and uncertainty.

8. When the dealer report is submitted, the trip ID is part of the data and it is now linked to the trip
9. No post trip matching required and a huge amount of effort and potential for error removed (which is known to occur)

It should be noted that this same process of handing off a trip ID and any other critical information electronically could also be leveraged in situations where harvest is offloaded to a truck and then takes the harvest to a specified dealer. It is known that there are many instances where the driver takes the catch to a dealer other than the one specified. If the vessel operator transferred the trip ID along with the ID or permit number of the dealer they want their harvest delivered to, it would be easy to spot when it was taken to another dealer because the intended dealer data could be part of the information the driver electronically transfers to the dealer they do deliver the harvest to.

3.4 Benefits of a Logical Equivalent Trip Registry

Know a trip even exists: As mentioned earlier it is very common for there to be no record of a trip even occurred until data for it starts to be submitted

Use to feed sub-systems that need prior knowledge of trips: One such possibility being observer systems used to schedule an observer(s) for vessels that require them.

Remove the need for separate trip start hails or trip declarations: In fisheries that use these forms of notification. The same details can be collected – just entered via the trip registry instead which can then pass those details to whomever needs them. This could occur automatically as soon as the trip is scheduled to depart. Even if the vessel does not depart on time, by using VMS GPS track data a more precise departure time can be derived.

Increased precision and timeliness requirements: This will benefit (or can't hurt) all fisheries, but specifically catch shares fisheries as they become more widely used.

Aid in the rapidly increasing volume of data requests: These requests are on the rise and encompass the desire for more accountability, full traceability, sustainable fishing claims, as well as NGO/NMFS regions/NMFS HQ/council/commission/state and other data requests.

Be used with paper and electronic reporting options: In both cases there are implementation options that will allow for either high levels of accuracy or 100% data linkage (getting trip ID and then transposing would have lower accuracy but can be mitigated using printed bar codes and similar concepts). Regardless of approach and anticipated levels of accuracy – any needed matching efforts would be tackling a smaller

volume of trips and would still have the trip registry to rely on to aid in the matching effort – making it easier and less labor intensive. Any mention of paper based workarounds are not condoning continuing to collect data on paper and are instead meant as temporary methods and should be phased out as soon as e-reporting can be implemented.

Can use modern technology which is far more efficient at handling critical pieces of data accurately: This benefit will be realized as e-reporting becomes mandatory and will most likely involve the use of mobile technology to convey critical unique identifiers and other trip details that should not be managed by humans. Near field communications, RFID chips, and Bluetooth should all be considered to name a few. One option to consider is that it is possible to have vessels identify themselves to dealers and to cross check that the vessel at the dealer is the same vessel as is identified on the trip in the registry – all with a few quick scans using mobile devices. What’s even better is this technology is readily available today.

Remove redundant reporting: All vessel details can be stored in a vessel profile in the trip registry or in other data source (linked by VRN) – no need for vessel operators, dealer, or observers to report those details). Profile details can be available to all sub-systems that need them in an effective dated manner so the details can be known as of the trip.

3.5 Detailed Look at Trip Registry Functionality

3.5.1 Overview

All of the functionality to be discussed will be accessed through a web application with the exception of the webservice which can be used with or without a user interface. In all cases an Internet connection will be needed to use the described functionality, as well as a modern web browser (with some exceptions for the webservice).

When designing a trip registry the timing of the normal order of data collection efforts around a commercial trip must be considered. This information is critical to being able to collect the correct core information about a scheduled trip so that the sub-systems that need to report the trip ID can easily retrieve it from the registry with only the VRN as an input.

The trip registry functionality will be laid out assuming the desired scenario of full electronic reporting is in use by all data collectors. This means that the following data is collected by the data collectors listed below:

- **Vessel Operators (who may also be vessel owners)** – collect/submit *logbooks* and send *trip end notices* electronically
- **Dealers** - record and report their *seafood purchases* electronically
- **Observers** - record and submit their *observer* data electronically

We are sticking with the full e-reporting scenario because it demonstrates how efficiently fishery dependent data can be collected and how e-reporting can change the speed at which data can be received.

In reality, a major shift can be undertaken (if desired) that allows for true real-time reporting. This is vastly different to allowing vessel operators and dealers to report their logs and purchases every week or month in aggregate – instead the data can come to NMFS

as soon as the data collectors are finished recording data (end of trip for vessel operators and purchase is complete). Observers require more time post trip to accommodate the debriefing process, but could submit the data as soon as it has been finalized.

By receiving data in real-time, errors can be spotted sooner and resolved while the memory of the event the data is still fresh and thus more reliable. If data collectors are not following protocols, they can be informed sooner as opposed to being left to operate incorrectly for a week, a month, and sometimes longer before any corrective action can be taken (making the issue worse in terms of data quality).

With real-time data coming in, data collection moves from being the mostly passive and reactive process it is now to a pro-active process that can be nimble and allow for fast corrections.

Using a trip registry represents a significant shift in data collection efforts in fisheries because the established order of events needs to be followed and if not, some steps may not be able to proceed unless a prior step goes according to plan. Being that data is currently collected in isolation and does not rely on any other parts of the collection process, the timing and order does not matter to any significant degree. This will change with the use of a trip registry, but the benefits are well worth the effort.

This shift towards a more timing and order sensitive process means that “what if” needs to be asked in places it never had to be before. For example, what if a vessel operator is unable to retrieve a trip ID from the registry prior to departing on a scheduled trip? These kinds of scenarios are manageable, but they do require a change in how data collection is looked at.

Really the shift is towards changing the requirements for logbook and dealer reports from being submitted in batches every X amount of time (often every week up to monthly) and instead expecting them to be submitted as soon as they are filled in and finalized. This means considering the ways in which data collection could be disrupted and having ways of making corrections when the issue is resolved or quarantining the trip or parts of it so that it can be checked later for accuracy. Observer data submissions can still occur in the timeframes they do now – mostly due to the time needed for the debriefing process. However, observer data is also often submitted in batches, so it is suggested that it too can be submitted as soon as an observer report for a given trip is finalized.

Before diving into the details, below in Figure 3 are the generic steps that would be taken on a commercial trip by vessel owners, vessel operators, observers, and dealers when interacting with a commercial trip. There are varying ways the trip ID can be retrieved and applied to the data being collected. Most differences tend to lean towards who is doing the data collection and what method is less burdensome, less error prone, and most accurate. As mentioned earlier, the first pass of the trip registry functionality will assume full e-reporting with various options being discussed after.

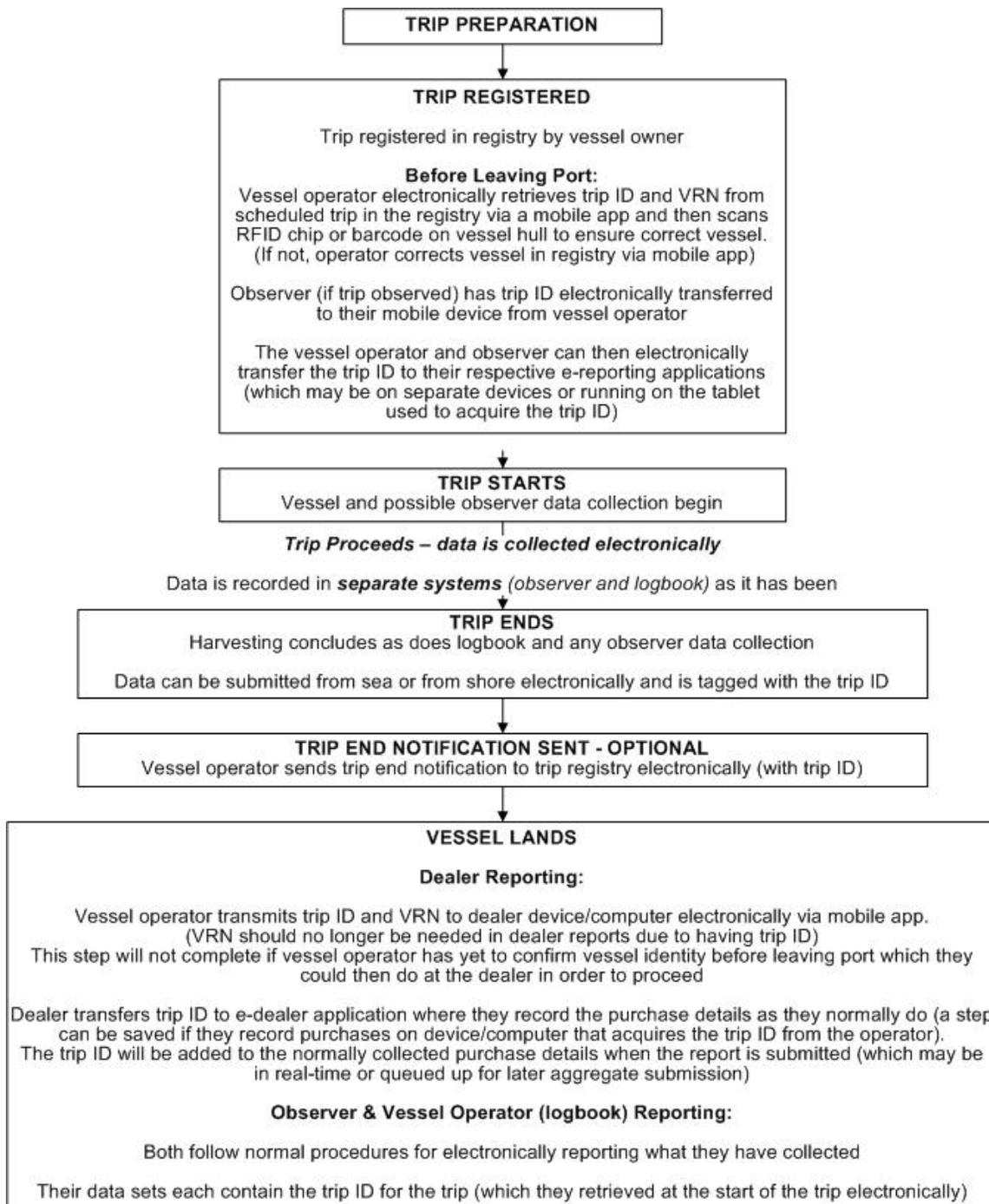


Figure 3 – Generic Trip Registry Steps for a Commercial Trip

3.5.1.1 General Data Structure Details to be Aware of

In the proposed solution below, database tables have been provided to better understand what is being proposed. Plain English names have been used instead of trying to impose any kind of table or column naming standards. Only basic data types are provided without length or precision details. Basic effective dating is being used which can be replaced with built-in database options such as journaling in Oracle. It is fully expected that additional columns may be needed by regions and perhaps even more tables. What is laid out are merely examples and not meant as a rigid blueprint.

3.5.1.2 Preserving Data History

All tables include “create” and “modify” date/time columns so these details can be known when needed. The date and time records were modified can be used to know when either the application updated the records or whether a database administrator has manually changed a record. It is better to have this information and not need it than to not have it in the first place. Database triggers can be added to all tables with these columns so the date/time values placed in them are handled automatically by the database with no need for the application code to touch them.

In some cases there is a “base” and a corresponding “detail” table suggested. The “base” table just holds an ID column (primary key) and a create date/time. The “details” table holds all the details for each ID in the “base” table. It may seem odd to use this data structure, but it exists because the table examples are using effective dating. It would be very inefficient to have a record ID change every time the details are updated by a user. The idea is to leave the base ID record unchanged and only ever end date detail records and create new ones when updates are made (preserves the data prior to the edit and lays down a new version with the updated data). This is done because the ID is associated with other related data and if the base ID changed every time details were updated (due to effective dating), those relationships would be lost or require too much extra effort to keep in sync.

Sample Tables to Illustrate

With the first option, every update to the user details would result in a new record because effective dating is in use. If a user ID was related to a vessel in an association table (links user IDs to the vessel IDs), then every update to the user details would mean updating those user/vessel association records with the new user ID (not very efficient).

User Table: Option 1

Column Name	Data Type	NULL?	Notes
User ID	Integer	NO	Primary key
First Name	Varchar	NO	User’s first name
Last Name	Varchar	NO	User’s last name
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

With the Option 2 set of tables below, changes can be made to the User Details Table as many times as needed, but each new record created due to a change will NOT result in a changed user ID. The user ID is maintained separately in the User Base table and thus never changes.

There are other methods available to maintain primary keys under this scenario. This example is illustrating the concept of managing this data in a clear way that does not require extra maintenance.

User Base Table: Option 2

Column Name	Data Type	NULL?	Notes
User ID	Integer	NO	Primary key

User Details Table: Option 2

Column Name	Data Type	NULL?	Notes
User ID	Integer	NO	Foreign key to User Base table
First Name	Varchar	NO	User's first name
Last Name	Varchar	NO	User's last name
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

3.5.1.3 Lookup Tables

There are also tables listed that may exist elsewhere in the region – perhaps because they hold core standardized data. These are referred to below as “lookup” tables. If they do reside elsewhere in the region, they can either be replicated in the Trip Registry database or use a common unique identifier so that Trip Registry versions can easily be mapped to the regional standardized data. If replicating or otherwise using a copy of regional data, a data management plan will be needed to ensure it stays up to date in the Trip Registry database.

These lookup tables all feature a “code” column. Although the tables themselves have an auto-generated primary key, that value will change over time as the data in these tables gets updated. So having a unique code in common across the changes is critical to know that a newly activated record is related to the old deactivated record. A simple example of this would be if a record in the fishery table had a fishery name of “Fishery ABC” (fishery ID is 1 and fishery code is “aaa”) and that gets used for 6 months for scheduled trips. After 6 months it is determined that “Fishery ABC” was not a clear enough name for some users and it gets changed to “Fishery XYZ”. When the change occurs the new record gets a new fishery ID of 2, but the unique fishery code remains as “aaa”. The name has changed, but it is still the same fishery. This allows for reporting to be done for all trips that involve this one fishery regardless of which version of the fishery record is recorded against scheduled trips.

3.5.2 NMFS Loading of Key Data to Avoid Errors

There is core data mentioned below that should be loaded (and verified) into the trip registry prior to starting operations. This is to facilitate the auto-generation of many user accounts using this core data to associate critical elements with the user accounts without the user creating those critical associations. If end users are allowed to key in or otherwise decide which data is associated with them, errors will occur (like the wrong vessel associated with a vessel owner and similar issues). Once loaded, all dealer, vessel owner, and vessel operator user accounts should be able to be generated automatically and credentials sent to the users (with passwords set with the requirement to change the password on the first login for better security). By auto-generating these accounts it will avoid a large amount of support because training is not required for large parts of what would otherwise be self-serve account creation. There may still be data issues found in the data laid down by NMFS, but it is best to hear about these issues and resolve them in a controlled environment. In some regions, some of this data is known not to exist yet. Data clean-up efforts as well as adding data not yet available may be a bit of a push to start with, but this is time and effort well spent as this data is the foundation of a data collection system with high levels of accuracy. It's very hard to change the foundation of a house once the house has been built!

3.5.2.1 Sample NMFS Loaded Tables

3.5.2.1.1 Lookup Tables

These are all of the lookup tables that will need to be loaded and most will be used to create user accounts (some support other trip registry functionality detailed later).

This data defines:

- The vessel and vessel owner data that will be associated as part of a vessel owner user account
- The dealer and port data that will be associated with a dealer user account
- User roles to be associated with every user account
- Units to be used for the hull length in the Vessel table
- The list of vessel operators that will need user accounts created and the unique identifier (operator permit number or a unique code for each operator)
- The list of vessel owners that need user accounts
- The list of dealers that need user accounts
- A list of species that may be encountered in the region
- A list of trip types used in trip scheduling
- A list of fisheries in the region used for trip scheduling

Vessel

A list of all known vessels and their associated details of interest that NMFS is able to access (either internally or perhaps via the Coast Guard or whomever is responsible for the vessel registry in the region).

Vessel detail needs will most likely differ from region to region, but should have some standard core details as follows:

- Vessel name
- Vessel registration number (VRN)
- Vessel length
- Vessel length units

Other data elements can of course be added to meet regional requirements. It is suggested that a review of vessel details that are typically requested of dealers, observers, and vessel operators be undertaken. With those details it should be fairly straightforward to add columns to the table below.

Column Name	Data Type	NULL?	Notes
Vessel ID	Integer	NO	Primary key
Vessel Name	Varchar	NO	Vessel common name
VRN	Varchar	NO	Vessel registration number (unique)
Vessel Length	Varchar	YES	Length of vessel
Unit ID	Varchar	NO	Foreign key to Unit of Measure table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Vessel Owner

A list of all known vessel owners with a unique identifier for each should be created. This list should include first/last name, middle initial, and name suffix as a minimum. It would be helpful if contact details could also be obtained. The table should be effective dated in order to track changes as dealers are added or removed from the list.

Column Name	Data Type	NULL?	Notes
Owner ID	Integer	NO	Primary Key
Owner Code	Varchar	NO	Unique code for the owner
First name	Varchar	NO	First name of the owner
Last name	Varchar	NO	Last name of the owner
Middle Name	Varchar	YES	Optional middle name of the owner
Name Suffix	Varchar	YES	Optional name suffix (i.e. Jr, Sr, II, etc.)
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

User Role

A list of the user roles in the trip registry.

Column Name	Data Type	NULL?	Notes
Role ID	Integer	NO	Primary key
Role name	Varchar	NO	Name of the role
Role Code	Varchar	NO	Unique code for the role
Sort order	Integer	NO	Ordinal value used to sort the list of roles
Is Admin	Boolean	NO	Indicates if the role has admin privileges or not
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Vessel Operator

Some regions already have this sort of list. In Canada all vessel operators have a FIN (fisher identification number) and some regions in the U.S. have or are heading in that direction (registering all vessel operators and deckhands). This list is VERY useful in using to allow vessel owners to select the operator operating their vessel on an upcoming trip in the registry. Many regions ask for an operator permit number to be submitted with logbooks and having this list would make the registry that much more useful in reducing reporting burden on industry (as they can just select the correct operator and thus won't accidentally mistype an associated operator permit number).

Column Name	Data Type	NULL?	Notes
Operator ID	Integer	NO	Primary Key
Operator Permit	Varchar	NO	Unique code or permit number for operator
First name	Varchar	NO	First name of the operator
Last name	Varchar	NO	Last name of the operator
Middle Name	Varchar	YES	Optional middle name of the operator
Name Suffix	Varchar	YES	Optional name suffix (i.e. Jr, Sr, II, etc.)
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Unit of Measure

A list of units that can be used to describe data that express a quantity, length, or other measurement (i.e. feet for use with vessel lengths).

Column Name	Data Type	NULL?	Notes
Unit ID	Integer	NO	Primary key
Unit Name	Varchar	NO	Plain English name of unit
Unit Code	Varchar	NO	Unique code for each unit (best to use unit short forms here like lbs for pounds)
Active DTT	Date/time	NO	Date/time record becomes active

Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Dealer

All known dealers in the region, their permit/license numbers, a unique dealer code, and any other descriptive details like the dealer operating name and the port ID in which the dealer operates.

Column Name	Data Type	NULL?	Notes
Dealer ID	Integer	NO	Primary key
Dealer Name	Varchar	NO	Plain English name of the dealer business
Dealer Code	Varchar	NO	Unique dealer code
Dealer Permit	Varchar (Integer if permit is numeric)	NO	Dealer permit number as of the active DTT and before the inactive DTT (if set)
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Port

All known ports in the region (or that may be landed at by vessels in the region), their unique port code, and any other descriptive details like the port name.

Column Name	Data Type	NULL?	Notes
Port ID	Integer	NO	Primary key
Port Name	Varchar	NO	Plain English name of the port
Port Code	Varchar	NO	Unique port code
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Species

All known species in the region, their unique species code, and any other descriptive details like the species common name.

Column Name	Data Type	NULL?	Notes
Species ID	Integer	NO	Primary key

Species Name	Varchar	NO	Plain English name of the species
Species Code	Varchar	NO	Unique species code
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Fishery

All fisheries in the region, their unique fishery code, and any other descriptive details like the fishery name.

Column Name	Data Type	NULL?	Notes
Fishery ID	Integer	NO	Primary key
Fishery Name	Varchar	NO	Plain English name of the fishery
Fishery Code	Varchar	NO	Unique fishery code
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Trip Type

All trip types (i.e. harvest, set gear, transiting ports, etc.) that may be applicable in the region, their unique code, and any other descriptive details like the trip type name. If trip types are fishery specific, an association table like many defined in the next section can be created to allow for the filtering of a trip type select box based on the fishery a trip is being scheduled for.

Column Name	Data Type	NULL?	Notes
Trip Type ID	Integer	NO	Primary key
Trip Type Name	Varchar	NO	Plain English name of the trip type
Trip Type Code	Varchar	NO	Unique trip type code
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

3.5.2.1.2 User Account Tables

These are the set of user related tables that will be populated when the auto-generation of accounts is performed. This auto-generation does not include observer or administrator

accounts. If a list of known administrators and observers can be created and loaded, the concepts put forward here can be used to generate those accounts as well (recommended if possible)

User Base

Described earlier and exists to preserve a single User ID for a user across effective dated account changes over time.

Column Name	Data Type	NULL?	Notes
User ID	Integer	NO	Primary key
Create DTT	Date/time	NO	Date/time the record was created

User Account

This table defines a user account. Additional details can be added as Regions' needs dictate. This table and the User Base table above can be populated using the appropriate role ID from the User Role table and the Vessel Owner, Vessel Operator, and Dealer tables to know who to generate accounts for. The username and password (encrypted) can be auto-generated using strong passwords set to require a reset on first login.

In the case of vessel operators and owners, the name and perhaps other tombstone data can be pulled from the Vessel Owner and Vessel Operator tables instead of populating the matching User Account columns. This is so NMFS keeps control of the official names on file for these user roles (as they came from NMFS data in the first place). If it is decided this is the correct approach, then the NOT NULL constraint on the columns in common with the User Account and either Vessel Owner or Operator tables can be removed (because not all records will have values for those columns) and the data can be displayed in the user account in the registry via table joins to the Vessel Owner and/or Vessel Operator tables.

There is a set of four individual names and a 5th organization name column – all allow NULL. This is because it is intended to either use an individual's name or an organization's name (for dealers) – but not both. That said, it may be desirable to have accounts for individuals that work at a dealer location because if anything goes wrong or is handled incorrectly, the individual that was involved can be identified instead of just the dealer. If this is the decided upon course of action, then the organization name can be pulled in from the Dealer table for all accounts associated with a given dealer, the organization name column can be removed from this table, and the first/last name columns can be set to not allow NULL.

Column Name	Data Type	NULL?	Notes
User ID	Integer	NO	Foreign key to User Base table
First name	Varchar	YES	First name of the user
Last name	Varchar	YES	Last name of user
Middle Name	Varchar	YES	Optional middle name of user
Name Suffix	Varchar	YES	Optional name suffix (i.e. Jr, Sr, II, etc.)
Organization Name	Varchar	YES	Organization name (for dealers)
Role ID	Integer	NO	Foreign key to User Role table
Username	Varchar	NO	Unique username for the user
Password	Varchar	NO	Password for the user – suggest storing

			encrypted or hashed
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Owner Vessel

Each vessel ID can be associated with the vessel owner ID of the owner of that vessel. By using effective dating, this table will hold a history of which vessel owner owned which vessel and when. This table can be populated using data in the Vessel Owner and Vessel tables. It serves to link vessel operator accounts with the vessel(s) each owner owns and avoids the users themselves being allowed to setup (and create errors in the process) this relationship.

Note that the Owner Vessel table needs a unique constraint applied using the combination of Vessel ID, Owner ID, Active DTT, and Inactive DTT to avoid scenarios where a vessel might be accidentally assigned to more than one owner at the same time (thus the date time fields being part of the constraint). The table also contains a foreign key reference back to the User ID in the User Base table. This is why the User Base table exists – to preserve the User ID every time the user account details get updated which maintains foreign key constraints while allowing the history of user account changes to be preserved.

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary Key (surrogate)
Vessel ID	Integer	NO	Foreign Key to Vessel table
Owner ID	Varchar	NO	Foreign Key to Vessel Owner table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Owner User

Each vessel owner ID can be associated with the user ID of the user account created for that vessel owner. By using effective dating, this table will hold a history of user accounts and their associated vessel owner(s). The data to populate this table come from the Vessel Owner and User Base tables.

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary Key (surrogate)
User ID	Integer	NO	Foreign Key to User Base table
Owner ID	Varchar	NO	Foreign Key to Vessel Owner table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become

			inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Dealer Port

Each dealer ID can be associated with the port ID of the port in which the dealer operates. By using effective dating, this table will hold a history of which dealers operated in which ports and when. The dealer permit number can be placed in this association table if each location a dealer operates has a unique permit number. If instead the permit number is for an entire dealer organization, then the permit number is best stored in the Dealer table. The data to populate this table come from the Dealer and Port tables.

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary Key (surrogate)
Dealer ID	Integer	NO	Foreign Key to Dealer table
Port ID	Varchar	NO	Foreign Key to Port table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Dealer User

Each dealer ID can be associated with the user ID of the user account created for that dealer. By using effective dating, this table will hold a history of which users were associated with each dealer over time. The data to populate this table come from the Dealer and User Base tables.

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary Key (surrogate)
Dealer ID	Integer	NO	Foreign Key to Dealer table
User ID	Varchar	NO	Foreign Key to User Base table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

Operator User

Each vessel operator ID can be associated with the user ID of the user account created for that vessel operator. By using effective dating, this table will hold a history of which users were vessel operators over time. The data to populate this table come from the Vessel Operator and User Base tables.

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary Key (surrogate)
Dealer ID	Integer	NO	Foreign Key to Dealer table
User ID	Varchar	NO	Foreign Key to User Base table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

3.5.2.1.3 Controlling When a Trip can be Scheduled

There may be other functionality needed depending on how each region may want to control how and if trips can be scheduled. This is mainly to avoid allowing trips to be scheduled for activity that is not allowed based on the authorizations granted to the permit involved on a trip. If a trip schedule is being created and data indicates that the vessel should not be involved in that trip for whatever reason (not allowed to fish in the specified fishery, specified fishery is closed as of the departure date specified, or having exceeded individual or sector quota), that schedule should not be allowed to be saved in the registry. Proper data structure is needed to facilitate this kind of functionality.

It is suggested that each region assess all of the reasons a given vessel/permit should not be able to start a trip because if allowed to start, the entire trip would be erroneous because it breaks very clear rules in the region. This is quite different than how most fishing is allowed to be conducted in each region where vessels can leave to fish (often with no notice) and could be fishing during closures, with an expired permit, or in fisheries they are not authorized to fish in. These situations become an enforcement issue after the fact that harvest was taken incorrectly and by then it's too late for the fish.

The information acquired as part of this analysis will allow other data structures to be put in place to define permit authorizations, openings and closures related to fisheries and permits, and any other structures needed to define when a given vessel/permit can be allowed to start a trip. This data should also be loaded and maintained by NMFS.

3.5.3 Secure Login

Modern web application security should be used – to what degree is up to the regions and may be dictated by existing security policies.

At the core, SSL should be used to encrypt all system pages, a username and password should be required to access the system, and data driven user roles should be used to restrict/allow access to appropriate system functionality.

Security should also ensure that no pages can be accessed via bookmarks when a valid system session does not exist. In these situations, a login screen should be forced on the screen and no other options presented to the user (with the possible exceptions of links to contact information, FAQ details, and other links or details that do not require a login to access).

Although every region will have its own security requirements, it is suggested that at the very least the logging of successful and unsuccessful login attempts should be part of the trip registry login process.

By logging the basic details listed in the table below, there are a few pieces of functionality that can be easily added to the login process:

- **Spotting signs of inappropriate access attempts** - multiple attempts with the same or varying credentials from the same IP address in a short time period
- **Locking an account after too many login attempts** - If the same IP address attempts to login with a valid username and incorrect password X number of times, the account can be locked so that nobody can login to it until the lock is removed and the password reset.

Sample Screen Mockups

Login screens are a commodity in systems design and thus a screenshot example has not been given.

Database Table

Login Log

Column Name	Data Type	NULL?	Notes
Log ID	Integer	NO	Primary key
User ID	Integer	YES	Foreign key to User Base table (defined below). Will be NULL if username provided does not match any accounts in the database
Username	Integer	YES	Username as submitted with login attempt – may be NULL if not entered
Password	Varchar	YES	Password (suggest it be encrypted) as submitted with login attempt – may be NULL if not entered
IP Address	Varchar	NO	IP address the login attempt was made from
Login Successful	Boolean	NO	Indicates if the login attempt was successful
Create DTT	Date/time	NO	Date and time the record was created (doubles as when login attempt was made)

3.5.4 User Account Management

The functionality required here is fairly standard for user accounts. In general, the administrator role has elevated permissions while the data collector roles will have varying capabilities in the registry.

User Role Access

Both the administrator role and all of the data collector roles will have access to this functionality – variations in access are explained below

Administrator Role

These will be needed so that an admin can make changes to user accounts as required (should an end user be experiencing any issues or there is otherwise a need to create or manage accounts for others).

This user role should be able to create, generate (from loaded core data from NMFS as described earlier), and edit user accounts for all available user roles that require access to the trip registry. That said, it must be taken into account how these accounts were generated in the first place. If they are meant to be created from core critical data loaded by NMFS, then it is strongly advised administrators should not be able to create those accounts on a whim. Instead formal processes would need to be developed to ensure proper handling of the core critical data and its usage in account creation.

Data Collector Roles

These accounts/roles are for users that collect and report data. Some of these roles are meant to be for individuals, but some can either be for individuals or an organization. All accounts will have the ability to login and change their own account credentials and make changes to data not loaded and maintained by NMFS. The data loaded and maintained by NMFS has to be controlled by NMFS because if it is entered incorrectly then the error will be perpetuated as data is collected and reported.

The data collector roles are as follows:

- **Vessel Owner**
 - A role to be used by vessel owners who may own one or more vessels. It will be used primarily for viewing their vessel list from NMFS, maintaining a list of vessel operators they work with, and searching/creating/editing trip schedules. Note that a vessel owner could also be the vessel operator and thus should have all the same capabilities in the trip registry that vessel operators do.
- **Vessel Operator**
 - A role to be used by vessel operators. It will primarily be used for reviewing scheduled trip details and possibly correcting discreet scheduled trip details should certain issues arise during the trip.
- **Dealer**
 - A role for each dealer business or each employee at each location for a given parent dealer business. It will be used to retrieve specific data about a trip via a webservice and to confirm details (like their permit number and port associations) loaded by NMFS are correct.
- **Observer**
 - A role to be used by individual observers. It will be used to retrieve the trip ID data about a trip they will be observing and to mark a trip as observed in fishery's that do not have 100% coverage).

Database Tables

These tables were all defined earlier and are involved in user account management, but not all are used for all user roles:

- User Role (all roles)
- User Base (all roles)
- User Account (all roles)
- Dealer User (dealer role only)

- Dealer Port (dealer role only)
- Owner User (vessel owner role only)
- Operator User (vessel operator role only)
- Owner Vessel (vessel owner and vessel operator via relationship to owner)

Owner Operator

Column Name	Data Type	NULL?	Notes
Record ID	Integer	NO	Primary key (surrogate key)
User ID Owner	Integer	YES	Foreign key to User Base table
User ID Operator	Integer	YES	Foreign key to User Base table
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known

Sample Screen Mockups

Basic User Account

This is a mockup of a generic user account and simply shows the security details and simple tombstone information. It is in an “add new account” state, but an edit state would not look much different other than to show the previously saved details in the form to be edited. As described earlier, many accounts can be auto-generated based on data loaded by NMFS, so it is expected that in most cases users would see this form in the edit mode. Dealer accounts can also show their business name, permit number, and port name as text (not editable as this comes from NMFS). The user role is not editable as no user should ever need to be moved into another role.

Technically a dealer could end up as a NMFS trip registry administrator, vessel owner, observer, or a vessel operator. The same could be said of any of the roles. The reason it was stated that no user should ever be moved into another role is that the roles are VERY different in most cases. The complexity of the trip registry could go up dramatically if a single user account had to be aware that it used to be a vessel owner account and thus still providing access to historical vessel owner data while also providing all the functionality of the new assigned role. Being that such job changes in the real world would occur infrequently, it does not seem wise to support role changes.

A hybrid solution could also be developed that allows for only certain role changes to occur over time (i.e. maybe a vessel operator could become a vessel owner). That would reduce the potential complexity to accommodate such changes as already mentioned.

Assigning multiple roles to a user at the same time would also drive up the complexity of the trip registry and especially the mobile application to interact with it. This is because the mobile application to be used by vessel operators, observers, and dealers is meant to know which role they are acting as (to facilitate capabilities such as “locate a dealer device” where knowing your device is broadcasting as a vessel operator and knowing other devices nearby are in dealer mode is critical). If a user could have multiple roles, the mobile application has to have an extra step of identifying what role the user is currently wanting to act as (which could easily cause confusion and raise support requests).

Account Details

* = required field

Username: *

Password: *

Repeat Password: *

User Role: ADMINISTRATOR

First Name: *

Last Name: *

Provide one or both of the following

Phone (land):

Phone (cell):

Save Account Details

Vessel Operators Associated with Vessel Owners

This screen would be used by vessel owners to maintain a list of the vessel operators they work with. The vessel operator user IDs would be selected from the list of known vessel operators as loaded by NMFS and saved in the Owner Operator table against the vessel operator's user ID. This list of operators will populate the operator select box in the form used by owners to schedule trips.

My Vessel Operators

Add Operator

PERMIT	NAME	ACTION
1234	Smith, John	REMOVE
4321	Jones, Sarah	REMOVE

Vessel List for Vessel Owners

A screen mockup is not needed for this as it is a simple list and cannot be edited. The data is loaded by NMFS (from regional permit databases) and associated with each vessel owner's account as the accounts are generated. A menu item should exist that allows vessel owners to see this list – perhaps the current vessel associations as well as past.

3.5.5 Trip Scheduling

This functionality is for the vessel owner role and optionally can be accessed by the administrator role (should administrators be needed to manage account details for vessel

owners – perhaps over the phone on a support line). In all cases the roles with access should be able to create and edit scheduled trips.

Scheduling trips and being able to accurately determine the correct trip ID when requested are at the core of the trip registry. Care must be taken to allow enough lead time for sub-systems to be notified of upcoming trips if notification is required by any sub-systems (perhaps for observer scheduling or to notify the OLE of a trip).

Each region will need to determine what information is collected about scheduled trips. The core suggested details will be listed in this section, but the regions are encouraged to review any hail or trip declaration requirements and integrate those details into the trip registry. This is to reduce the burden of reporting on vessel operators by having the trip registry notify whomever needs to be notified (often the OLE) about scheduled trips as opposed to making vessel operators do this via VMS or other means.

A key aspect of the registry is the ability to accurately determine a trip ID while a trip is active (defined as not yet landed and offloaded and is either next to depart or has departed). In order to know this, it must be known when each trip is over (landed) or else it will never be clear if a past trip could still be active or not. Part of the flow of events is that the dealer receives the trip ID and VRN from the vessel operator. The recording of such an event in the registry (via the device the dealer uses to acquire the data) can be used to indicate a trip is effectively over (even though there may still be harvest sales if selling to more than one dealer). By knowing a trip is over and knowing which trip is next, it is very simple to derive the active or next to be active trip ID. Note that trips can be known to be over without the use of a trip end hail.

Using the registry is really not much different than sending a trip start hail – in fact it is much like a trip start hail on steroids. The “on steroids” comment refers to all the extra benefits of knowing about the trip ahead of time and the ability to use the registry to notify sub-systems or organizations as opposed to leaving the burden on the harvester.

There is potential for some more sensitive timing issues to be resolved when considering trips that require observers and if the registry is going to feed observer scheduling systems. Those trips will likely need to be recorded in the registry a specified amount of time before an observer is needed. This of course facilitates scheduling an available observer which can involve requests to change the planned trip itinerary so it lines up with an available observer. The reason this may be more time sensitive is because of the time between observed trips for a given vessel (especially in 100% coverage situations which means every trip is observed). If the time between a vessels landing and wanting to go out on a subsequent trip is shorter than the advanced observer notice for scheduling one there will be a conflict with the rule only allowing one trip to be scheduled in advance. The conflict is that it would artificially force vessel operators to change their plans in order to use the trip registry.

A solution is to allow the scheduling of “tentative” future trips. If the vessel returns in time, the new tentative observed trip can be confirmed. If not, the new observed trip start time is rescheduled based on the late arrival. This would still allow for the trip registry to feed observer scheduling systems and to adjust trip itineraries as needed to ensure an observer is available. In fact another benefit could be realized with a fairly easy adjustment to observer scheduling systems. They could initially receive notice about scheduled trips from the trip registry and when an observer is booked for those trips, the

observer scheduling systems could indicate that fact back to the registry (via a webservice). That way no observers have to manually flag trips as observed in the registry and instead they conduct their normal business via their scheduling system.

The details that need to be collected may be the same for any trip in a given region or it may differ from fishery to fishery or otherwise be split in ways that need varying details to be recorded. Each region will need to determine what factors need to be known in order to determine what details to present in the form for scheduling a trip.

Trip schedule needs will most likely differ from region to region, but should have some standard core details as follows:

- VRN
- Departure date
- Departure port
- Fishery (indicating a fishery helps identify the trip activity)
- Trip Type (indicating a trip type may allow prior knowledge of what data sets may or may not be collected on the trip)
- Target species or species complex (like groundfish)
- Whether an observer is needed or scheduled

Rules are required to constrain the trip schedule editing process. It is conceivable that over the course of the trip it is noticed that the fishery or the vessel or other core critical details are seen to be wrong and may need to be corrected (and it's better for the reporter to do so as soon as the error is realized). Each region will need to assess the allowable edits and the timeframe policies around when such edits should be cut off and ensure the business logic in the registry is set correctly. A way to handle very late edits (perhaps after several more trips have gone by or the timeframe passes in which the trip harvest has been used as part of a stock assessment) is to have a more formal process where the user must request (or NMFS spots the issue and requires an edit) an unlock code in order to make the edit (which can then be tracked and/or the late data change can be dealt with according to the rules of the region). Another approach is to only allow such edits to be performed by NMFS staff and the new version of the trip schedule is flagged as entered by NMFS – perhaps tracking a reason code to go with it. This is not a simple task and does take a fair amount of scenario running to make sure the edits are dealt with correctly. This is time well spent because the result is a data system that has hooks to handle the data the correct and consistent way.

Consistent handling of the same scenario(s) is ideal as the data quality can be known. When database administrators manually force data corrections and changes into a system that is not tracking those changes nor dealing with the reality that such changes could occur, there will be data problems and the data may simply not be reliable.

Sample Screen Mockup

Schedule a Trip

This is how the form would look when a vessel operator first navigates to this page to schedule an upcoming trip. If a trip is already scheduled and is still editable, then the form would be filled in and edits could be made (and a new schedule could not be added).

Notice the user can't type any data and must use proper form controls – thus reducing errors by design. A trip can be scheduled very quickly and should not be seen as a barrier to the registry concept.

There are other interactions possible with a scheduled trip, but they would be via mobile applications that are detailed later as well as via a standard web interface in the trip registry. If edited via mobile applications, only discreet changes would be allowed and other details of the schedule would not need to be presented in those situations. In other cases like observers indicating the trip was or is to be observed could happen with this form, but only the checkbox about the trip being observed would be editable and the rest of the details could be shown in plain text as opposed to editable form prompts.

Schedule a Trip

* = required field

Vessel:	<input type="text" value="SELECT VESSEL"/>	▼	*
Vessel Operator:	<input type="text" value="SELECT OPERATOR"/>	▼	*
Departure Port:	<input type="text" value="SELECT PORT"/>	▼	*
Departure Date:	<input type="text"/>		*
Fishery:	<input type="text" value="SELECT FISHERY"/>	▼	*
Trip Type:	<input type="text" value="SELECT TRIP TYPE"/>	▼	*
Trip Observed?:	<input type="checkbox"/> YES		

Save Trip Schedule

Database Table(s)

These are the required and optional tables needed for trip scheduling – several are defined earlier:

- Trip Schedule (required)
- Trip Type (required)
- Fishery (required)
- Port (required)
- Vessel (required)
- Vessel Owner (required)
- Owner Operator (required)
- Owner User (required)
- Species (optional)
- Trip Target Species (optional)

Regardless of a trip being associated with one or many target species or fisheries, a one-to-many data relationship is suggested for storing that data against the trip. Even when current needs only dictate a one-to-one relationship, it is often wise to store the data in a one-to-many relationship in order to “future proof” the application. This makes updating

the system much easier as there would be no need to restructure the database when a one-to-many relationship becomes needed.

Trip Schedule

Column Name	Data Type	NULL?	Notes
Trip ID	Integer	NO	Primary key
Vessel ID	Integer	NO	Foreign key to Vessel table
Departure DTT	Date/time	NO	Date and time trip is scheduled to depart
Port ID	Integer	NO	Foreign key to Port table
Trip Type ID	Integer	NO	Foreign key to Trip Type table
Fishery ID	Integer	NO	Foreign key to Fishery table
User ID Created	Integer	NO	Foreign key to the User Base table – the user that created the record
User ID Last Update	Integer	YES	Foreign key to the User Base table – the user that last updated the record
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

NOTES:

1. The Trip Type table could also have a foreign key to the Fishery table in the event trip types are fishery specific and not all types are valid for each fishery
2. The Species table could also have a foreign key to the Fishery table in the event species are fishery specific and not all species are valid for each fishery. This may depend on each region's stance on whether to allow the reporting of species not to be harvested in a fishery vs. wanting to see the event reported so they can communicate the proper rules to the vessel operator post trip.
3. The two user ID columns are meant to track the creator and last updater of the record. This is useful when potentially vessel owners/operators, administrators, and observers could make changes to a scheduled trip.
4. As mentioned earlier, it may be wise to allow observers to login and flag trips as observed that they are scheduled to observe. An alternate method is to have observer scheduling systems automatically update the registry (via a webservice provided by the registry). That approach would then automate the step. If feeding observer sub-systems, a Boolean field ("Is Observed") could be added to the Trip Schedule table ("false when the trip is recorded and could be set to "true" if the trip is to be observed). This would not be required for fisheries that are 100% observed as observation could be assumed. It would be done in partial coverage scenarios and done to make sure it is known an observer report is expected for the trip at some point. If additional details about the observer (maybe the ID of the individual observer, or just to know which of possibly many observer contracting companies is

doing the observing) are desired, then observer user accounts could be expanded to contain this data. This is best accomplished by adding an “Observer Details” table that would be related to the User Account table (one-to-one relationship) via the user ID. This allows all required columns to be required for observer data without impacting the required data in the User Account table (i.e. do not have to allow NULL in observer only columns should they be added to the User Account table). With this setup, the observer account user ID could be recorded in the Trip Schedule table and would be populated only when an observer user edits the trip schedule. Although it may not need to be said, but an observer (or scheduling system) would NOT be able to edit any details about the trip schedule other than the flag that indicates the trip is to be observed or not.

3.5.6 Submitting Trip End Notifications

This functionality is meant for both the administrator and vessel operator roles should it be needed (more on why below).

Trip end notifications are meant to be sent from the vessel either while at sea or once landed at port. These notifications in addition to the departure date/time in the trip schedule serve to “bookend” the trip so that both the departure and completion dates/times of the trip can be known. While this does not aid in the lookup of trip IDs for active trips, it provides a formal end of the trip that will aid in looking up trip IDs for past/future trips with just the VRN and a date.

Each region will need to assess its needs and geography to determine what communication methods of sending trip end notifications are appropriate and whether they need to be sent from sea or port.

An application that contains an electronic form to send the notification as an e-mail, as an e-mail attachment, or via a webservice to the trip registry would be acceptable for trip end notifications. This application could be integrated with an e-logbook application, a standalone application, integrated into the mobile application suggested later to handle unique identifiers, or part of a VMS system or software that uses the VMS system to communicate data to shore.

Trip end notification needs will most likely differ from region to region, but should have some standard core details as follows:

- Trip ID (acquired prior to departure from trip registry)
- Landing port (if needed in advance of landing – if not it can be derived instead)
- VRN (not actually sent – trip ID identifies the vessel already)
- Trip end date/time (auto generated by the application – not user entered)
- Dealer selling to at landing port (optional – may only be needed if dealers require advanced notice of a vessel coming to sell – that can of course be handled between the dealer and vessel operator outside of the trip registry)

Sample Screen Mockup

There are many ways in which these notifications could be sent (see above) and the interface will be very basic in all cases. It simply conveys the trip ID (not displayed) – sent with the other data, the intended landing port, and the anticipated date/time of landing.

It would be very much like a stripped down version of the trip scheduling form already provided. For this reason no sample screen has been provided.

Sample Database Tables

There is one required table and one optional table needed in the trip registry to support trip end notifications. Trip End Notification is required and Dealer is optional. The previously defined Trip Schedule, Port, and User Base tables are also used as they hold ID columns which will be used as foreign keys in the Trip End Notification table described below.

Trip End Notification

Column Name	Data Type	NULL?	Notes
End Notice ID	Integer	NO	Primary key
Trip ID	Integer	NO	Foreign key to Trip Schedule table
Port ID	Integer	NO	Foreign key to Port table (landing port)
Dealer ID	Integer	YES	Foreign Key to Dealer table (optional)
Trip End DTT	Date/time	NO	Date and time trip is scheduled to land/end
User ID Created	Integer	NO	Foreign key to the User Base table – the user that created the record
User ID Last Update	Integer	YES	Foreign key to the User Base table – the user that last updated the record
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

3.5.6.1 Alternate Approaches to Determining the End of a Trip

It appears that the requirement of trip end hails needs a good long look as there seems to be little need for them. With e-reporting in place and a trip registry in use, most reasons to send a trip end notification will fall away. There can be some benefit to them as they can help to communicate landing information to dealers, dockside monitors, and electronic monitoring (EM) technicians. The benefit is advance notice of the vessel getting to shore in order to schedule a sale, have offloading monitored, or EM equipment serviced respectively.

If this kind of scheduling is seen as a reason to require trip end hails, it may be time to reassess that line of thinking. Those kinds of notifications don't seem to require the involvement of NMFS and are instead private business conversations between the parties mentioned. Those parties should be left to manage their affairs as they see fit (while following regulations of course).

There is nothing wrong with allowing for trip end notifications as part of e-reporting systems (perhaps part of an e-logbook application). After all they could allow one communication to be leveraged by all three parties mentioned. Making trip end notifications mandatory seems to be going too far though.

Here are some alternative approaches to consider that allow it to be known when a trip is over without anyone sending a trip end notice.

One alternate method of determining the end of a trip was put forward in section 2.4.3 above. This method involves using VMS GPS track data and the known VRN and departure date to determine the landing port and the trip end date and time without the vessel operator sending anything. Using this method will depend on many factors and each region should assess whether this is tenable. It has been put forward as it reduces reporting burden and will be more accurate than what is currently collected via trip end hails/notifications.

A 2nd method was suggested earlier and involves the trip ID acquisition dealers will undertake when purchasing harvest. The fact that a vessel's operator transmits the trip ID to a dealer will be recorded against the trip in the registry via the mobile application the dealer uses. Once it is known the vessel has initiated the sale of harvest, it is safe to consider the trip is over (but may still sell to another dealer). This is the suggested method as it is very concrete evidence of a trip being over and is less likely to be impacted by technical issues (a dead VMS unit would ruin the first suggestion above).

3.5.7 Trip Schedule Reporting

This functionality is meant for both the administrator and vessel owner roles. It could also be accessed by observers if it is decided they will be marking scheduled trips as being observed.

The functionality is fairly basic in that it is an ad hoc report with basic search criteria. A user would fill in the criteria as they see fit (within the options presented) and press a "search" button to produce a paginated report that they can browse through.

Most of the criteria will be the same for all roles, but a few extra criteria can be added to help administrators and possibly observers to find the trip(s) they are looking for. Vessel owners/operators should be restricted to only searching trips that they have scheduled or are listed in respectively. Administrators should be able to search all trips past and future. Observers may need to be locked down to only seeing a subset of trips (perhaps only certain fisheries or trip types).

An optional table will be listed to show how data can be structured to restrict observer access as suggested. It is only one example, but the concept is flexible and regions can adjust as and if needed.

The columns in the search results to be displayed should be assessed by each region in terms of data privacy. It may be that one or some of the roles will see more/less/different columns than the other role(s).

A main use of the report will be to find and interact with already scheduled trips (that have not yet departed) and to review details of past trips, an "action" column could be added to hold buttons to navigate to the trip. There should be an "edit" button (only shown for trips yet to depart) and a "view" button for all trips that take the user to a read only view of the trip schedule.

For clarity it should be mentioned that if data corrections are ever needed post trip, those would occur via two main methods. The first would be administrative interfaces to make

changes for common issues that follow known patterns. The second would be via direct manual changes to the database less common or more complex fixes.

Sample Screen Mockup

Trip Schedule Report

The search results could include all the same columns as the search criteria that are listed. A sub-set has been shown simply to ensure this image fits within this document. It should be noted that the results represent both active and completed (over) trips. This is to demonstrate that this report can also be used to edit the active trip schedule (perhaps by the vessel owner, and observer, or the vessel operator as detailed already).

The results are also an excellent way to allow for a vessel owner to be given access to all of the data collected on the trip (a single trip report). This is described right after the sample database tables for this section.

This search and its criteria could be modified based on the user role that accesses it and that role's needs and provide a great deal of utility.

Trip Schedule Reporting

Use the search criteria below to find the trip(s) you are looking for

Vessel:	<input type="text" value="SELECT VESSEL"/>	▼
Vessel Operator:	<input type="text" value="SELECT OPERATOR"/>	▼
Departure Port:	<input type="text" value="SELECT PORT"/>	▼
Departure Date:	<input type="text"/>	
Fishery:	<input type="text" value="SELECT FISHERY"/>	▼
Trip Type:	<input type="text" value="SELECT TRIP TYPE"/>	▼
Trip Observed?:	<input type="checkbox"/> YES	

Search Results

VESSEL	OPERATOR	STATUS	ACTION
Vessel 1	Smith, John	ACTIVE	VIEW/EDIT SCHEDULE
Vessel 4	Jones, Sarah	OVER	VIEW SCHEDULE

Database Tables

The tables to drive this search functionality have already been listed above in other sections with the exception of an optional table should observers not be allowed to view and interact with every trip in the region (most likely they will be locked down to only fisheries and trip types that require or possibly require trips to be observed).

The tables that have already been defined include:

- Trip Schedule (the core table that holds all scheduled trips – past and future)

- Vessel Profile (to retrieve vessel details from)
- Unit of Measure (to retrieve the vessel length units from)
- Trip Type (to retrieve the name of the trip type for the trip)
- Species (to retrieve the target species name(s) from)
- Fishery (to retrieve the fishery name from)
- Port (to retrieve the departure port name from)
- User Account (to retrieve the name of the person that scheduled the trip)
- User Role (to retrieve the role of the user that created/edited the scheduled trip)

In order to setup data that can be used to indicate what trips an observer can interact with (i.e. to set a flag to indicate the trip will be observed), the Trip Type, Fishery, and User Account tables are used as well as the newly defined Observer Access table below.

The Observer Access table simply relates the observer role with the fishery and trip type combinations (using their unique codes to avoid data change issues over time as discussed earlier) that the role is allowed to flag a trip as observed for. When/if an observer searches the trip registry; they will only be allowed to see trips in the results where the fishery and trip type is listed in the Observer Access table.

The data can also be used on the trip schedule form to ensure observers can only view trip schedules they are authorized to see. This can be done by adding validation when the form loads. This means that even if an observer managed to get to a trip they shouldn't (perhaps via a bookmark or by altering the URL), they would not be allowed to see the form and instead would just see an error on the page stating that they do not have correct permissions to view the page.

Observer Access

Column Name	Data Type	NULL?	Notes
Access ID	Integer	NO	Primary key
Trip Type Code	Varchar	NO	Unique trip type code
Fishery Code	Varchar	NO	Unique fishery code
Role Code	Varchar	NO	Unique role code (for the observer role)
Active DTT	Date/time	NO	Date/time record becomes active
Inactive DTT	Date/time	YES	Date/time record became or is to become inactive – if known
Create DTT	Date/time	NO	Date/time the record was created
Modify DTT	Date/time	YES	Date/time the record was last modified (NULL when created)

3.5.7.1 Extending this Functionality – Single Trip Report

Once trips can be searched and found, a natural extension to consider is to provide access for vessel owners to see all data (or all data deemed appropriate to show them) about each past trip they have completed. This allows for a single trip report with sections that cover all of the logbook, dealer, VMS, and observer (should the trip have been observed) to be displayed.

This sort of functionality would be more difficult with the logical equivalent model due to having to reach into each sub-system to pull the data into this sort of report, but with true integrated reporting it is straightforward.

Harvesters often have a need to see exactly what data the government has about the trips they have taken. If both the harvester and the government can view the same report, this will fulfill that need and help resolve data quality issues.

3.5.8 Communication and Handling of Trip ID and VRN

This functionality is meant for both the administrator and all data collector roles. There isn't a real world need for administrators to have access, but having access will help facilitate testing. Alternately administrators can use special vessel owner/operator testing accounts to facilitate this kind of testing. Those accounts would need to be generated based on dummy test data in all the NMFS loaded tables – that data is best to be flagged as being for testing purposes only so it can be excluded from real production data.

The main purpose here is to ensure the accurate transmission of critical unique identifier data (VRN and trip ID) from the trip registry to the parties that need that data.

The proposed uses are as follows:

- **Transfer of trip ID to an electronic logbook.** Vessel operators need to get the trip ID from their mobile device to an e-logbook application if it is not running on the mobile device that acquired the trip ID already.
- **Transfer the trip ID to each dealer sold to.** Vessel operators will need to transfer the trip ID to the dealers they sell to in order for the dealer(s) to be able to report the trip ID along with the details about the purchase they collect. The trip ID is transferred from the vessel operator mobile device to the electronic dealer system (most likely running on a laptop or PC, but could be on a mobile device).
- **Transfer the trip ID to an electronic observer application.** Observers would be able to acquire the trip ID from the vessel operator in the same way dealers could. The trip ID could then be transferred to the electronic observer application and would be submitted with the observer report post trip.
- **Transfer the trip ID to VMS data once identified.** As the VMS data is made available and the logged GPS track records can be identified for the trip, those records can be tagged with the trip ID. This can happen at the database level and does require a user interface or special application – unlike the three uses already listed.

To facilitate all but the VMS transfer of the trip ID, a mobile application can be built that would work on tablets and smartphones. This application would serve the user roles that will require it – vessel operators (who can also be owners), dealers, and observers. Each user role would have access to different functionality as is the case in the trip registry – in fact logging into the application would be logging into the mobile side of the trip registry.

The application would have the following functionality to support the transferring of data listed above as well as to perform certain trip schedule corrections from the field.

To start it should be made clear that the vessel operator would follow these steps after a trip has been scheduled and before departing on the trip while an Internet connection is available to the mobile device they are using.

Vessel operator steps prior to departure:

- Vessel operator logs in to mobile application
- The application interface shows a “Get Trip ID and VRN” button (one of a few buttons) to be used to request the active trip ID and associated VRN for the current trip the operator is associated with
- The button is clicked and the trip ID and VRN are encrypted and the encrypted strings are copied to the mobile device for later use
- The vessel operator is then prompted to verify the vessel identity. They can find the RFID chip or barcode affixed to the hull and scan it using the mobile application. Once scanned, the mobile application can quickly compare the known VRN from the registry and the scanned VRN. If a mismatch is spotted, the mobile application will provide a way for the vessel operator to correct the vessel associated with the scheduled trip in the registry (assuming they know they are at the correct vessel meant to be used on that trip). If they are in fact at the wrong vessel, they would be able to indicate that to the registry via the mobile application. That would reset the state of their vessel verification process and they could re-initiate that once at the correct vessel.
- Vessel operator logs out and departs on the trip as planned.

Note: The vessel operator will NOT be able to transfer the acquired trip ID or VRN to a dealer or an observer if the trip registry sees that the VRN verification (by scanning the RFID chip or barcode on the hull) has not yet been completed by the operator. The operator will need to perform that step and handle any vessel mismatch issues (as described above) before proceeding.

Transfer of trip ID to an electronic logbook:

- Vessel operator starts a new trip in their electronic logbook software (which could be running on a tablet that also runs the trip registry mobile application or on a PC/laptop).
- Part of the starting of that logbook involves the transfer of the trip ID to the logbook software.
- The mobile device can transfer the encrypted string to the laptop, PC, or tablet running the logbook software via various means:
 - Bluetooth file transfer directly to the logbook software
 - Mobile device can be connected (as a mass storage device) to the device running the logbook software which can then browse and find the file to be transferred on the mobile device and the file will be loaded into the logbook software
 - Near field communication (NFC) file beaming
 - There are other options, but the above are the most convenient
- Once the logbook software has the file containing the encrypted trip ID, it can extract the encrypted string and add it to the data the logbook software collects for later transmission.
- The above steps can be streamlined if the logbook software is running on the same mobile device that acquired the trip ID. In that case there is no step to transfer the

trip ID between devices. As all of the software being discussed could be built by different developers, standards need to be determined so that an acquired trip ID can always be located in the same place so that any software running on the same device can access it with ease. Standards like this are very important as there can also be many authors in a region for software that performs the same task (i.e. there can be multiple e-logbook developers and they'll need to follow the specifications designed to allow for interoperability regardless of the author.

Transfer the trip details (ID and possibly VRN) to each dealer sold to:

- When at a dealer, the encrypted trip ID (and VRN if still needed for the dealer report – to be clear though, the trip registry already has that detail and it does not need to be collected by dealers unless they need it themselves) can be sent to the dealer's device or computer on which their e-dealer reporting system resides. This can be triggered by pressing a "Send Trip ID to Dealer" button which becomes active once the two devices recognize they have paired with each other (i.e. vessel operator device knows it is connected to a dealer device/computer because each device broadcasts what mode it is in (dealer or operator))
 - The trip ID will be transferred from the vessel operator device to the dealer device/computer via Bluetooth, near field communication beaming, or scanning of an onscreen bar/QR code on the vessel operator's device by the dealer's device.
- The comment about standards put forward in the section above about transferring the trip ID to the electronic logbook software applies here as well.

Transfer the trip ID to an electronic observer application:

- This can be accomplished in the same way the trip ID can be acquired from the registry by a vessel operator and then transferred to their electronic logbook application.

Sample Database Tables

There are no new tables required to facilitate this functionality.

3.6 Impacts to Consider

With all system design, there are many things to consider and what is mentioned below should help with that effort. The reality is each region is different and will have different needs due to how they currently do things, political pressure, financial and technical resource availability, and so on.

Below are items that should be appropriate for most regions and thus are brought forward just like core data needs for the trip registry have been mentioned above.

Changes to regulations will be needed to require all commercial trips to be registered.

It is understood that this may be a difficult change to request of the industry. In some regions, very few trips are known about ahead of time unless they are being observed or are undertaken in a catch shares managed fishery. The benefits of such a change should be made clear.

Some benefits to put forward are:

- **Less reporting burden while fishing.** This comes in the form of having the registry report the trip to whomever needs notification (as opposed to hail and trip declarations), not having to report vessel or vessel operator details (already in trip registry), having less far less data mismatches of varying forms – meaning less time spent post trip dealing with data issues.
- **A true step towards real traceability.** Traceability has been shown to increase the value of harvest once sustainability can be proven. It can't be proven without knowing about every trip and ensuring it is properly monitored. Harvesting and the verification of harvest is the first step in the supply chain after all.
- **Easy access to reported data.** As described earlier, the trip registry is an excellent avenue to make a single trip report that encompasses logbook, dealer, observer (if observed), and even VMS GPS track data to the vessel owner/operators.

Trip definitions may need to change. There are usually legal and regulatory definitions of trips in each region. These need to be reviewed and most likely one or both will need to be changed in order to facilitate greater control over the flow of commercial fishing trips (to facilitate high levels of accuracy) or to identify custom requirements of the trip registry functionality for the region to accommodate differing trip definitions.

Most of the data sets involve data collection on/from the vessel during the trip (observers, VMS, and logbooks). It is dealer purchases that can be more variable in that they can occur directly post trip or in some cases possibly days later (meaning a second trip for the vessel may start before the catch from the first is sold). The data collection that occurs on the vessel during the trip is more easily related to the correct trip because those collecting the data are on the vessel and know when the trip has concluded.

Situations where not all harvest is sold prior to a new trip starting are currently happening and are quite problematic. The cleanest regulatory solution is to require the fish to be sold before starting another trip. Some regions may object to this restriction but it should be made clear that the consequences are significant and they come in the form of more complex systems. This means more variation in rules to be explained to fishers who already have a hard time understanding all of the caveats they are asked to learn in order to fish in the fisheries they deal with. It also means higher system design and maintenance costs for a system with a greater chance of inaccuracies.

Regulations and system design must go hand in hand in order to work properly.

Professional systems designers that already intimately understand the system(s) in use in a region should be at the table as new regulations are discussed. This is to ensure that if the regulations come into effect, that the system(s) will be able to handle the changes without sacrificing data quality, will not add undue burden to any data reporters, and that the cost to implement and maintain is understood.

To put this into easy to understand terms – you would not tell someone you could add wings to your car without checking with a mechanic first.

Adjustments will be needed to support a move to near real-time data collection. As mentioned near the beginning of section 3.5.1, a move to the solution being proposed means some significant changes to the flow of data from batches to single reports in near real-time as well as the need to allow for all the forms of proactive data corrections that

need to be determined and addressed. The proposed trip registry has been designed to be easy to use and to address some such errors (most notably a vessel mismatch between trip registry and the actual vessel used on the trip).

Some areas to consider include:

- Establishing a help line may need to be established to field calls from trip registry users that may require assistance (this is why the administrator user role is of use).
- Methods of back entering data known to be missing
 - If a vessel operator arrives at a dealer with no scheduled trip, the mobile application mentioned could be used for them to enter the missing trip schedule. Alternately the vessel owner could be contacted and asked to enter the missing trip and a method added to the mobile application could allow the vessel operator at the dock to retrieve the trip ID.
 - With the trip ID being common to all data sets and dealer reports required to be submitted much faster than is now the case (say within 24 hrs of the purchase) it will be easy to know if dealer reports are missing for a given trip. The vessel can be contacted to indicate who the sale(s) were made to and those dealers can be asked to provide the missing data.
- Tracking revisions to data is critical in large complex data systems. There are many reasons and times at which data may be altered and knowing who altered what and when is very important. This comment is not just about reported data, but also core critical data as laid out in the section detailing tables to be loaded by NMFS. The NMFS loaded data is used to know what vessel/operator/owner/observer/port(s)/dealer(s) were involved with the trip. That data may have minor or major changes over time (consider the change of a ports name). When reporting on past trips, the data that was active at the time of the trip is to be displayed. So if the departure port involved had its name changed after the trip occurred, the report should show the old name and not the new one. This is why effective dating changes to data is so important. If these changes are not tracked then all that is known in the future is the current name of the port, not the name as of the time of the trip. Now a port name may not be critical, but the unit of measure associated with a quantity of fish is!

VMS may need to be required on all vessels. There have been some novel uses of VMS GPS track data mentioned in this report. They were put forward as different ways to think about this data stream and how it could be used in new ways to reduce reporting burden and increase accuracy. As these are suggestions and not an essential part of the proposed trip registry, this impact will be mitigated if a region opts not to use any of the VMS suggestions.

That said, perhaps it should be considered that the existing VMS infrastructure has a great deal of rigor built in to the type approval process for VMS vendors. A not insignificant part of the reason for this is VMS units must be tamper proof so the data produced can hold up in court should it ever need to.

Are there alternate ways of recording GPS tracks on commercial trips outside of the current VMS model?

In fact there are and they will become easier and cheaper as location aware devices continue to pop up on the consumer electronic landscape. One option that currently exists is the GPS data logger the Gulf Shrimp fisheries use to track a vessel's activity from port to the fishing ground and back to port again. It uses a cellular connection to send the data to shore automatically when the vessel is in cell range and identifies the vessel in the data it transmits to shore. This may very well occur soon enough that the OLE could intercept the vessel when it lands – making the current VMS systems obsolete.

Using a far less expensive data logger or similar device could replace the need for VMS GPS track data in fisheries management (it will most likely still be needed for other regulatory and legal reasons – but not on all vessels as is the case now).²

² Baker Jr, M. S., M. Sciance, and J. Halls. (In Press). Potential for a simple GPS-based binary logit model to predict fishing effort in a vertical hook and line reef fish fishery. *Marine and Coastal Fisheries*. Vol. 8, Iss. 1.

In fact, a similar device could be created and have the added capability of being able to interact with a mobile device. This interaction would allow a vessel operator to transmit an acquired trip ID to the device via the same mobile application the vessel operator uses to acquire and transmit the trip ID. The logger device could send status messages back to the operator's mobile application to confirm proper receipt of the trip ID (this is required as there is no monitor connected to such a basic logging device).

When considering alternate technology, what seems like a regulatory hurdle could become a low cost solution that provides many benefits in fisheries data collection.

3.7 Methods of Unique Identifier Handling for Paper Reporting

Once again – paper reporting is not being condoned as it can't constrain and validate data entries at the source (which makes QC efforts too unwieldy). It is understood that it may still need to be used while transitioning to a logical equivalent model that leverages full electronic reporting in all data streams. The following advice assumes that a trip registry is put in place but some or all of dealer/logbook/observer data is collected on paper.

The suggestion is to place a barcode or QR code on the logbook, observer, or dealer report forms that identifies the trip ID from the trip registry. Retrieving and printing the barcode or QR code can be accomplished in many ways.

Some of these ways are:

- **Printing from the registry** to a standard label using any device that can access the registry and connect to a suitable desktop, mobile, or label printer. This could also take the form of an Internet enabled kiosk at each port for retrieving and printing the trip ID codes.
- **Use of mobile technology as already** mentioned to retrieve and transfer the trip ID and VRN. The major difference being any of the steps to transfer the trip ID to an electronic reporting system. Those steps could be replaced by printing the label and attaching it to report forms

When data is submitted to NMFS, a barcode or QR code scanner could be used to retrieve the correct trip ID for data entry purposes. It should be possible to scan the trip ID value behind the code directly into data entry applications or develop a bridge to do so.

3.8 General Steps to a True Integrated Reporting System

Once a regional trip registry is up and running and stable, the transition to a true integrated reporting system can be undertaken. This transition is all about the incremental migration of the databases used in sub-systems into a centralized (for each region) integrated trip reporting database (which will already be started with the creation of the trip registry).

Data migration will be very straightforward if the sub-system and trip registry databases use the same kind of database (i.e. if they both use Oracle). If the two sides use different databases, then a data migration plan is needed to correctly transform the data from one format to the other. The hardest transition is when the two databases differ in the kind of SQL they use. Most enterprise level databases use what is known as T-SQL (SQL Server for instance as well as many others), but Oracle uses PL-SQL. Migrations from T-SQL to PL-SQL or from T-SQL to PL-SQL will take the most effort.

The transition should integrate the sub-systems that have the least potential for ill effect to regular operations first and then transition the more potentially disruptive systems (logbooks and dealer purchases) – hopefully having learned some lessons with the initial migrations that can aid the later migrations. The observer and VMS integration should be undertaken first as they do not involve industry participation. These two data sets can be migrated at NMFS leisure and time can be taken to make sure all goes to plan. This allows freedom to learn the lessons that need to be learned before making changes to systems that face the industry – those migrations need to happen smoothly and have the greatest chance of encountering issues (so some earlier practice is wise).

By the time migrations occur, each sub-system will already have the concept of shared trip IDs in place and functioning. The data is moved into the integrated reporting database, trip ID foreign key relationships are established (linking to the trip ID in the Trip Schedule table), and the application code for each sub-system will need upgrades to accommodate the new data location and possibly format.

The users of those systems should not have too many changes to what they already do when migration officially occurs and the “switch is flipped” (those came with the addition of the trip registry and the new ways of handling unique identifiers). They still use the same applications they always have, but the data ends up in the new centralized location.

If changes do impact users, it is less likely that data collection efforts are impacted, but is quite likely that the various ways in which the collected data in each sub-system is accessed could be impacted. Most notably would be a change to ad hoc and pre-canned reporting of the data, the way in which data extracts may be provided both internally and externally to NMFS, and the way other data systems may reach into a sub-systems database and retrieve data. Most of these issues are essentially the same as the re-wiring of the application code to account for the new database location and possibly format.

4 Southeast Region

4.1 Regional Overview

The Southeast region covers a large geographic area as well as a large number of states. Dealer data is collected on both paper and electronically, logbooks and observer data are collected on paper, and VMS GPS track data is captured electronically. One exception to logbooks being recorded on paper is the Gulf Shrimp GPS data logger. It is not technically a logbook as it does not record harvest, but it is part of the data that makes up what is considered a logbook data set (it is still a novel use of technology that other regions should review). An electronic logbook pilot has recently concluded (results and feedback are ongoing).

Data storage for the sub-systems is going through a consolidation effort right now in order to reign in some past duplication of effort and artificial separation of like data. This effort is an excellent direction for the region as it starts to move towards electronic reporting in the coming years.

As most regions experience to varying degrees, there is a great deal of political influence on many aspects of fishery management and monitoring. This is exaggerated in the region due to the higher number of states involved than most other regions in the U.S. This means consensus is often harder to come by. As mentioned in this report, ensuring system design professionals are involved earlier in the regulatory change process should help the region to mitigate the political pressure if it can be clearly shown why a change being discussed could negatively impact operations.

The region mostly uses logbooks and dealer purchases for in-season catch accounting and there are various observer programs (some species specific and others broader like all pelagic species). Matching efforts take place using the fuzzy matching techniques in this report with the goal to link logbooks and dealer reports for the same trip together for analysis. When “matched”, the linkage is not stored because the matching is performed as and when needed. The data that supports the matching effort is stored in various databases within and outside of the region. NMFS pulls this data into local tables in order to perform their matching efforts. The data stored externally to NMFS may change (without notice or any clear history of changes in some cases) over time and this is why NMFS pulls in fresh copies each time they need to match the data. The issue of external data changes without preserving history is slated to be addressed soon.

VMS data is used in an ad hoc fashion when needed to determine various vessel activities (i.e. did it visit a certain port and thus dealer? did it fish in a restricted area? Was it at sea during a given time period when it appears there are missing data submission about that trip?). Any linkage to a commercial trip that is derived via that process is not stored for later use.

Observer data is primarily used for discard rate calculation in terms of in-season management (it of course has many other uses – especially in terms of bio samples and other data of value to science efforts).

4.2 Current Data Flows

Below are the current data flows for the four data sets of interest. This information has been provided as a jump off point for making changes and to show the amount of manual QC and data entry effort that exists and could be replaced with e-reporting and a well designed trip registry.

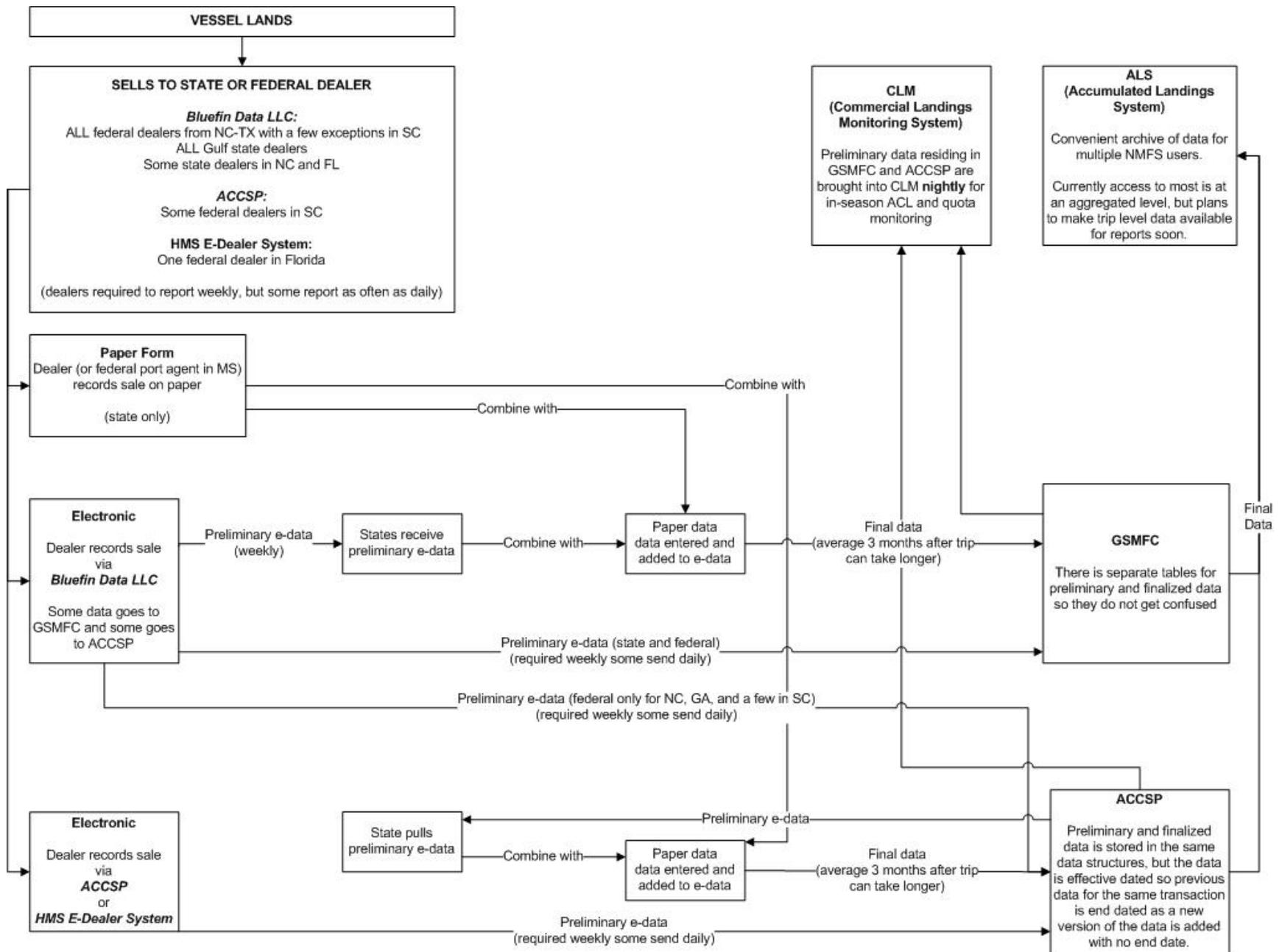


Figure 4: Southeast Dealer Data Flow

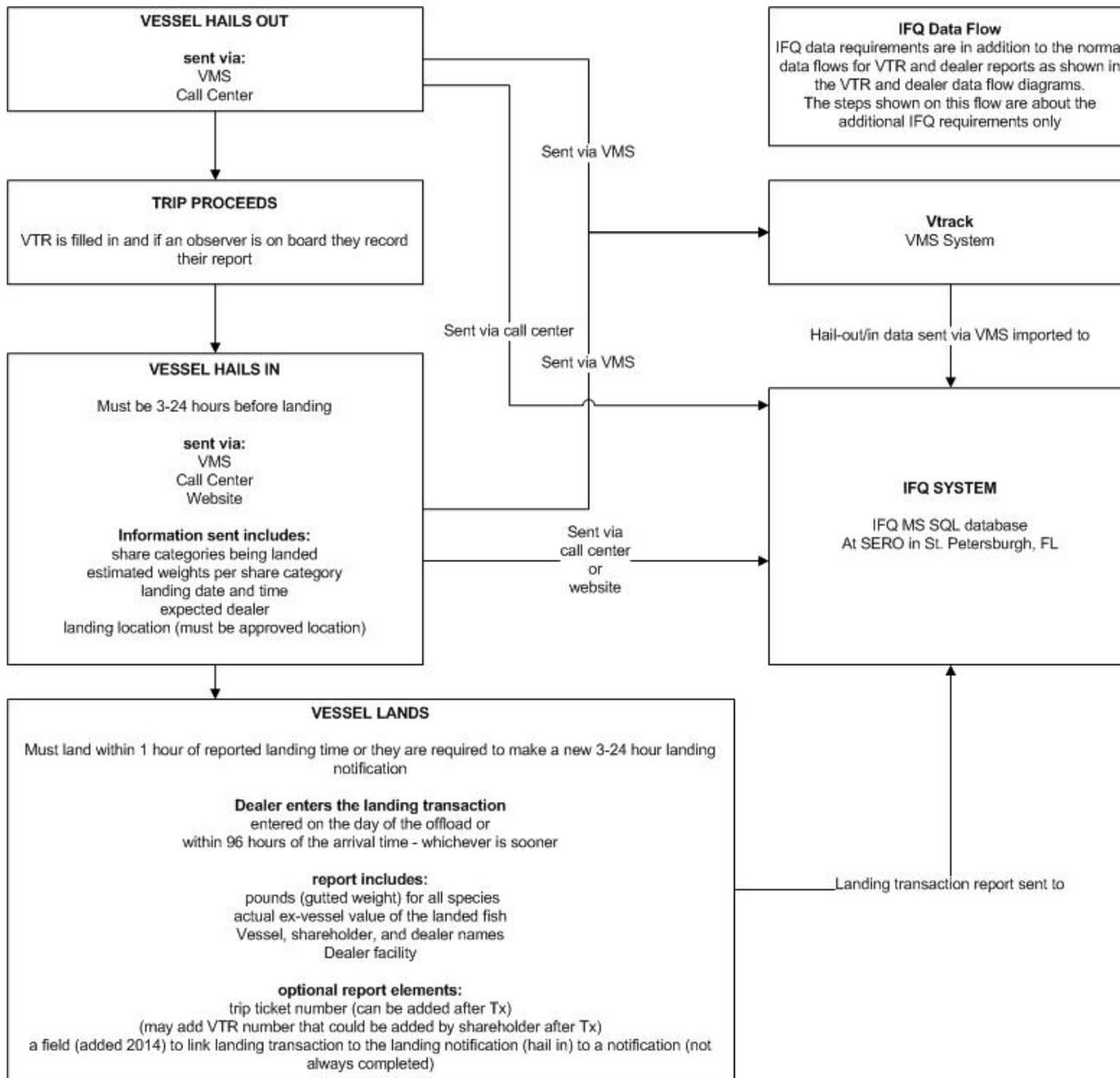


Figure 5: Southeast IFQ Data Flow

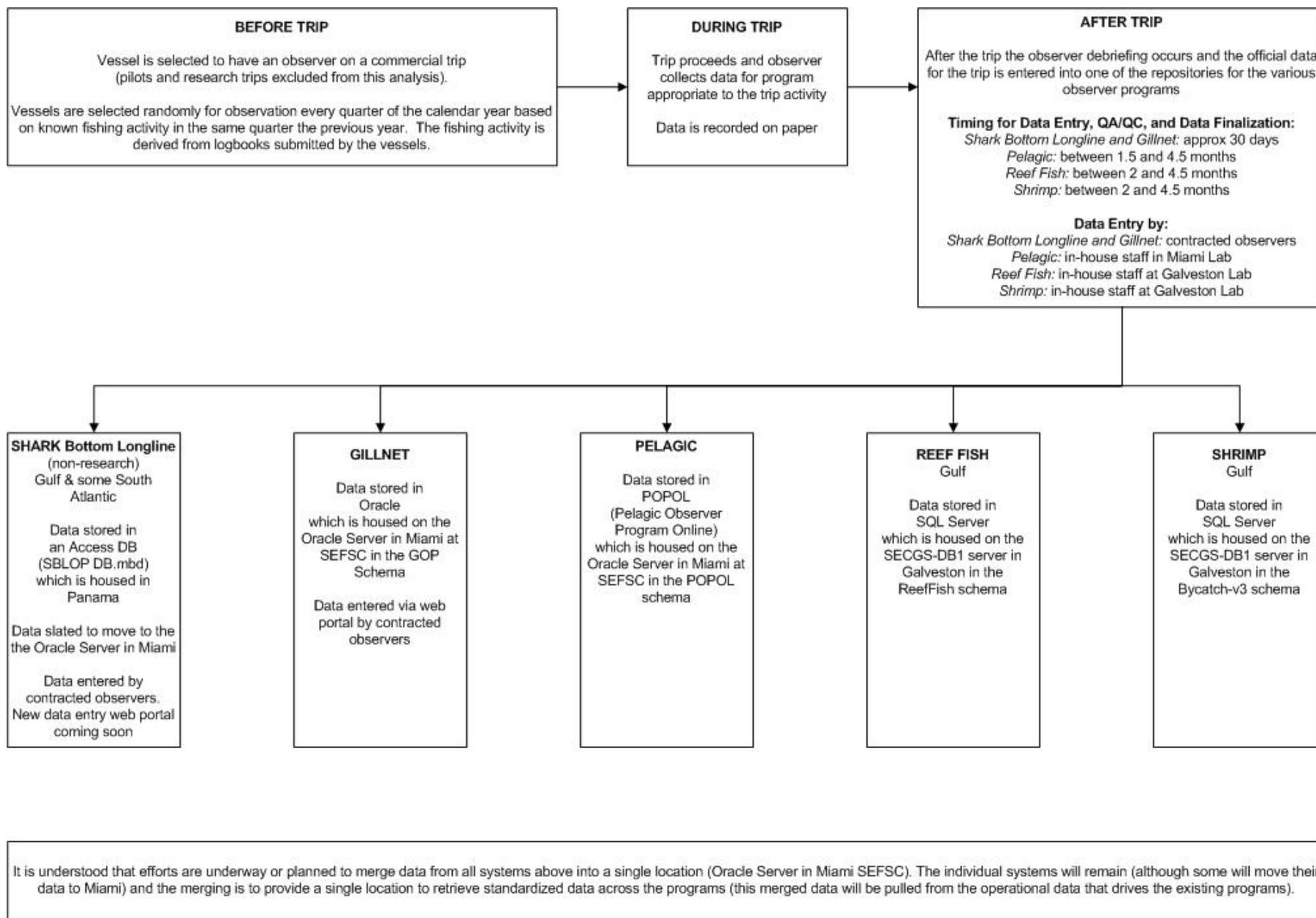


Figure 6: Southeast Observer Data Flow

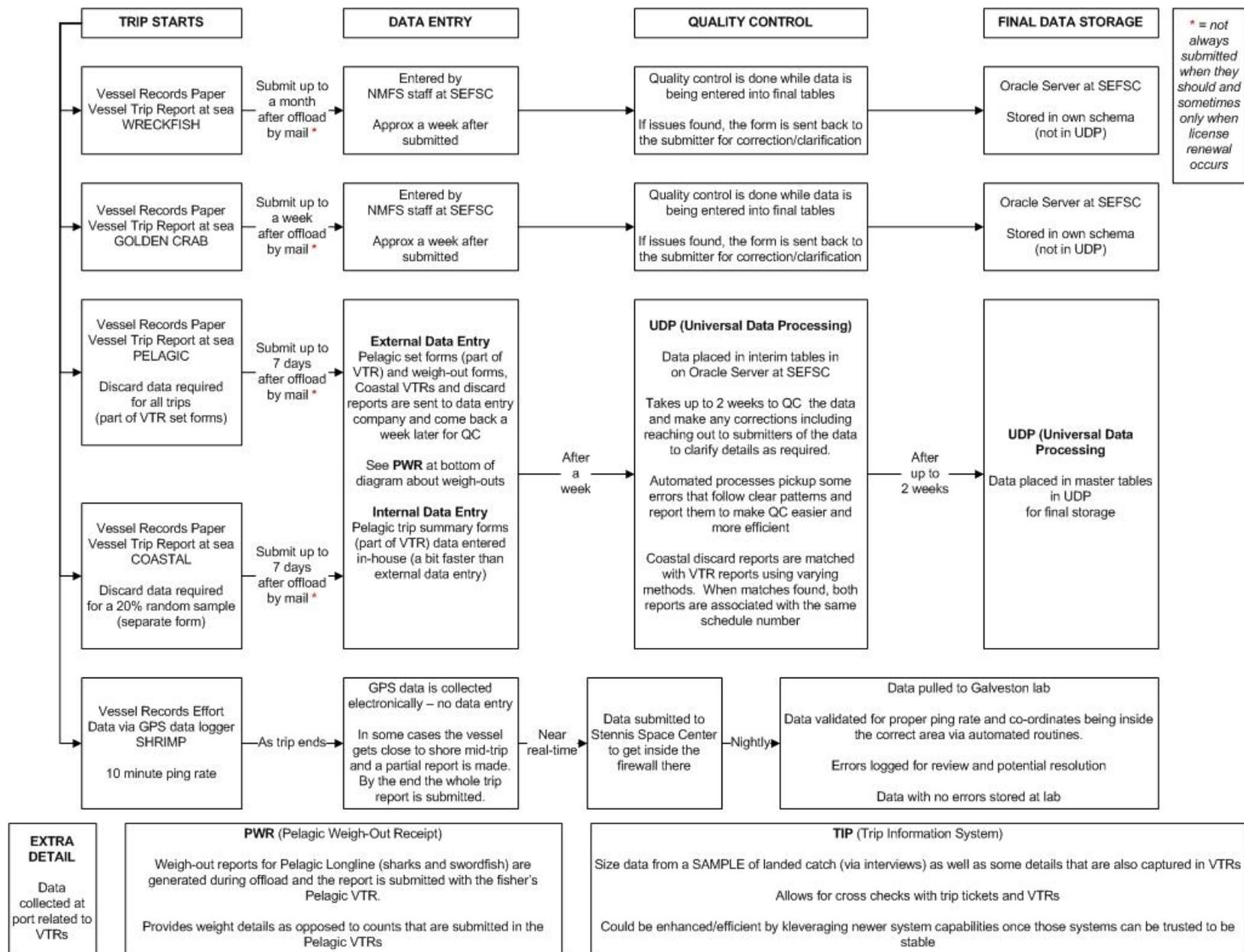


Figure 7: Southeast VTR (logbook) Data Flow

Pelagic longline, purse seine: VMS on 24/7/365 (unless declared out of fishery)
Pelagic bottom longline and gillnet : VMS only required with directed shark permits. VMS on 24/7 during specific times of year and in specific locations (unless declared out of fishery)
Rock Shrimp: VMS required when fishing for rock shrimp (east coast of Florida)
Gulf Reef Fish: VMS required to be on 24/7 (unless a power down exemption has been granted).

NOTES: "Harbor Mode" can be used in some fisheries (not Pelagic) while in port which reduces the ping rate from 1 hour to every 4 hours
 There are other details about VMS usage, but they are IFQ/IBQ specific and are not about the GPS track which is the focus in this data flow



Figure 8: Southeast VMS Data Flow

4.3 Implementation Details and Steps Specific to the Region

The region is on the correct path and should continue to consolidate its data storage and systems. There are known issues with the current method of asking dealers to provide VTR numbers from vessel operator logbooks and these should be resolved by making this mandatory as opposed to optional if possible. This paves the way for the reality of a trip registry system in that details such as unique identifier handling will take a more prominent role.

Once data and system consolidation efforts are complete and perhaps once electronic logs are in regular use, a logical first step for the region would be to introduce a trip registry into the fisheries currently using IFQ. The required reporting timeliness is already increased a fair amount over other fisheries and the vessel operators and dealers are already used to these speedier requirements. A trip registry would aid in the co-ordination of the data efforts needed for the IFQ monitored fishing and it provides a smaller subset of vessel operators and dealers to start with.

Once electronic logbooks get up and running in the region, a second step might be to expand to some of the smaller fisheries like Wreckfish and Golden Crab. From that point it would be a matter of bringing other larger fisheries into the registry like Coastal and/or Pelagic fisheries – or just go region wide at that point.

5 West Coast Region

This region is large geographically, but only encompasses three states. It appears to have consistent handling of each data set across each of the three states which in most cases result in data ending up in PacFIN. There is a solid group of IT professionals on staff which will be very advantageous in terms of assessing the impacts and changes required to introduce a trip registry system.

There is also catcher processor and mother ship operations in this region (in at-sea hake) as there are in Alaska, but not to the same degree.

Dealer data is mostly collected on paper with the exception being seen with the electronic reporting of IFQ Trawl trip landings. Logbooks are collected on paper and again are only used with the IFQ Trawl fishery. Observer data is split between WCGOP using paper data collection and A-SHOP using electronic reporting. The WCGOP program has nearly 100% observer coverage while the A-SHOP program has 100% coverage.

There is co-management of salmon and crab between the states and NMFS.

5.1 Current Data Flows

Below are the current data flows for the four data sets of interest. This information has been provided as a jump off point for making changes and to show the amount of manual QC and data entry effort that exists and could be replaced with e-reporting and a well designed trip registry.

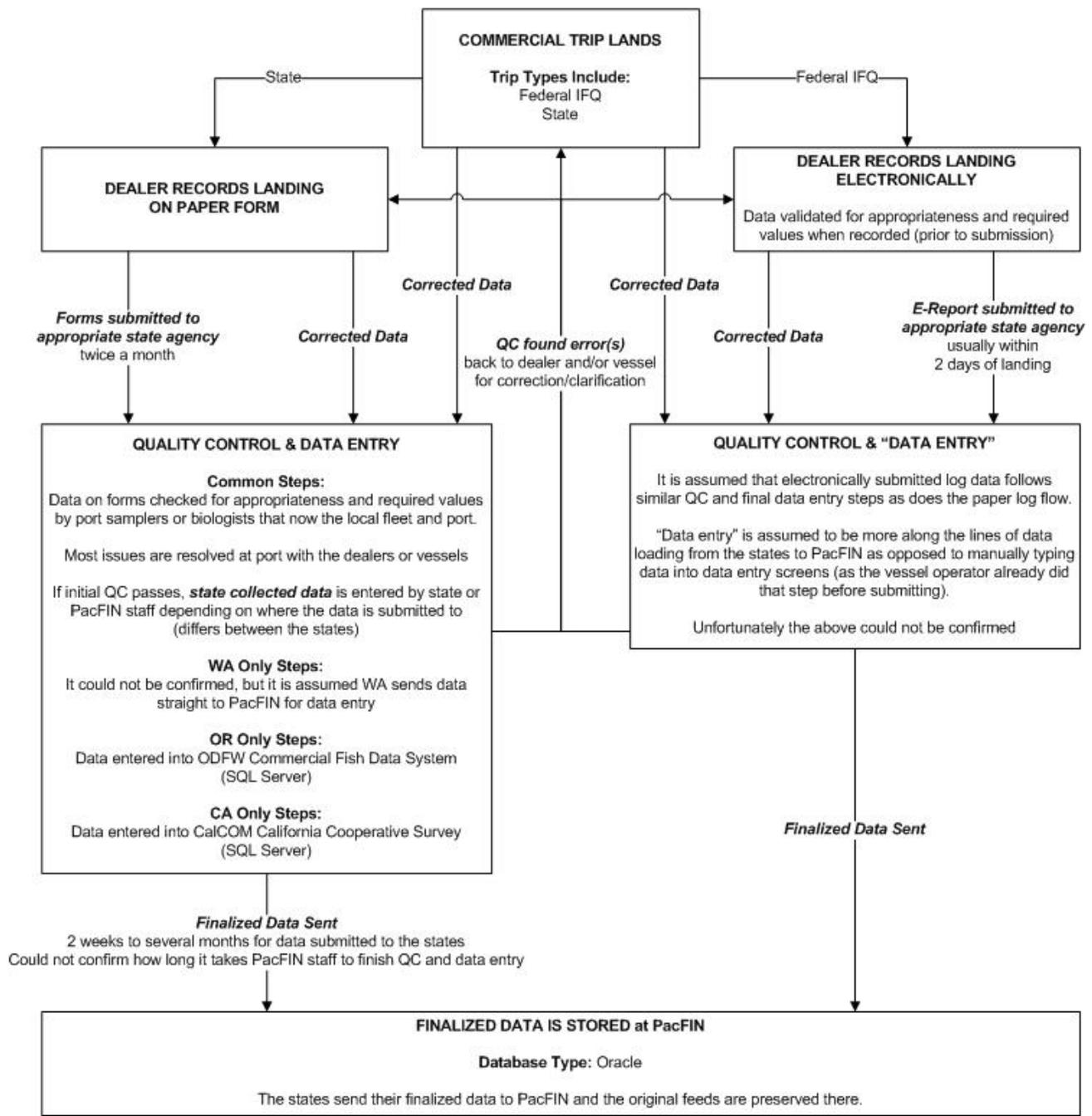


Figure 9: West Coast Dealer Data Flow

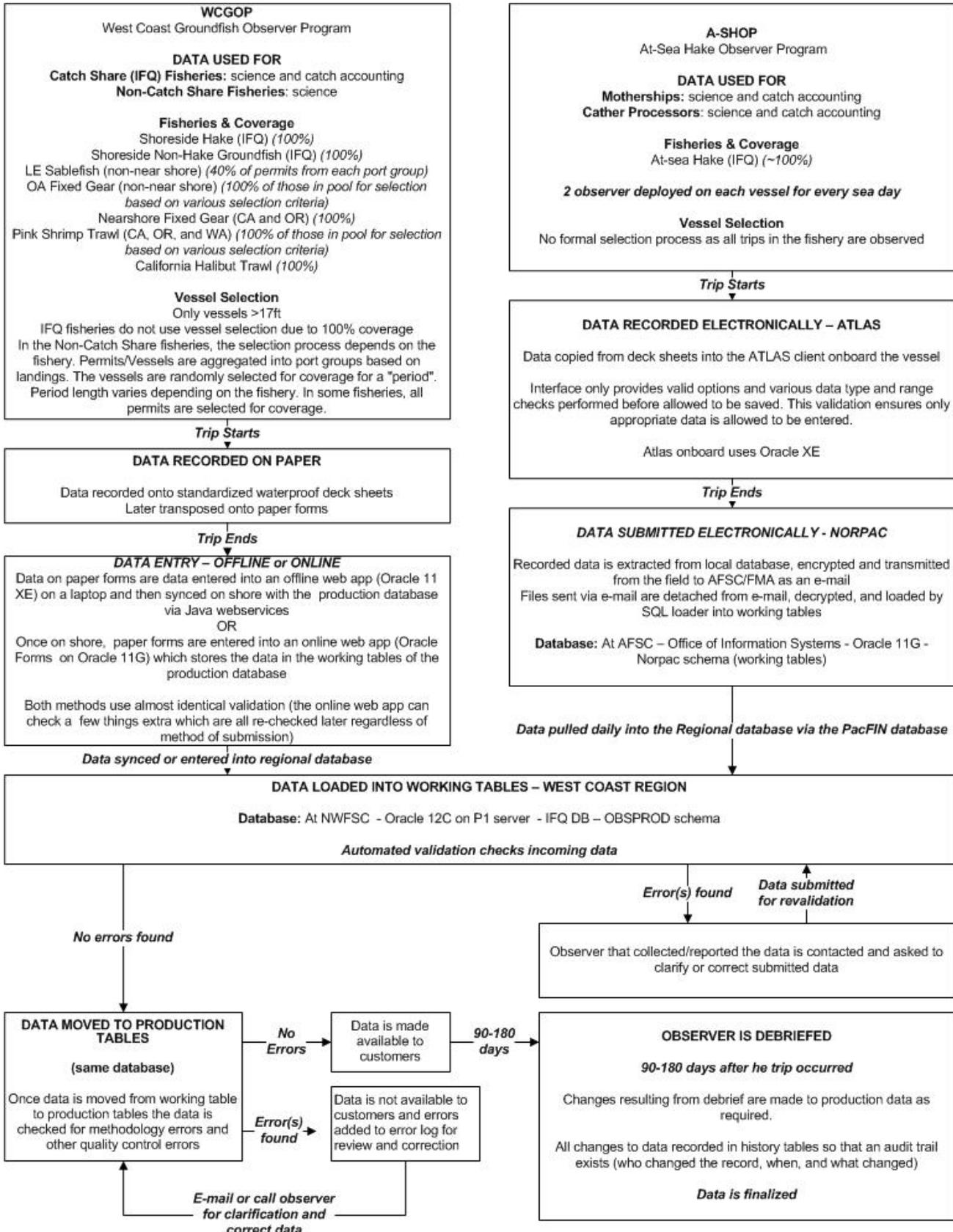


Figure 10: West Coast Observer Data Flow

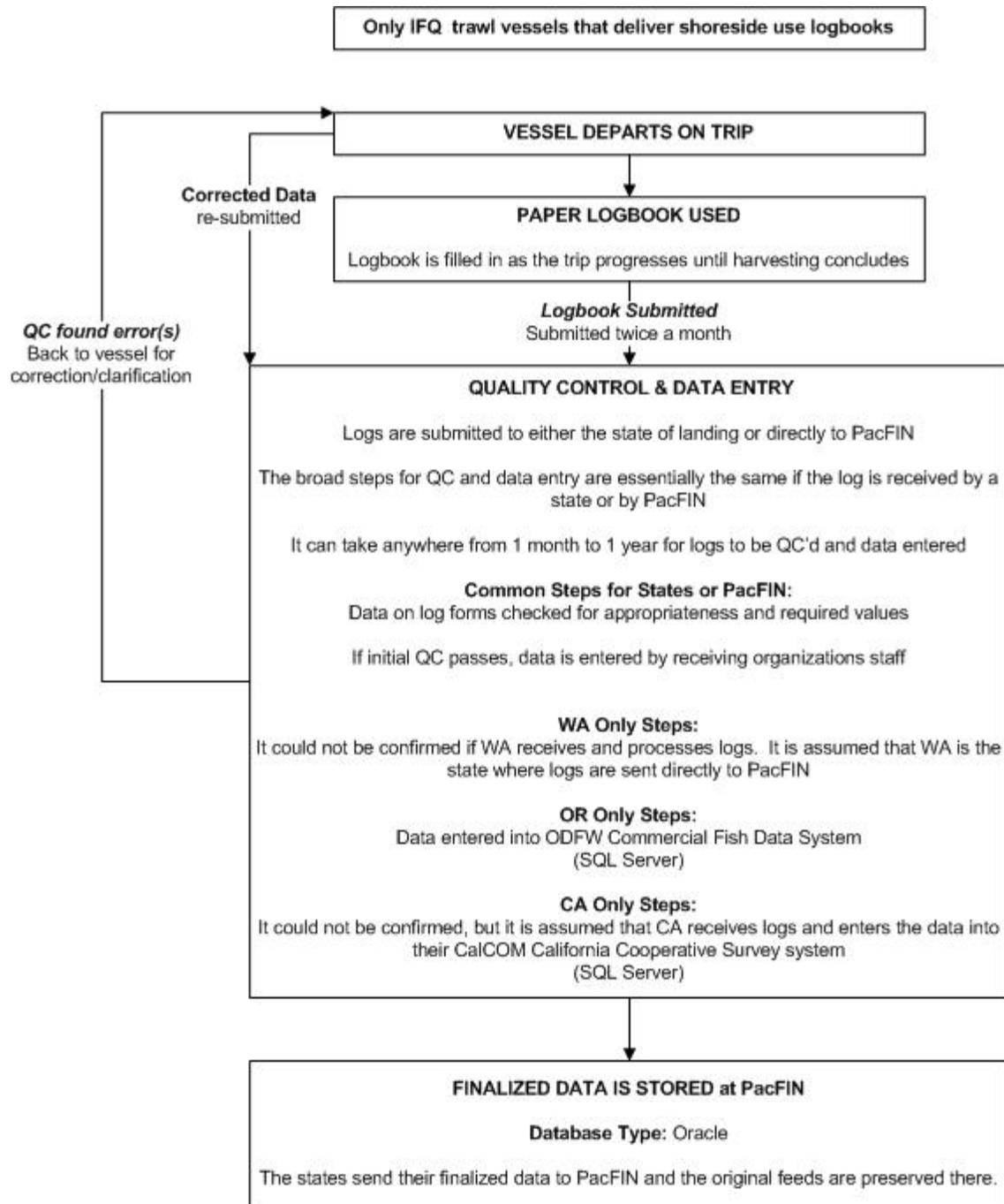


Figure 11: West Coast Logbook Data Flow

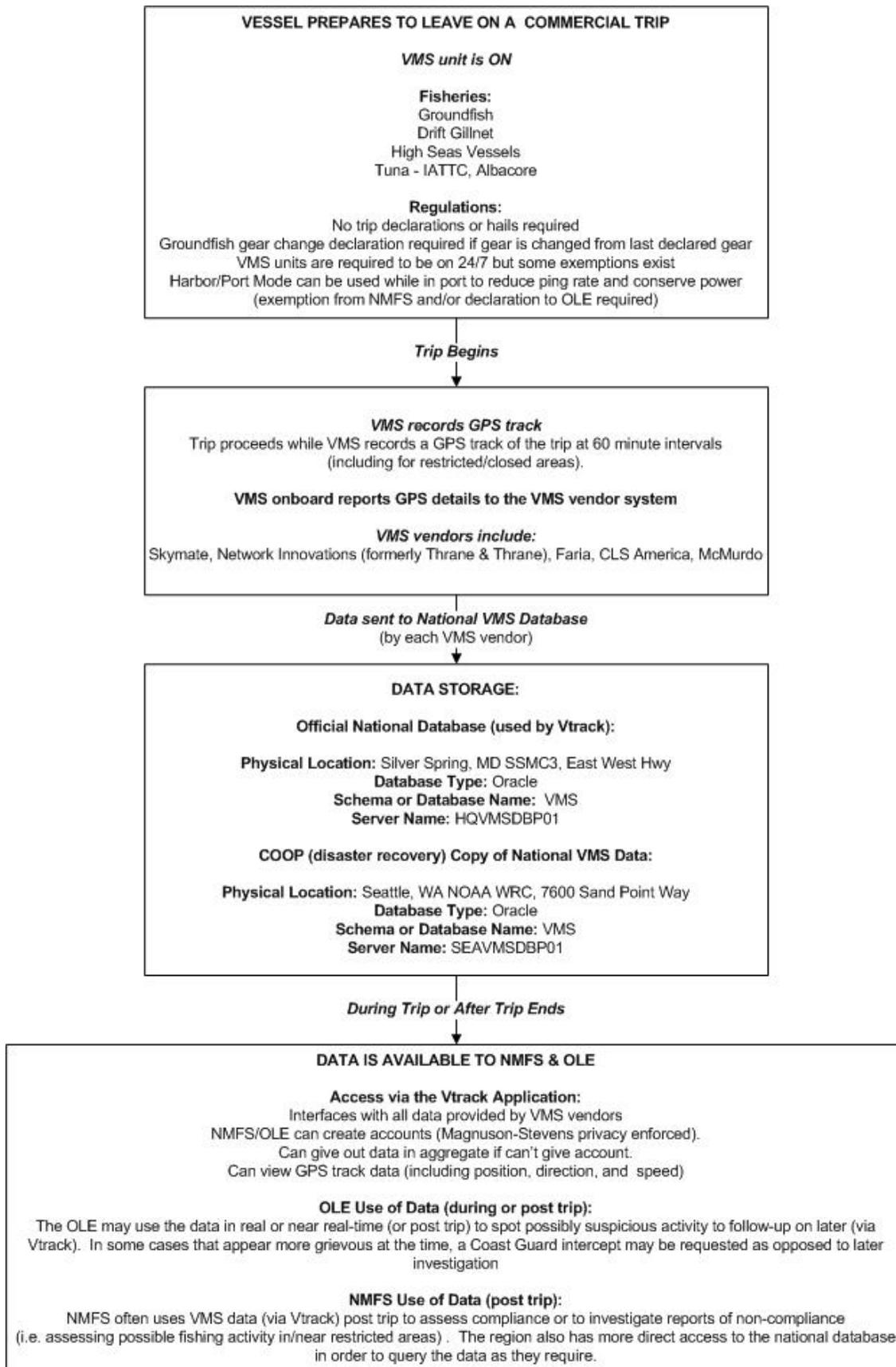


Figure 12: West Coast VMS Data Flow

5.2 Implementation Details and Steps Specific to the Region

As with the Southeast region, the West Coast region may want to consider introducing a trip registry to their current IFQ efforts. The IFQ reporting requirements are already greater and timelier than other fisheries in the region, making the introduction of a trip registry less impactful to current operations there than in the non-IFQ fisheries.

If trying the trip registry concept in IFQ fisheries is not desirable, then it is suggested that transitioning to full electronic reporting in all or some fisheries first would be beneficial.

6 Alaska Region

Alaska has a unique situation in U.S. fisheries and that is there is only one state in the region. This situation appears to have resulted in good co-ordination of data collection efforts in the region as well as solid data sharing. In fact, state and federal landings are put through the same system. There is co-management of salmon and crab between the states and NMFS.

Alaska has 100% electronic dealer reporting, a mix of paper and electronic logbooks, either 100% or partial observer coverage in most fisheries, and has the only known copy of VMS GPS track data (pulled from the disaster recovery copy of that data housed in Seattle).

A unique aspect in Alaskan fisheries is the heavy use of catcher processor vessels and mother ships. This means larger vessels with expanded technical capabilities – making electronic reporting initiatives an easier effort than in regions with smaller vessels with less technological capabilities in the mix.

Catch accounting in the region uses mostly landings (dealer reports) and observer data. This is different than most other regions which tend to rely on logbooks as opposed to observer reports for harvest verification.

The systems in use are very centralized and well thought out. There is a great deal of automated processes to aid in QC efforts and issues are pushed to the forefront to be dealt with. This provides an excellent infrastructure to facilitate the integration of the trip registry concept.

6.1 Current Data Flows

Below are the current data flows for the four data sets of interest. This information has been provided as a jump off point for making changes and to show the current details of the systems in place in Alaska.

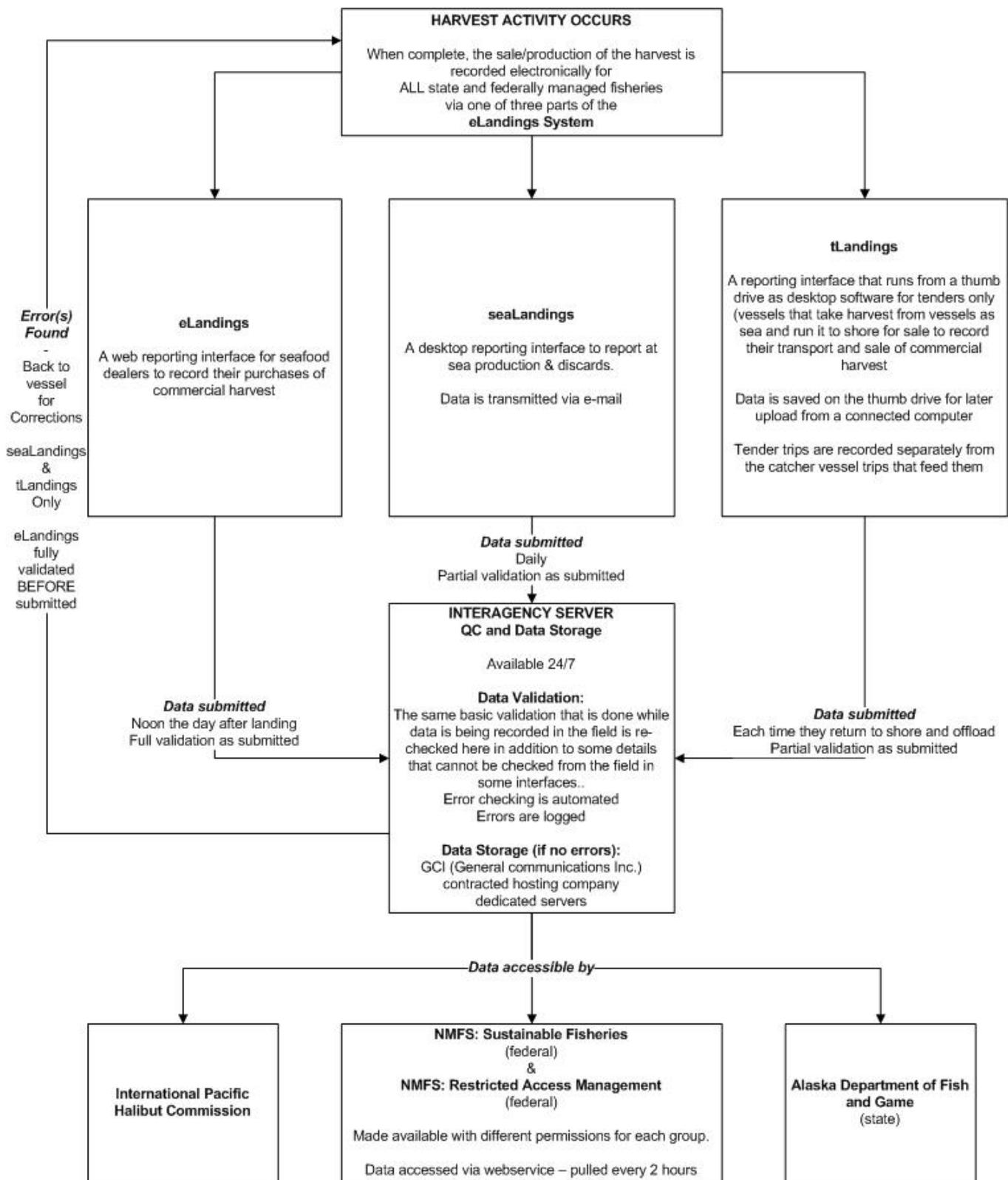


Figure 13: Alaska Dealer Data Flow

The following fisheries have no observer coverage:
 IFQ Halibut & Sablefish < 40' LOA CV
 Longline & Pot <40' LOA CV
 Jig

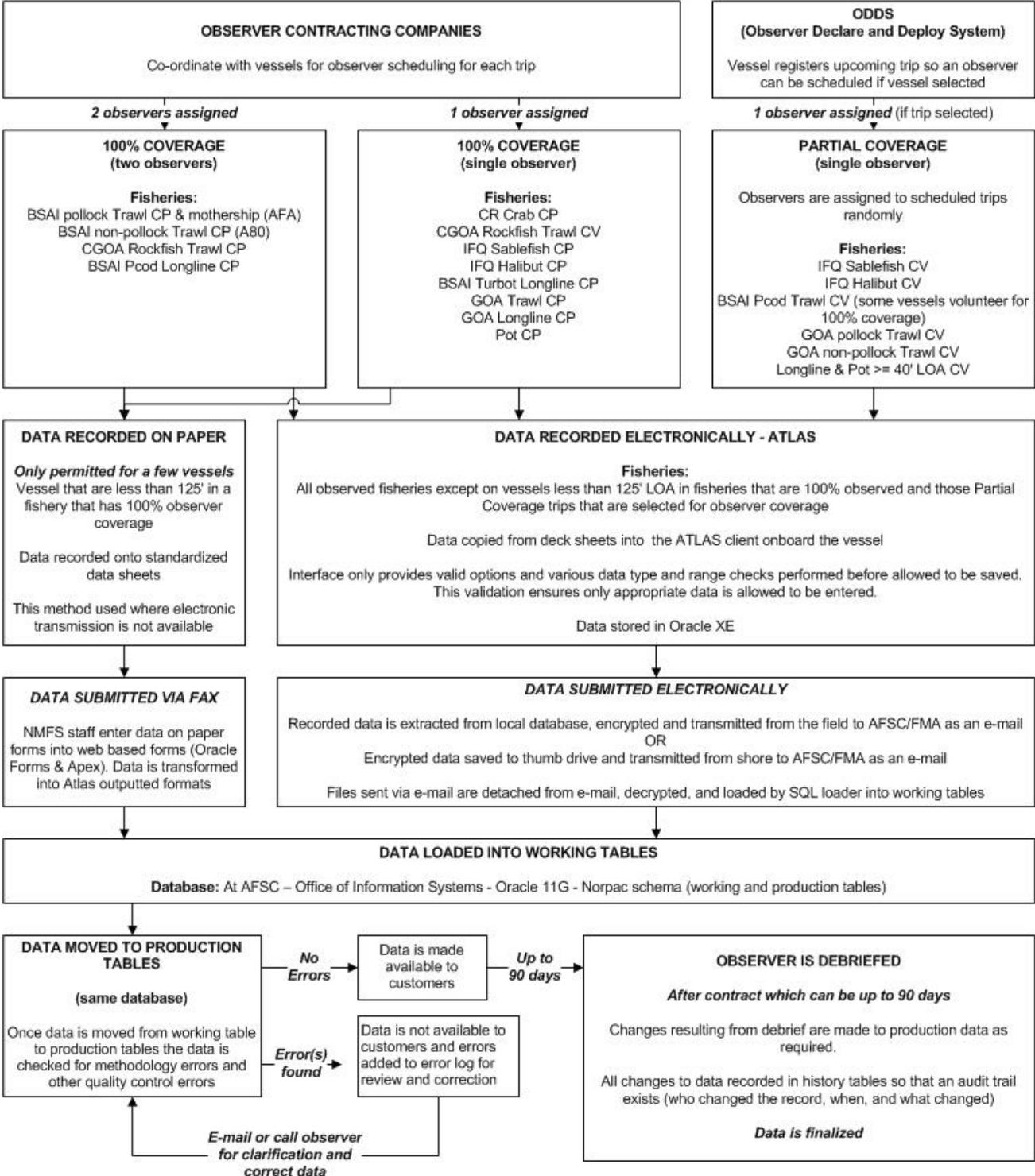


Figure 14: Alaska Observer Data Flow

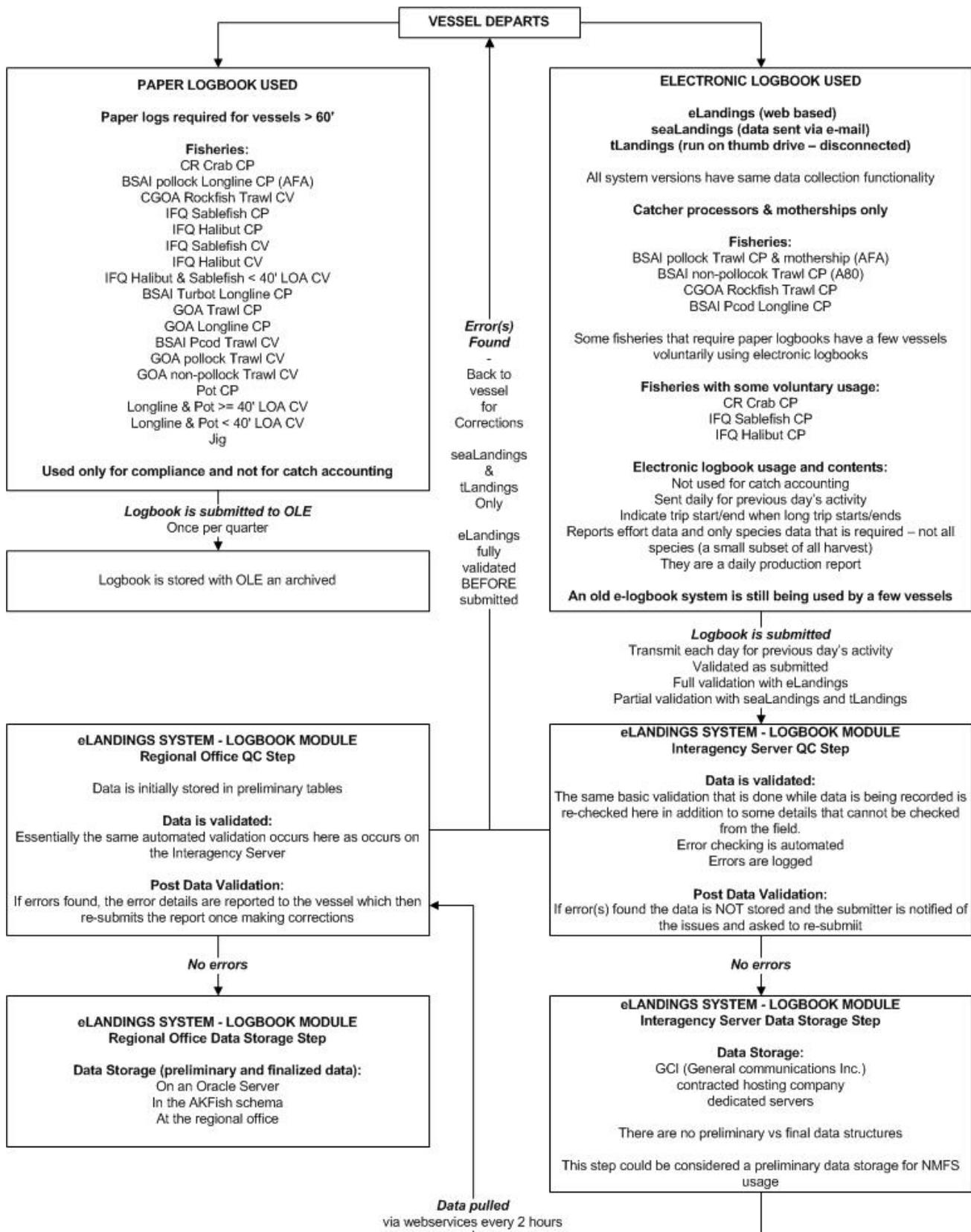


Figure 15: Alaska Logbook Data Flow

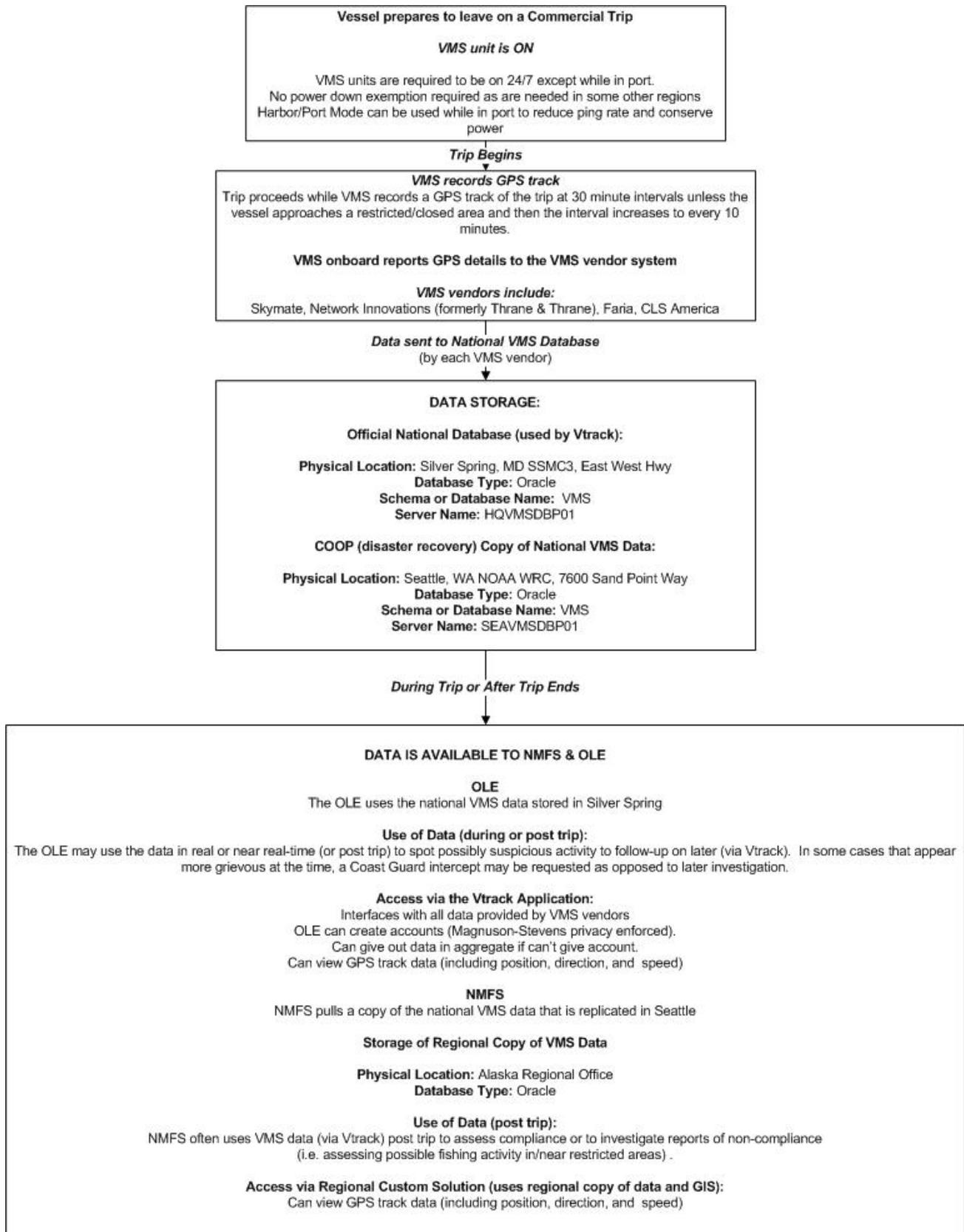


Figure 16: Alaska VMS Data Flow

6.2 Implementation Details and Steps Specific to the Region

7 Pacific Islands Region

This region is unique in that there is really only one fishery being managed federally – HMS longline (deep & shallow sets). All data is currently collected on paper (with the exception of VMS), but there are both an electronic logbook and observer reporting pilots currently underway. Although some dealer data is submitted electronically, that is just on spreadsheets and follows no standard format – essentially these are treated as printed documents and are still manually data entered by people as opposed to loaded automatically into databases in a fully electronic system. There is talk about implementing some form of electronic dealer reporting system, but no formal plans nor pilot as of yet.

Currently, a single individual plays a key role in linking logbooks to observer reports as well as the collection and initial processing of logbook data. This is mentioned as a warning and a benefit. It is a warning because a lot of knowledge sits with this one person and to lose it could be damaging to operations. It is a benefit in that this person knows the current data systems across at least a few of the data sets and will be an excellent source of information when architecting new electronic systems in the region.

7.1 Current Data Flows

Below are the current data flows for the four data sets of interest. This information has been provided as a jump off point for making changes and to show the amount of QC and data entry effort that exists and could be replaced with e-reporting and a well designed trip registry.

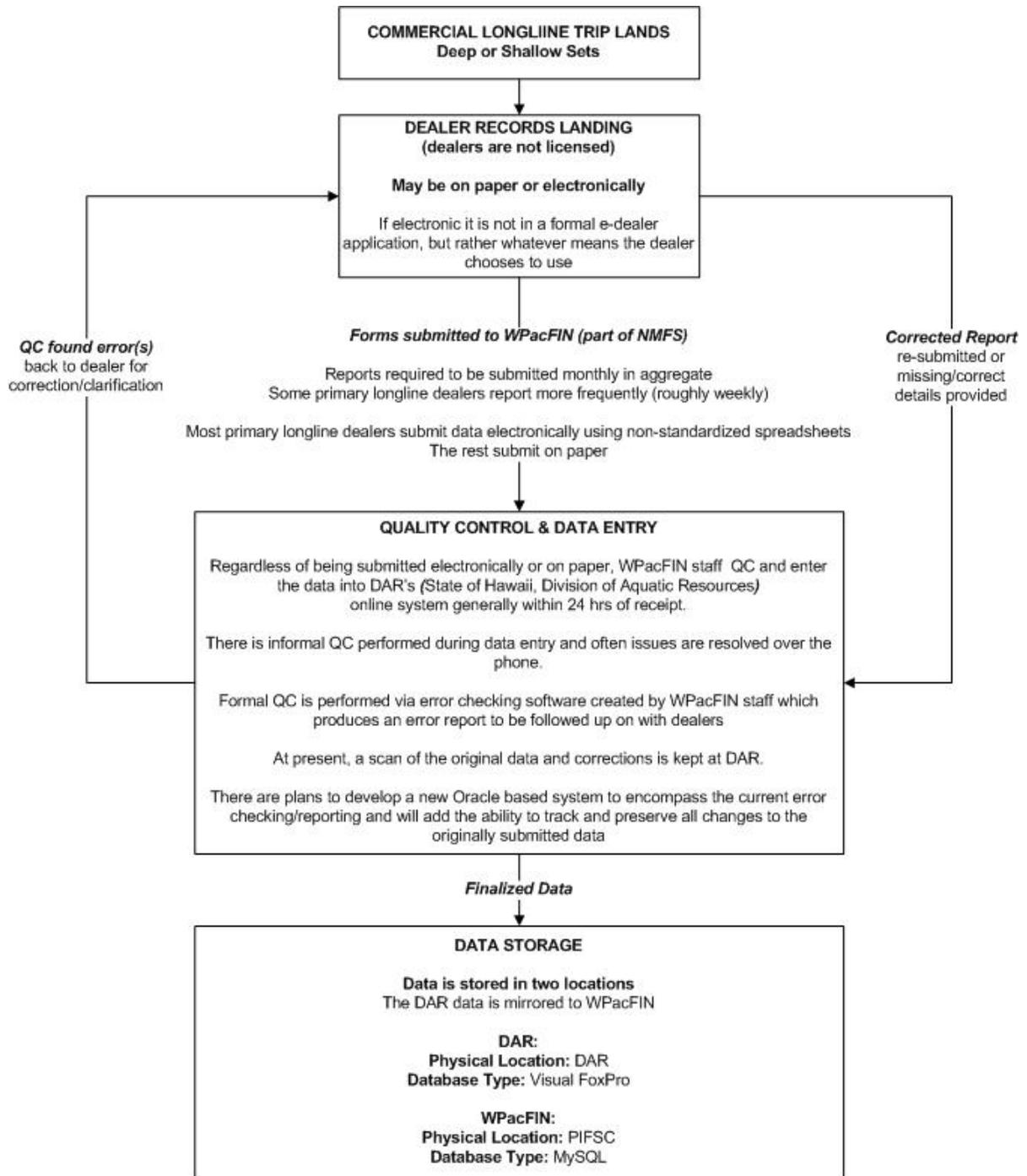


Figure 17: Pacific Islands Dealer Data Flow

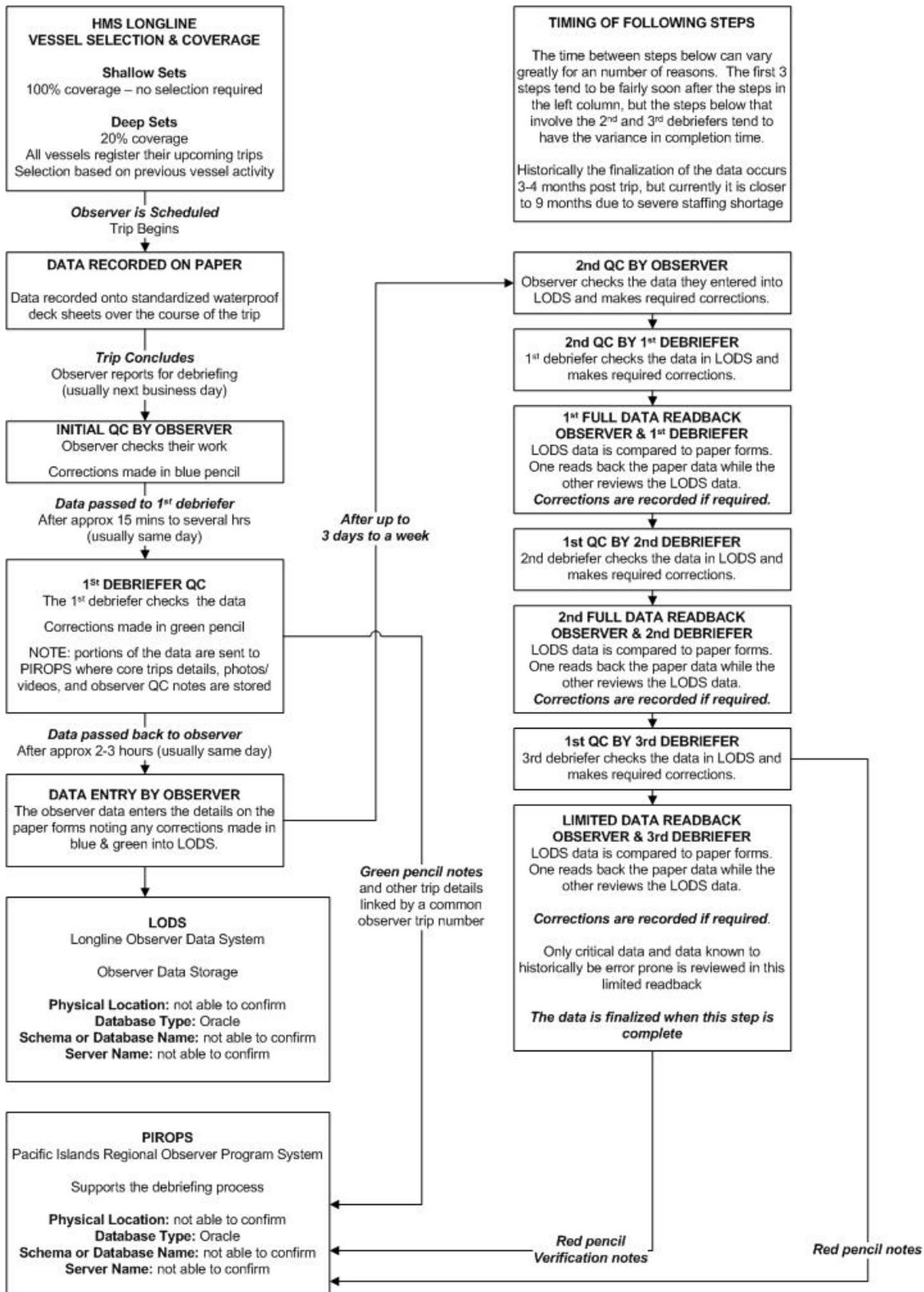


Figure 18: Pacific Islands Observer Data Flow

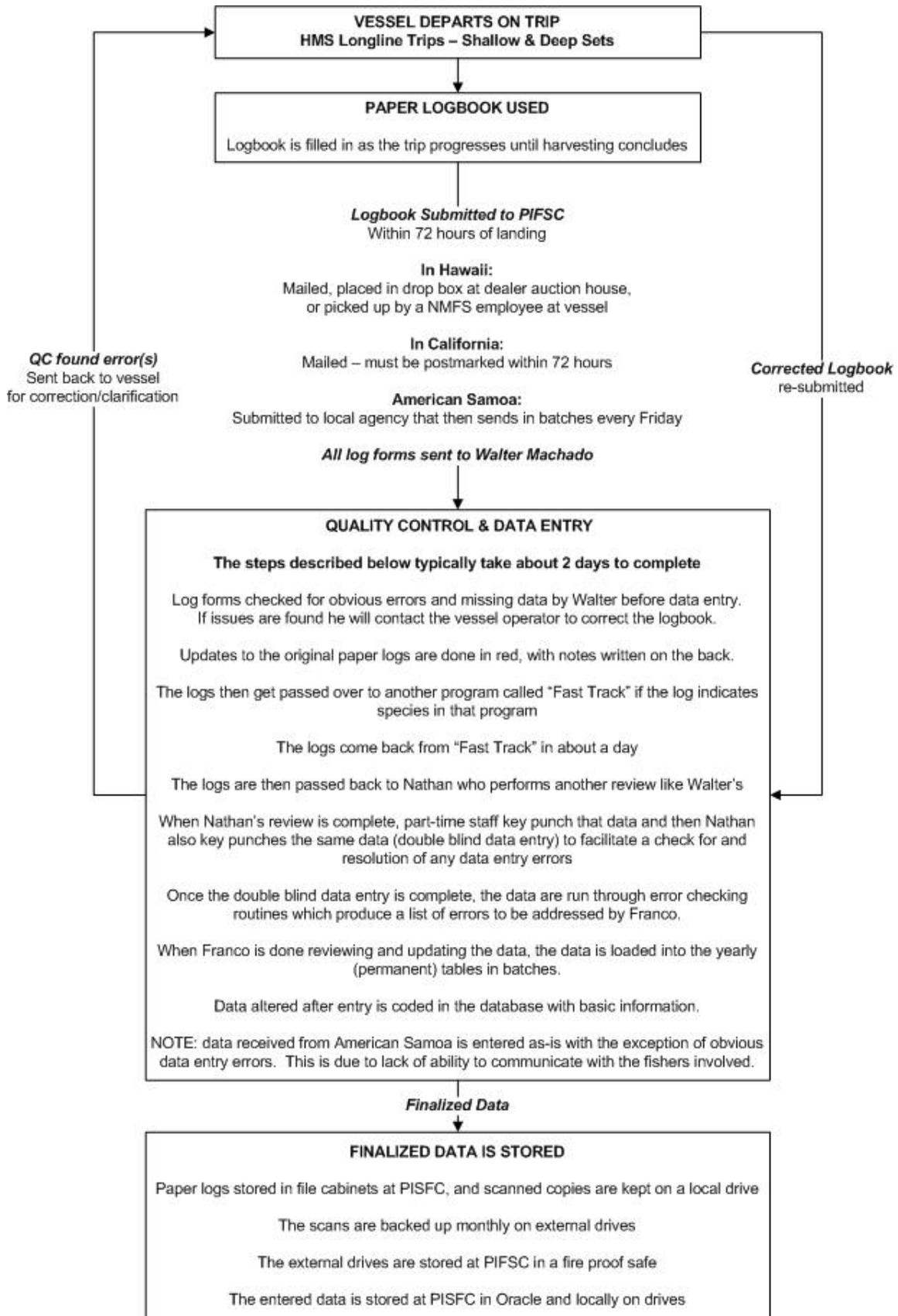


Figure 19: Pacific Islands Logbook Data Flow

Regulations for Hawaii Longline (deep and shallow sets):

NOTE: Focus of this project was domestic fisheries and other fisheries have some differing regulations due to treaties etc.

No hails or declarations made as trip begins, but register with observer contractors 7+ days in advance of trip (list sent to OLE)
 VMS unit must be on prior to leaving on a commercial trip (verified by OLE)
 VMS units are required to be on 24/7 "while at sea" but some exemptions exist (dry dock, in port, etc.)

Vessel prepares to leave on a Commercial Trip

VMS unit is ON

Trip Begins

VMS records GPS track

Trip proceeds while VMS records a GPS track of the trip at 60 minute intervals. In rare situations (like medical emergency onboard) the ping rate has been increased. In this region, the government owns the VMS units and pays a flat monthly rate for service – thus there is no cost savings to decrease the ping rate in port.

VMS onboard reports GPS details to the VMS vendor system

VMS vendors include:
 Skymate, Network Innovations (formerly Thrane & Thrane), Faria, CLS America

*Data sent to National VMS Database
 (by each VMS vendor)*

DATA STORAGE:

Official National Database (used by Vtrack):

Physical Location: Silver Spring, MD SSMC3, East West Hwy
Database Type: Oracle
Schema or Database Name: VMS
Server Name: HQVMSDBP01

COOP (disaster recovery) Copy of National VMS Data:

Physical Location: Seattle, WA NOAA WRC, 7600 Sand Point Way
Database Type: Oracle
Schema or Database Name: VMS
Server Name: SEAVMSDBP01

During Trip or After Trip Ends

DATA IS AVAILABLE TO NMFS & OLE

Access via the Vtrack Application:
 Interfaces with all data provided by VMS vendors
 NMFS/OLE can create accounts (Magnuson-Stevens privacy enforced).
 Can give out data in aggregate if can't give account.
 Can view GPS track data (including position, direction, and speed)

OLE Use of Data (during or post trip):
 The OLE may use the data in real or near real-time (or post trip) to spot possibly suspicious activity to follow-up on later (via Vtrack). In some cases that appear more grievous at the time, a Coast Guard intercept may be requested as opposed to later investigation

NMFS Use of Data (post trip):
 NMFS often uses VMS data (via Vtrack) post trip to assess compliance or to investigate reports of non-compliance (i.e. assessing possible fishing activity in/near restricted areas). The region also has more direct access to the national database in order to query the data as they require.

Figure 20: Pacific Islands VMS Data Flow

7.2 Implementation Details and Steps Specific to the Region

Given that the region is already involved in electronic logbook and observer reporting pilots and there is talk of a possible electronic dealer system being developed, it is an excellent time to consider the addition of a trip registry.

Regardless of whether the region feels the need to consider a trip registry now or later, assessing their current pilots while looking at similar needs for an electronic dealer system in terms of integration with a trip registry would be wise. This may spot some changes to how the current pilots perform certain tasks that could make them more ready for future integration with a trip registry. In addition, lessons may be learned that would aid in the formal development of an electronic dealer reporting system.

8 Greater Atlantic Region

As mentioned earlier, a full analysis was not possible for this region.

It is known that dealer reporting is a mix of electronic and paper reporting (appears to be heavier on the electronic side). Logbooks are still mostly on paper, but electronic logbooks can be voluntarily used in all (or at least almost all) fisheries. It is known that observer data collection is performed on paper, but efforts are underway to incrementally transition to electronic reporting. VMS GPS track data is collected as it is in all regions, but it is not clear which fisheries require it and which do not.

It is known that the region is planning something that appears very similar to the proposed trip registry (at least in terms of the high level flow of a commercial trip). The author of this report saw a presentation by Doug Christel from the region at the 2015 AFS conference in Portland, OR which touched on this effort. What is not known is any details of how they plan to implement this for any of the data sets or if it in anyway handles unique identifiers in the ways proposed in this report. It is entirely possible it is simply another form of a hail system. The name for this new system (at least at the time of the AFS presentation) is the Vessel Activity Census (VAC).

When the region was initially visited (well before the VAC was being thought of), discussions revolved around the need in the region to know more about commercial trips in advance of the trip to facilitate higher levels of accuracy. At that time it was in fact hails that were being discussed.

Signs seem to point to a decent approach in the region and it is hoped that the concepts put forward in this report are of value to the efforts being undertaken.

9 Error Reduction by Design

There are many ways to reduce errors when collecting data electronically – this is the main and major difference between electronic reporting and reporting on paper when electronic reporting is done correctly. Unfortunately, time and time again the focus when introducing electronic reporting in any environment (not just fisheries) is more on what needs to be collected as opposed on ensuring it is collected with the highest levels of accuracy possible.

It is always better to spend time ensuring high levels of data accuracy as opposed to complex data quality checking routines after data is submitted – the latter is often the way NMFS currently operates as seen in the current data flow diagrams for each region.

Error reduction falls under the following categories:

1. Avoiding and limiting human input whenever possible
2. Quality assurance of the input beyond form controls

9.1 Avoiding and Limiting Human Input Whenever Possible

This kind of error prevention falls into two categories – avoiding and limiting.

There has been lots of discussion in this report already about avoiding human input. This is how the proposed trip registry design came about; the need to remove human handling of unique identifiers in order to drastically reduce errors that are currently occurring in fuzzy matching efforts around the country. No further description of the avoidance of human input in this regard is needed here.

The limiting of human input can be accomplished in many ways, but those methods all have one goal – to avoid the free form entry of data in all cases possible. This is in fact not always possible, but when it is not there are methods to mitigate the potential for error.

9.1.1 Form Controls other than Text Fields

These form controls allow the user to provide data, but not by typing. In all cases only the appropriate data can be put forward for the user to choose from as opposed to letting them type any response they want. By using these form controls, data accuracy is raised considerably simply because users cannot provide invalid responses. They can still provide a combination of responses that is invalid even though the individual responses are valid – this is where validation of an entire submission is needed.

It is also important to accept that they can provide a valid answer that is incorrect (i.e. they meant to select the 1st option in a select box, but accidentally selected the 2nd). These can sometimes be caught when cross checking a user's submission against other data that may also answer the same question (i.e. comparing a logbook to an observer's version of the harvesting observed for the same trip).

9.1.1.1 Select Boxes (single and multi-select)

These allow for a valid list of options to be provided for the user to select from. Single select boxes allow for a single entry while multi-selects allow for multiple options to be selected at once. Multi-selects should be used carefully as often end users are unfamiliar with how to hold down the CTRL key and use the left mouse button to select the desired options (clicking an already selected option deselects it). Further to that, multi-selects often allow for any number of selections to be made. If a limited number of selections are required, then checkboxes may be a better solution or adding a later validation check that ensures no more than the maximum number of selections has been made.

Filtering of the contents of select boxes is often performed based on other user input or other known details (perhaps about the user via their user account and associated details). For example, most mailing address forms that serve more than one country will perform this kind of filtering – often called “related” select boxes. In this example, the user selects a country in the address form and that selection then alters the list of states to provinces (if the user selected Canada as their country). This is an excellent way to pre-filter data before a user interacts with it so they can only select valid entries and keeps the number of options presented to them as low as possible. Keeping user interfaces as clean and easy to read as possible helps the user to be more focused and accurate.

9.1.1.2 Checkboxes

These are simple controls that are an excellent way to allow a user to select multiple options without needing to scroll through a select box list that is maybe too long. A great example is the options that can come on a car. This control is often seen in interfaces designed to help someone customize a car purchase in order to pre-determine the price. The user is presented with sometimes dozens of car options with a checkbox beside each – all options are specific to the car being assessed. The user then checks all options they want and a quota can be generated.

Checkboxes are an excellent way to ask for answers which are Boolean (true/false, yes/no) in the form of a single checkbox where if checked it means either true or false and being unchecked is the reverse.

9.1.1.3 Radio Buttons

These can be used for Boolean responses as checkboxes can be as well as for short lists of options where the user must only select one option and it is better for them to see all the options at once (unlike a select box).

9.1.1.4 Date pickers

There is nothing more frustrating to deal with for programmers and data managers than dates that were not properly recorded and thus the date format is not known (i.e. 2016-02-12 vs. 2016-12-02 – both may be in Feb or Dec unless you know the format). Date pickers solve this issue as the user can select a date using a calendar format they are used to and that date is of a known format. The selected date is usually placed into a disabled text box (disabled from direct editing by the user) in the format specified by the programmers. This alleviates most of the frustration and inaccuracies that can so easily slip in when free form date entries are allowed as well as the need to validate every date entered by a user for the correct date format.

9.1.1.5 Sliders & Range Selectors

These two controls are often used to interact with ranges of values (i.e. 1-10 or strongly disagree to strongly agree).

Sliders are often used to select/define a single value that is part of a range of values part of a range of values and range selectors allow for a subset of the range of values to be selected (i.e. a min/max value in a range of numbers from 1-10).

These things can be done with other form controls as well, but these two controls provide an interface that users are used to seeing when providing this kind of data (i.e. most surveys ask for the strongly disagree to strongly agree range of answers). They are often

used in mobile interfaces as it is easier to move a slider with your finger than to toggle a small radio button.

9.1.1.6 Auto Complete

While auto complete is not technically a form control, it is tied to text inputs. As the user types, a list of valid values can be queried and the most likely matches displayed (either the most likely option is appended to what is being typed or several likely options show in a list under the text box for the user to select).

Auto complete makes sense for search interfaces as it helps the user to find what they are after fairly quickly (ease of use is known to raise accuracy). They may not be of much use for data quality assurance because of how easy it is for a user who is unfamiliar with the interface and how it behaves to make an errant keystroke and accidentally select the wrong suggested item without noticing. That selection is a valid one, but not what was meant to be reported – similar to a typo that is still a valid value.

9.2 Quality Assurance Other than via Form Controls

Restricting user input via proper form controls should be considered only the first line of defense in terms of data quality assurance. It goes a long way to ensuring the user can only provide valid entries. More is still required for various reasons, but the most significant is that so many provided details need to be cross-checked with one another or other rules apply that may not even be about what was entered in the form (i.e. is it even appropriate for that form to be filled in given other events).

To ensure the highest level of accuracy requires some bigger thinking:

1. Defining the business rules that impact the collection of the data in question
2. Defining as many of the business rules in the database as possible
3. Using of the above details and data to:
 - a. Provide only valid options for the user to choose from wherever possible. This is an excellent way to avoid invalid entries (but can't stop them all).
 - b. Verify and validate the data that the user submits.
 - c. Feed processes meant to quality check previously submitted data
 - d. Produce reports showing outliers or responses that are out of range but not impossible.

9.2.1 Defining Business Rules

Business rules come in many forms. They range from very basic to quite complex, can be very fine grained (like how a piece of data relates to another or if the data is required or not), or not even be about the data being collected and instead when it can be collected or edited and so on. The rules help to validate data as well as control the flow of data and many other things in a system.

Business rules most certainly do not stop at the defining of what data is to be collected. The author's own experience with NMFS requirements is that they often do stop at what data needs to be collected and the format of that data without specifying other details that

aid in the control of data flow or avoidance of reporting when it is not appropriate. Instead the approach tends to lean towards accepting all submissions and sorting out the details after the fact.

Some examples of business rules are as follows:

- A start date/time should be before an end date/time
- A piece of data asked for should be numeric
- Changes to an entire report can only be made up until a certain time period (say midnight the day of the event or up until 72 hours after the event)
- Cod must be recorded in either pounds or kilograms, but Halibut can only be recorded in pounds.
- A valid entry for weight of a given species harvested by a specific gear is between 5 and 10,000 pounds per trip
- A permit that is expired should not be allowed to be involved in a new trip until such time as the permit is renewed
- A new trip cannot be started in a fishery that does not have an opening as of the date/time a trip is desired to be scheduled for
- A new trip cannot be started until all harvest is known to be offloaded
- A definition of the capabilities of a permit so a system can know if that permit has a special condition that allows for some harvesting activity that requires additional data to be collected about the trip and thus a dynamically adjusting form can show the inputs for that additional data.

As you can see from the examples, the rules can be directly about the data entry or about the entire entry effort itself (controlling the flow).

First we'll look back at the free form text issue and how business rules can be used to address that scenario.

9.2.1.1 Data Driven Business Rules for Free Form Text Entry Scenarios

Let's reflect on the first section above about using form controls to reduce human free form entry of responses. You can't always get away with entirely removing the need for a human to type something.

In fact this happens often in fisheries in the form of lists of harvested species and the corresponding quantities that were harvested on a trip (or purchased by a dealer or observed by an observer). To be clear, this example is about addressing the free form entry of quantity values and not the selection of valid species from a select box. The solution is to define the ranges of values that are deemed acceptable for each species and unit of measure in order to ensure wildly out of range quantity values cannot slip into the submitted data without being spotted.

We'll use these examples from above to show how this should be handled "A valid entry for weight of a given species harvested by a specific gear is between 5 and 10,000 pounds per trip" and "Cod must be recorded in either pounds or kilograms, but Halibut can only be recorded in pounds."

These rules need to be defined in data so that the free form responses provided by the user can be checked to ensure the rules are being followed.

The following data assumptions can be made:

- A table exists that lists unique species and each species has a species ID
- A table exists that holds the valid units of measure and each has a unique unit ID
- A table exists that lists the unique list of gears and each gear has a gear ID

A “Species Harvest Rule” table can be created that holds records that relate each species to each gear, the allowable unit, and the acceptable normal quantity range for the combination of species/gear/units. The “Note” column is for clarity and should not be in this table. We can add to the examples and say that the range in pounds for Halibut is 40- 120,000 pounds and the other range mentioned was for Cod.

Species ID	Gear ID	Unit D	Min Value	Max Value	Note
2	8	3	5	100000	Cod in pounds
2	8	7	2.272	45454.545	Cod in kilograms
3	8	3	40	12000	Halibut in pounds

The table above can now be used to verify that the units and quantities reported of Cod and Halibut are appropriate given the data driven business rules in play. If a user enters 200,000 pounds of Halibut, an error can be displayed to inform them that Halibut should be within 40-200,000 pounds.

Now consider adding effective dating to the table above and you have a table that defines the rules around the reporting of harvest over time. This makes it VERY simple to know which rules were in effect when a given report was submitted – this is critical in unwinding data situations that arise over the course of maintaining and supporting complex systems.

You may be asking yourself “What about if they report in the wrong units or what if they really caught that much Halibut?” The next section will address the other ways in which the above table can be used.

9.2.1.2 Other uses of Data Driven Business Rules

It should be pointed out the sample table in the previous section is one of MANY ways to define data driven business rules. This report is not intended to teach how to design systems, but rather to show some concepts that are of great use that NMFS can consider when undertaking new systems projects. We won’t delve into every possible way that rules can be defined in data, but wanted to make sure it is clear that these concepts can and should be leveraged to enhance NMFS data systems (or those built by contracted third parties). That said, the author (Bryan Stevenson – the creator of a true integrated reporting system for fisheries) is available to discuss other ways rules can be defined in data in the event that the reader would find this useful.

The author of this report has software in production and pilot use in a few NMFS regions and can attest to the need for clearer communication of how NMFS data is related and what rules should be applied to its data collection.

Consider the simple relationship of acceptable units to be used for harvested quantities. That very detail has been seen as free form text in a spreadsheet as opposed to proper structured data that is crystal clear. Rules defined in data do not require the interpretation

of free form typed rules or data relationships on the part of third party IT contractors or whomever they are intended for. Being able to communicate these details in a structured way in data is far more preferable than leaving so many critical details open to interpretation – it also make conveying these details far simpler and less time consuming to document.

Let's look at other ways the sample table in the last section can be leveraged.

Presenting only valid options. It has been mentioned that the use of various form controls can allow for the presentation of only the valid options. The way to present those options is from data driven business rules. Consider the sample table in that last section. It can also be used to display a table of harvest for a vessel operator to fill in. If the user is asked for the gear type that was used before the table harvest table is displayed, the data above can be used to populate the appropriate unit select box options for each row in that table. As the species that was harvested is selected for each row from a select box, the correct list of valid units for that species and gear combination are retrieved and used to populate the initially empty units select box for that row. The presenting of valid options is not just at the form level, and can be used throughout the application – perhaps to present only the currently active fisheries to start a new trip in given the authorizations of the current user's permit and the known openings.

Automated validation of an entire form submission. So the electronic form has been filled in and it was presented using all the best concepts for limiting invalid entries – so what? Well there are still plenty of things to validate (i.e. is a start date before an end date, were all the required fields filled in, was a special condition recognized in the submission and handled the correct way, the list goes on.

By using the sample table from that last section, the units of measure specified can be double checked in a similar way to the correct units being presented in the first place and any errors can be displayed to the user for their correction (more of presenting and handling errors later).

9.2.1.3 Automated Data Quality Checks and Outlier Reports

Once again we look to the sample table in the last section. Just as it can be used to present only valid data and allow for validating an entire form submission, but it can also be used to drive automated QC efforts and produce reports on outliers seen in the data based on the defined rules.

Another way to look at this process is to consider that not all data may have been entered via the form under your control. There may be a need to import external data that is of unknown quality, and using data driven rules can allow that incoming data to be quality checked automatically. The results could be placed in a readable report or data issues could be reported back to the sender of the data (even in a format that their system could read and perhaps auto-fix some or all of the data issues discovered).

9.2.2 Avoid Overloading the Meaning of Data

There are often cases where shoe horning a piece of data where it does not belong seems like a quick fix and “OK in this case”. This is an incorrect assumption and will inevitably

come back to haunt you in various ways (including conveying business rules that are defined in data).

Here are a few examples:

- “Large gutted fish” or similar are often found in lists of data that are meant to contain the condition of a fish when sold. It in fact describes both the grade (large) and the form (gutted). Grades and forms should be kept in separate lists as they describe different things. If you keep them separate then they can be applied discreetly and reported on discreetly (all fish that were large when sold and all fish that were gutted when sold). If they are combined into a single entry in a list, things become more difficult to program for and you end up doing things the hard way that should be simple.
- “Home Consumption” in a list of dealers. It is not a dealer, it is a catch disposition. It may seem like home consumption could be considered a dealer – after all it is where the fish end up. That would be the incorrect way of looking at the data. A dealer is something you identify AFTER the catch disposition of “sold” is determined to be what happened to the fish. By adding that entry to a list of dealers you have overloaded the meaning of a dealer to include home consumption. This makes reporting and programming harder in the same ways put forward in the example above.

9.2.3 Errors – How to Present and Manage Them

We’ve now looked at how to reduce invalid entries as much as possible, but we still have to consider how to present errors to the user and think about what should happen if an error can’t be resolved at the time of data entry.

The golden rule of presenting errors is to provide clear details about the error that non-IT professionals can understand. Part of the explanation should clearly discuss how to resolve the error.

Here are some examples of errors and a note about them being acceptable or not:

- **An error has occurred**
 - Unacceptable – does not explain the nature of the error
- **Please fill in all required fields**
 - Unacceptable – does not identify the required fields
- **Missing codepage or helper program, or other error (could be IDE device)...**
 - Unacceptable – this is an unhandled exception and is meaningless to end users and even many IT professionals – instead handle these exception with a generic error that indicates it is not known what happened and perhaps to contact a helpline because there may be no way around it
- **The harvested value of Cod should be a whole number greater than zero (currently -1234). Please correct and re-submit the form** – acceptable – details what is incorrect and what to do about it.

Now there is still the issue of what to do when an error can’t be corrected. This can occur when the provided entry options do not allow for what the user really did or the error is the result of free form entered text being out of range (as per our earlier example).

Those two situations can be handled differently, but share a common concept.

When the user enters a combination of X and Y that form validation recognizes as invalid. This is not the user's fault and they just want to submit what really happened. An override mechanism can be built that can let the user know that they can use it if they encounter errors they cannot get past. The entire submission as well as the parts of it known to be in error can be accepted, but marked as quarantined (even going as far as to store the data in alternate tables). This lets the user move on (perhaps being told that using the override may have consequences to downstream system functionality and/or that they may be contacted to resolve the error(s)). It also provides a clean way to separate good data from questionable data that needs to be reviewed and resolved).

There is also the 2nd kind of error that stems from the validation of free form text. Using the earlier example, a user could provide 105,000 pounds of Cod – 5000 pounds over the max value defined in the table. In these situations it is wise to make the errors into dismissible warnings. The reason being that the user may have actually caught that much Cod regardless of the assumed min or max values usually encountered. The warning is still shown in case the user mistyped the value and needs to make a correction. It can be dismissed because maybe it did happen and there is not actual rule about harvesting over 100,000 pounds of Cod on a trip.

The decision then is what to do with the dismissal. It can be recorded that an out of range quantity entry came in for Cod. It may also be easier to just run an automated QC routine as described above and using the data driven rules to find the out of range entries and report them. Depending on how proactively this data is being managed, notifications about these events can be sent to those responsible with dealing with them in real-time.

10 Device and Communications Considerations

Currently most data that is collected electronically is occurring on laptops, PCs, via VMS systems, and more recently via mobile devices such as tablets and smartphones.

10.1 Mobile Technology for Unique Identifier Handling

Mobile technology allows for simple and quick handling of data transfers that can't be done easily (at least by the average user) with other devices that have historically been used for data collection. This is what makes them a good fit for the handling of unique identifiers as part of a logical equivalent model. They contain sensors that other devices don't have which allow for mobile applications to be built that leverage these sensors to interact with other technology for data acquisition. Mobile applications can access and use the sensors without the user of the device knowing how to use them. In addition, connecting mobile devices to Wifi connections is simpler than on more traditional computers. All this means it is an easier computing platform for the non-technical user and should increase uptake while keeping support requirements lower.

By using RFID chips affixed to the hulls of vessel, the vessel VRN can be read by a mobile device in close proximity to the chip. This would facilitate the transfer of the vessels VRN to a dealer as described in the proposed solution section above. There should be no privacy concerns as all the chip will identify is the hull number already on the hull.

There are many free RFID chip reading mobile applications available and this means that producing such a mobile application as described above should not be overly burdensome and it should work with devices that many will already have.

To facilitate the varying unique identifier interactions between the vessel operator and the dealer, Bluetooth and/or NFC (near field communications) can be used to great effect.

To handle the interactions with the trip registry by the vessel operator (to perhaps correct the vessel) or by the dealer (to compare the acquired VRN to the VRN on file for the scheduled trip), Wifi or a cellular data connection will be great solutions. If Wifi is available for free at each port it would allow for a standard trainable way to connect to be provided to all parties involved. Having free Wifi would also reduce the cost to participants that could be incurred using their cellular data plans (perhaps a good PR aspect of Wifi availability at ports).

10.2 Mobile Technology for Data Collection

As mobile technology has become more prevalent, many fishers have started to use it in their daily lives. Often those that did not take well to more traditional computers have found mobile technology very user friendly and they want to use it in their fishing activities as well.

As with any tool, selecting the correct one for the job is important. With the smaller screen size of mobile devices (especially smartphones), care must be taken to ensure that reasonably usable interfaces can be built and used successfully to produce high quality data.

Mobile interfaces are designed to be vertically stacked as opposed to horizontally laid out. This makes it difficult to provide tabular data layouts (rows and columns of data) for data entry and review. When complex data is asked for in a vertically stacked format it can become very confusing and cumbersome for the user because they cannot see their entire entry on one screen and make sure they have done things correctly. This speaks directly to the quality of the data being provided. So special attention should be given to what kind of device software will be deployed on. Sending a trip end notification or interacting with unique identifiers is an excellent use of the technology, but entering complex multi-species logbook data on a mobile device with a screen smaller than 7 inches is not wise and should not be promoted as a high quality solution. This is a requirement that NMFS may want to consider as part of data quality improvements.

10.3 Data Transmission from Sea

The requirement to transfer data from sea should be re-evaluated in light of concepts put forward in this report that could be used to replace most current needs to send data from sea. It is costly to industry and adds more burden to their fishing activities.

If it is still found that data needs to be sent from sea, perhaps it can be sent when close enough to shore (state waters) to allow for cellular connectivity instead of satellite.

VMS systems or other satellite based solutions (phones used as modems or dedicated modems) are still the clear winners for data communications from sea where cellular connectivity is not available or too unreliable.

These systems tend to be costly, but alternatives are starting to surface like AIS.

10.4 VMS Alternative for Positional Data

As mentioned in this report, the Gulf Shrimp data logger or similar low cost technology could be used to record positional data on commercial trips. It must be ensured that the device is a) on and b) tamper proof. These loggers get to the heart of what is being asked for in terms of position, direction, and speed without all the extras and cost that come with current full flown VMS systems.

There would be cell charges incurred, but would be less costly than VMS systems currently in use. These data loggers could fit on vessels that really don't have room for the current VMS systems.

11 Conclusion

Electronic reporting is on the rise in all regions and NMFS has clearly made this transition a priority in order to be more accurate, receive and process data faster, and re-think “the way things have always been” which is often the mantra in fisheries.

In discussion with those in the trenches of data collection and matching efforts in the regions it is clear they see the need for a better way to link data sets from commercial trips. There are many efforts underway to tackle these challenges, but many are missing the control that electronic reporting provides and instead rely on humans to follow best practices and not make mistakes.

It is hoped that the proposed solutions and new ways of looking at old obstacles in this report can be leveraged by each region or spark new solutions to their specific situations.

Although fisheries data collection has some difficult and complex challenges, they can be overcome. In fact, in many instances not only can they be solved, but made simpler, faster, cheaper, and reduce the burden of reporting on the participants.

There have been a great deal of details put forward in this report, but it is understood that the written word is only as clear as the reader's interpretation. I (Bryan Stevenson the author of this report and creator of a true integrated reporting system) am available to discuss any part of this document with any and all of the regions in order to ensure the correct message is being received. Please do not hesitate to contact me at bryan@fisheryfacts.com.

I would like to thank all of the people in the regions I met during this project – I couldn't have come up with solutions without first understanding your challenges and current methods of collecting fishery dependent data.

12 Appendix A – Terminology & Definitions

Data Sets of Interest:

Any mention of “data sets”, “data of interest”, “four data sets of interest”, “fishery dependent data”, or similar will be in reference to the four fishery dependent data sets that this project investigated.

The data sets are as follows:

- VMS - both standard GPS track data and hauls, trip declarations, and any other non-standard VMS data sent through a VMS system
- Dealer Sales – aka fish tickets, landings, offloads
- Logbooks – trip logs filled in by the captain or crew that record area, gear, and harvest information
- Observer – typically discards are the interest with this data set, but does include all observer data that may be collected on a trip (although more focus was given to the data that supports catch accounting as opposed to the scientific data like biological samples etc.

VRN: vessel registration number (aka vessel identifier)

FIN: fishery information network

Finalized Data: This is data that has passed through any quality control steps that are required before the data is usable. Data should almost never be considered to be 100% finalized and static because there are many reasons why it may still be changed (i.e. a QC issue may have been missed or a reporter of the data may realize a typo and so on) post quality control. It is advised that all data changes be tracked and logged to accommodate situations where it is needed to be known what the state of the data was at any particular point in time.

13 Appendix B - Regional Data Flow Diagram File Access

Please note that all original Visio files for all the regional current data flows are available upon request to Mark Brady (mark.brady@noaa.gov).