Metadata

Min-Yang Lee

February 03, 2021

Contents

This has a purpose and a target audience. I'm not quite sure what that is yet. One of the audiences is "future me."

General things

Many data tables are live, with nightly or continuous updates. For example, when working with the dealer data, expect approximately 300 changes or additions to the current and previous year of data per day. Data is "complete" 6-9 months after the end of the calendar year; however, small changes are always occurring.

This has consequences for reproducibility if you do not store a copy of the data.

Make sure the table that you're using is not "stale." Alot of the data is copied from GARFO to NEFSC servers. Sometimes, they stop getting copied. Sometimes they are updated monthly. One way to check this is to get the maximum DE, DC, or some other date field.

There are table and there are view. Sometimes, the sql that generates a view can help you figure out why you're getting an unexplainable result of a query. For example, the following bit of code shows that SECTOR_PARTIPANTS_CPH is based, in part on permit.vps_owner, permit.vps_vessel, and mqrs.mort_elig_criteria.

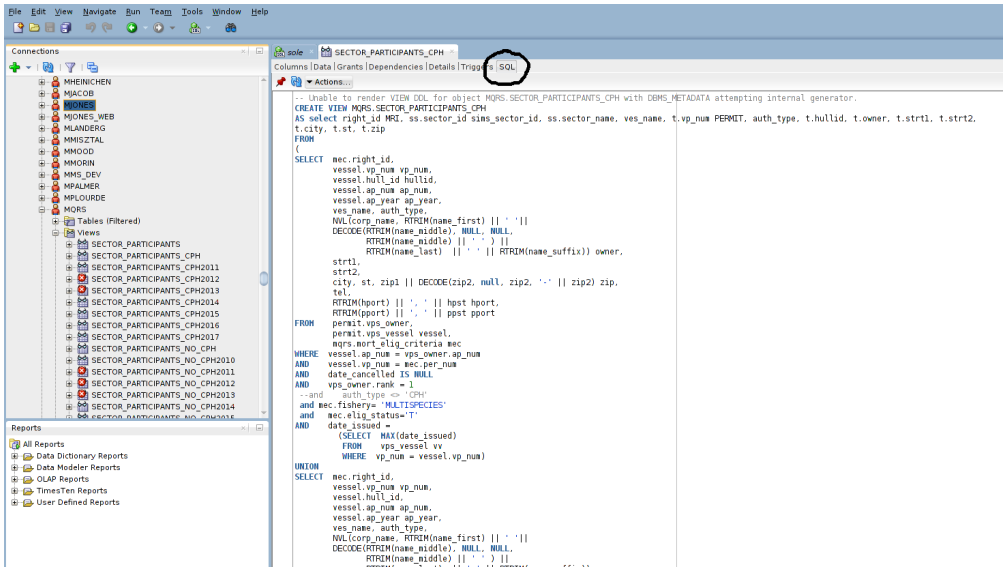

Figure 1: An illustration of looking at the underying SQL code that generates a view.

Dealer Databases

Tables: CFDETSyyyy; CFDERSyyyy

Location: Sole

Schema: CFDBS

Overview

The dealer data are transaction-level pricing at the level of the "market-category." These data are primarily generated through mandatory reporting by federally-permitted fish dealers. The federal reporting is supplemented with data from non-federally-permitted (state-only) fish dealers. Data are currently reported electronically in partnership with ACCSP through SAFIS.

- CFDETSyyyy contains "detailed species data" for 1994-2003
- CFDERSyyyy contains "detailed species data" for 2004-present
- CFLENyyyy -fish-level port sampling data for length

Additionally, APSD has a table of CFDERS for all years, so no need to loop through iterative lists in R or or use UNION in SQL. This can be accessed with

select * from fso_admin.cfders_all_years@garfo_nefsc;

Current Collection Methods

These data are the result of mandatory federal dealer reporting at the "trip-level", supplemented by state-level, aggregated reporting. Federally permitted fish dealers that are required to report purchases of all fish to NMFS.

Changes to Collections Methods

- The number of species triggering these requirements have increased over time, which has implications for completeness (50 CFR 648.6). For example, mandatory dealer reporting for Monkfish, herring, and hagfish began in 1999, 2001, and 2007 respectively.
- Mandatory electronic reporting began in 2004. This improved quality of data, in particular, the collection of VTRSERNO,which improves matching to VTR data.

Tips 'n Tricks

- A dealer-veslog link can be made reasonbly well starting in 2005. To make this link, match the the CFDBS.VTRSERNO to VESLOG.SERIAL_NUM. Chances are that you care about Trip-level outcomes: be careful, because a vessel may have more than one SERIAL_NUM per TRIPID in the VESLOG tables.
- Here is a slick way to check confidentiality using dealer data:

```
 select year, port,
CASE WHEN LEAST(COUNT(distinct COALESCE(NULLIF(permit,'000000'),NULLIF(cf_license,'0'))),COUNT(distinct
```

```
from fso_admin.cfders_all_years@garfo_nefsc
where NESPP3 in (081, 082, 120, 122, 123, 124, 125, 147, 152, 153, 155, 240, 269, 512, 159, 250)
and year > 2018
group by year, port;
```

General Caveats

- Dealers are only required to record one VTR serial per trip.
- Outlier prices are always possible. Filter these out carefully.
- The following species are sketchy:
  - Surfclam and Ocean Quahog dealer reports are contained in the SFCLAM schema (separate from CFDBS). It is unclear whether reports of SC and OQ in CFDETS and CFDERS are duplicates or not, particularly for landings of Maine Mahogany Clam [Walden].
  - Giant Bluefin Tuna dealer are supposed to be reported individually and should be in a different schema. Giant Bluefin Tuna in CFDERS are either misreporting or duplication [George Silva, NMFS HMS].
  - Herring stock assessments do not necessarily use Dealer data as the source for assessments [Jon Deroba, PDB]. This is because Maine DMR has collected herring data and comprises the population of catch. The dealer data does not match the ME DMR data for herring quantity landed.
- There are many species that have two NESPP3 codes (As of March,2018).
  - Cod (080, 081)
  - Monkfish (011, 012)
  - Winter flounder (119, 120)
  - Yellowtail flounder (122, 123)
  - Plaice (124, 125)
  - Haddock (147, 148)
  - White hake (153, 154). 155 is Red/White mixed
  - Pollock (269, 270)

This bit of code may help:`select * from cfspp order by doc desc;`

- Some species were/are grouped together, but subsequently split apart.
  - Tilefish, which starts as NK, but becomes Blueline and Golden.
  - Skates

This bit of code may help:`select * from cfspp order by doc desc;`

- Data derived from "state" reporting may not include all fields that are populated by "federally reported" dealer reports. This may affect the PORT, COUNTY, PORT2, PERMIT, HULLNUM, VTRSERNO, MONTH, and DAY fields.
  - Permit numbers that do not correspond to a single vessel are:
    1. 000000, which means either "no vessel" (ex. from shore or aquaculture),or "unknown" federal permit, which could be "no federal permit".

2. 190998,390998 correspond to different size classes of vessels

- Many NESPP4 codes will not match well to VTR's SPPCODE. For example, VTR cod is all 0818 (unclassifed round). Almost all Cod will eventually be classified when sold; there is very little 0818 in dealer data.
- Ports are inconsistently encoded over time.
  - Some "port groups" were split into mutiple ports. (Hampton, Seabrook, and Hampton/Seabrook, NH is a good example).
  - Many records are entered only at the "state" or "county" level. This is particularly frequent for "older" records and non-federal reports that are received through SAFIS.
  - The names corresponding to the port codes may or may not match to Census "units." The 2 digit state code does not correspond to FIPS codes.
  - There is a table POPLACE_BASE that contains some Census Places. I'm not sure who made this, or if it's currently maintained.
  - month=0 or day=0 mean 'unknown' month or day. I believe that both are due to state-level reporting requirements that allow for monthly or yearly level reporting, instead of 'trip level' reporting.
- Live and Landed weights are recorded.
  - Scallop in-shell prices can be quite variable.
  - miscellaneous "parts" like monk liver, cod cheeks, or skate racks will have zero or null SPPLIVLB.
- The length distribution of landed fish might be useful. This code extracts the length distribution for cod in 2010:

```
select year, state, nespp4, length, sum(numlen) as num_len from cflen2010
where nespp3=081
group by nespp4, year, state, length;
```

Sample Projects

- Construct prices for fish [Lee, Demarest, Ardini].
- Construct trip revenues, post 2005 [Demarest]
- Commercial Landings and Revenues for the "Community Profiles." [Olson, Colburn]
- "The record" of commercial landings for use in stock assessment. Sort of. An "Area Allocation" usually need to be performed, because some species have multiple, spatially distinct stocks.[PopDy]
- Construct entity-level gross revenues from commercial fishing for Regulatory Flexibility Act Analyses [Lee].

Update Frequency and Completeness

- Nightly updates. Expect approximately 300 changes or additions to the current and previous year of data per day.

- Data is “complete” 6-9 months after the end of the calendar year; however, small changes are always occurring.
- This has consequences for reproducibility if you do not store a copy of the data.

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/12205
- NEFSC’s Data Dictionary http://nova.nefsc.noaa.gov/datadict/
- Preceded by: “Weighout” (WODETSyy and WODETTyy)
- Succeeded by: n/a

Related Tables

- CFDERSyyyyAA tables - “perform some Area Allocation”
- CFDETSyyyyAA tables - “perform some Area Allocation”
- CFDETTyyyy contains “detailed trip data” for 1994-2003
- CFSUMTyyyy, CFSUMSyyyy - “summary tables” for 1994-2003
- CFAGEyyyy - fish-level port sampling data.

Support Tables

- PORT, VALID_PORTS
- GEAR
- SPECIES_ITIS_NE decodes into names, links to the species_itis system
- CFSPP - decodes NESPP3 and NESPP4 into names

Table 1: Unique fields

Column	Description
SPPLNDLB	
SPPVALUE	
UTILCD	Quality unknown
DISPOSITION_CODE	Quality unknown
REPORTED_QUANTITY	
UOM	
GRADE_CODE	
MARKET_CODE	
SPPLIVB	Certain NESPP4 codes (monkfish livers, cod milt) convert into zero “live pounds.” This is done to prevent potential double counting during the stock assessment.

Table 2: Primary Source fields - These fields are firsthand data.

Column	Description
YEAR	This may not be the same as the year in which fish was caught.
MONTH	This may not be the same as the month in which fish was caught. May be zero.
DAY	This may not be the same as the day in which fish was caught. May be zero.
DEALNUM	Dealer Identification number
NESPP4 [3]	There are 4 species codes (NESPP3, NESPP4, WHSPP, SPECIES_ITIS).
WHSPP	
SPECIES_ITIS	
STATE_DNUM	

Table 3: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Description
PORT	Concatenation of state, port, county
COUNTY	Data dictionary claims this is a string, but it is a 2 digit code.
NEGEAR	
NEGEAR2	
NEMAREA	
AREA	
HARVEST_AREA	
DEPTHCD	
SUBTRIP	
TONCLASS [TONCL1,TONCL2]	
FZONE	
PERMIT	
HULLNUM	
VTRSERNO	
SPRATIO	
FIPS_STATE	
FIPS_PLACE	
FIPS_COUNTY	
CF_LICENSE	
NEGEAR_VTR	

Table 4: QAQC columns. Quality Control or Auditing fields.

Column	Description
LINK	
DOCN	
EFFIND	
SOURCE	
DERSOURCE	
PARTNER_ID	
DEALER_RPT_ID	
DOE	
LANDING_SEQ	

VESLOG Databases

This is partially complete. Let's call it 8 out of 10.

Tables: VESLOGyyyyT; VESLOGyyyyG; VESLOGyyyyS;

Location: Sole

Schema: VTR

Overview

The veslog data contains everything collected through the Vessel Trip Report System. These data are primarily generated through mandatory reporting by federally-permitted fishing vessels.

- 1994 to present
- one table per year

Current Collection Methods

These data are the result of mandatory federal vessel reporting. Federally permitted vessels are required to submit one VTR report per "gear-mesh-area" fished. 50CFR648.7(b)(1).

Changes to Collections Methods

- The VTR form has changed slightly over time. You can get the VTR form, with instructions at https://www.greateratlantic.fisheries.noaa.gov/aps/evtr/vtr_inst.pdf
- Electronic VTRs start in 2011.
- Some fisheries allow for (or require) reporting through a different system.
- In 2019, the PORT code and PORTLND1 variable has been frequently NULL. > When the (e-VTR) app was programmed the intent was to replicate the paper VTR but landed port was left off the fields to be entered. I'm not sure when/if this will be corrected and this was not well advertised as APSD did not know about this until July. The sail port was included so to the extent to which sail port and landed port may be the same would be an alternative work around. For recreational evtr's this is very likely since passengers need to get back to where their cars are parked. –Eric Thunberg, October 1, 2020.

Tips 'n Tricks.

- A dealer-veslog link can be made reasonbly well starting in 2005. To make this link, match the the CFDBS.VTRSERNO to VESLOG.SERIAL_NUM. Chances are that you care about Trip-level outcomes: be careful, because a vessel may have more than one SERIAL_NUM per TRIPID in the VESLOG tables.
- Vessels may declare out of fishing. The NOT_FISHED column in VESLOG_T can be used to filter these out.

- The TRIPID is unique to a "fishing trip."

  - A TRIPID to should match to at least one GEARID.
  - A TRIPID to should match to at least one CATCH_ID (if a vessel caught any fish)
  - The NRPAGES and NSUBTRIP Columns will be something other than `1` if there are more than one GEARID corresponding to a particular TRIPID.
- The GEARID is unique to the "gear-mesh-area" fished. There should be a one-to-one correspondence between GEARID and SERIAL_NUM.
  - A GEARID to should match to exactly 1 TRIPID
  - A GEARID to could match to 0 CATCH_ID if no fish was caught.
  - A GEARID is far more likely to match to 1 or more CATCH_IDs though.
- The CATCH_ID is unique to the "SPPCODE- DEALNUM-dealer-gearcode-mesh-area" fished.
  - A CATCH_ID to should match to at exactly 1 GEARID. And therefore exactly 1 TRIPID).
- DATESAIL, DATELND1, DATELND2 include clock time. You can sometimes get negative trip durations due to reporting or data entry errors.

General Caveats.

- 1994-1995 are kind of sketchy
- Electronic VTRs have very long SERIAL_NUM, TRIPID, GEARID, and CATCH_ID. Some software doesn't like this (Excel, stata) – you might want to do this:`select to_char(g.serial_num) from veslog2014g g`
- Some of the older numbers (from 1994-1995) are non-numeric.
- All quanties are "Hail Weights," which are the operator's best estimate of catch.
- SPPCODES will not match well to dealer's NESPP4 codes. For example, VTR cod is all 0818 (unclassifed round). Almost all Cod will eventually be classified when sold; there is very little 0818 in dealer data.
- The following species are sketchy:
  - Surfclam and Ocean Quahog dealer reports are contained in the SFCLAM schema (separate from VESLOG). This might create inaccuracies for trips that land SF/OQ and other species (two un-linkable logbooks).
  - Lobster coverage is not full. All lobster landings by vessels holding other federal permits are supposed to be in VTR.
  - Giant Bluefin Tuna dealer are supposed to be reported individually and should be in a different schema. Giant Bluefin Tuna in CFDERS are either misreporting or duplication [George Silva, NMFS HMS].
- PORTLND1 and PORTLND2 are inconsistently encoded over time.
  - The names corresponding to the port codes may or may not match to Census "units." The 2 digit state code does not correspond to FIPS codes.
  - The PORTLND1 and PORTLND2 fields are data entered and error-corrected on the fly[Lee, Dentremont] This means that data entry of `NEW ROCHELL, NY` is autocorrected to `NEW ROCHELLE,`

NY and coded as PORT=350739
  - This might provide more insight:`select * from vtr.vlportsyn order by doc desc;`

- CAREA, CNEMAREA, CLATDEG, CLATMIN, CLATSEC, CLONDEG, CLONMIN, CLONSEC, CERRNO, AREA_IND, TENMSQ
  - The "C" stands for calculated. The calculation is pretty complicated. [Lee, H. McBride]
  - The "AREA" refers to the Northeast region statistical areas.
    * AREA in the 300s and 400s are in Canada
    * AREAS under 200 are Inshore. They shouldn't show up, but occasionally do.
  - "NEMAREA" includes "inshore areas"
    * The latitude and longitude points reported on a VTR are first binned into a Ten Minute Square, then checked against a lookup table that converts Ten Minutes Squares into statistical areas [LOC2AREAS].
    * If there is a match, then the reported latitude and longitude points are accepted.
    * If there is no match and the reported lat-lon is in an AREA adjacent to the reported AREA, then the lat-lon points are accepted and the reported statistical area is replaced by the AREA corresponding to the lat-lon.
    * If there is no match and the reported lat-lon is *not* in an AREA adjacent to the reported AREA, then the lat-lon points are rejected and converted to `NULL` values. The reported statistical area used as the CAREA.

  - Technically, only degrees and minutes are required to reported.
  - Some vessels report LORAN readings. These are converted to lat-lon prior to that entire process.
  - This QAQC step is performed differently in the GARFO CATCH, IMAGES, DOCUMENT data.

- Some dealer numbers (DNUM in VESLOG_S) indicate that catch was not sold to a federally permitted dealer (DNUM<=8 or DNUM=99998)

- Recreational trips report numbers of fish, not pounds.

- PORTLND1 and PORTLND2 are recorded. Sailing port is not recorded
  - Use landing location from previous trip?
  - Assume round trip to same port?

- The code for WOLFFISH is CAT. The code for BLUE CATFISH and NS Catfish are CATB and CATNS respectively. This is hilarious. If you need to pick WOLF apart from CATFISH, wolffish is not caught inshore.

- The code for WHITE HAKE is WHAK. White HAKE is not Whiting. Sometimes operators write "WHAK" for "WHITING - HAKE." They should be writing SHAK, HAKOS, or WHB. You can separate based on mesh size if necessary. [Thunberg]

- There are some errors on OPERNUM and OPERATOR. This happens mostly because of transcription mistakes
  - The captain writes his name, but not the number and that name is incorrectly linked to the OPERNUM.

– The captain's writes his number, but it doesn't agree with the NAME This type of error seems to happen infrequently, but often enough to be a problem.

Sample Projects

Update Frequency and Completeness

- Nightly updates. Expect approximately 300 changes or additions to the current and previous year of data per day.
- Data is "complete" 6-9 months after the end of the calendar year; however, small changes are always occurring.

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/1423
- NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/
- Preceded by: "none"
- Succeeded by: n/a

Related Tables very incomplete.

- CATCH, IMAGES, DOCUMENT - these are the "GARFO" version of VTR. C,I,D are a single table, not one per year. I can't even.

Support Tables very incomplete.

- VLSPPTBL decodes SPPCODES into names and NESPP4 codes. So does VLSPPSYN_94_95, which looks deprecated.

- TENMINSQ, LOC,LOC2AREAS
- PORTSYN, VLPORTSYN
- VLGEAR - decodes gear codes into english

Table 5: Fields in VESLOG

Column	Location	Description
CAREA	G	
CATCH_ID	S	
CLATDEG	G	
CLATMIN	G	
CLATSEC	G	
CLONDEG	G	
CLONMIN	G	
CLONSEC	G	
CNEMAREA	G	

Column	Location	Description
CREW	T	
DATE_SIGNED	T	
DATELND1	T	This includes time
DATELND2	T	This includes time
DATESAIL	T	This includes time
DATESOLD	S	
DEALNAME	S	
DEALNUM	S	
DEPTH	G	
FZONE	G	
GEARCODE	G	
GEARID	G S	
GEARQTY	G	
GEARSIZE	G	
HULLNUM	T	
MESH	G	
NANGLERS	T	
NHAUL	G	
NOT_FISHED	T	
NRPAGES	T	
NSUBTRIP	T	
OPERATOR	T	
OPERNUM	T	
PAGENO	G	
PERMIT	T	
PORT	T	
PORTLND1	T	
PORTLND2	T	
QDSQ	G	
QTYDISC	S	
QTYKEPT	S	
SERIAL_NUM	G	
SOAKHRS	G	
SOAKMIN	G	
SPPCODE	S	
STATE1	T	
STATE2	T	
SUBTRIP	G	
TENMSQ	G	
TIMELND1	T	
TIMELND2	T	
TIMESAIL	T	

Column	Location	Description
TRIP_ACTIVITY_TYPE	T	
TRIPCATG	T	
TRIPID	T G S	

Table 6: Primary Source fields - These fields might be useful for linking, either exactly or fuzzy.

Column	Location	Description
DATELND1	T	
DATELND2	T	
DATESAIL	T	
DATESOLD	S	
DEALNAME	S	
DEALNUM	S	
HULLNUM	T	
TIMELND1	T	
TIMELND2	T	
TIMESAIL	T	
PERMIT	T	
PORT	T	
PORTLND1	T	
PORTLND2	T	
SERIAL_NUM	G	

Table 7: QA/QC columns. Quality Control or Auditing fields.

Column	Location	Description
AREA_IND	G	
BATCHID	T S	
DATE_RECV	G	
DC	T G S	
DE	T G S	
FILENAME	G	
IMG_DATE	G	
IMGTYPE	G	
UC	T G S	
UE	T G S	
SIDEID	G	

PERMIT Databases

This is very incomplete. Let's call it 6 out of 10.

Tables: VPS_VESSEL; VPS_FISHERY_NER

Location: Sole

Schema: PERMIT

Overview

The permit data contain information about fishing vessels, the fisheries they can participate in. These data are generated through permit applications that are submitted to GARFO by vessel owners.

- VPS_VESSEL
- VPS_FISHERY_NER
- BUS_OWN
- VPS_OWNER

Current Collection Methods

Changes to Collections Methods

Tips 'n Tricks

General Caveats

- There is an AP_NUM. This is different from the APP_NUM in the MQRS system.
- The same permit number does not necessarily link to the same vessel through all years, as vessels may be upgraded or replaces.
- The same permit number does not necessarily link to the same owner over time, as vessels (with accompanying permits) can be sold.
- The are 5 date fields in VPS_FISHERY_NER
- There have been "clean ups" – sometimes an entry gets put into VPS_VESSEL where the DATE_ISSUED is equal to the DATE_CANCELED. This may cause mis-merges. You also may want to exclude these completely from your query.

```
select * from vps_vessel where (date_issued<>date_canceled or date_canceled is null);
```

- Similarly, you may want to exclude rows where start_date >=end_date or start_date>=date_expired
- BUS_OWN and VPS_OWNER are used to aggregate multi-vessel firms together for Regulatory Flexibility Act Analysis.
- Scallop and Groundfish "ownership" has been databased going back to 1996 (OWNER_HISTORY_PIDS). Not exactly relevant, but could track "income" or something else at the firm level for a sbuset of those two fisheries. Be careful when thinking about

- If a vessel is owned by one or more people or businesses in a partnership, then we use rank 2, 3, etc. for the additional owners in vps_owner. The rank field does not represent any percentage ownership, hierarchy in the company, or majority ownership, it is simply used for mailing purposes. The rank = 1 owner name shows up on the permit and we use that name for our mailings. (Ted Hawes - Sept 23, 2020)
- In the vps_owner table, we only populate the business_id for the rank = 1 records. Add a condition that rank = 1 to ensure that there are no missing business ids. If you do, that would be an error. (Ted Hawes - Sept 23, 2020)
- The owners/shareholders of Company A are entered in the Client system and given a person_id. Each person in the system has a unique id that can be reused for any of the businesses that they may be a part of. For new companies or changes in corporate owners, we then assign each person to the company. Each year with our renewal applications, we collect the owner information and compare it to what we have in the system. If there are no changes in owners, we don't alter any of the business information. If a person is added to a company, we generate a person_id for them, if they don't already have one, and assign them to the business. If a person leaves a company or dies, we don't delete them from the business. We place an end_date on their relation with the company so that we can have a history of the owners. (Ted Hawes - Sept 23, 2020)

Sample Projects

Update Frequency and Completeness

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/12985
- NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/

Related Tables

- BUS_OWN contains ownership data that is often linked to these PERMIT data. We'll put them in a separate section.
- MQRS tracks the ability of a *thing* to participate in a moratorium fishery. We'll put them in a separate section. Allocations are tracked in other places (DAS, DAS2, and AMS).
- CPH
- OPERATORS (captains) are permitted independently

Support Tables

- VALID_FISHERY translates the PLAN and CAT into words.

```
select * from permit.valid_fishery where descr like '%SCALLOP%' order by plan, cat, permit_year
```

Table 8: Unique fields

Column	Description
Placeholders	

Table 9: Primary Source fields - These fields are firsthand data.

Column	Description
Placeholders	

Table 10: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Description
Placeholders	

Table 11: QAQC columns. Quality Control or Auditing fields.

Column	Description
Placeholders	

DMIS Databases

This is complete. 10 out of 10.

Tables: APSD.t_ssb_trip@garfo_nefsc APSD.t_ssb_catch@garfo_nefsc APSD.t_ssb_discard@garfo_nefsc

APSD.t_ssb_trip_current@garfo_nefsc APSD.t_ssb_catch_current@garfo_nefsc APSD.t_ssb_discard_current@garfo_nefsc

Location: GARFO super secret server

Schema:

Overview

The Data Matching and Imputation System is maintained by GARFO. Just email Michael Lanning at GARFO, he's got all the answers. j.michael.lanning@noaa.gov

Linking to Veslog

Linking to VESLOG with DMIS DOCID has a few issues. (Chad Demarest, May 14, 2020) 1. JML adds digits to DMIS DOCID to denote subtrips. 1. Another is that some data handling protocols (varies by file type) truncate a digit off the end of the EVTR serial numbers. 1. A third is that DMIS will Give positive (mostly correct) match's where the DOCID fails for these and other reasons.

On Home Consumption Fish

> The code for this stuff is BHC_ (either _LIVE_POUNDS or _LANDED_POUNDS). "BHC" stands for "Bait and Home Consumption". A few years ago they added LUMF (Legal UnMarketable Fish) to this category as well. LUMF, at lease for here, and at least as I understand it, is derived from observer trips only. But I'm not 100% sure on that.
>
> If you use DLR_LIVE or DLR_LANDED (or DLR_DOLLAR) you won't get the BHC fish. If you use LANDED or POUNDS or DOLLAR_SSB/GDP you'll get 'em, plus imputed values for the DOLLAR field.
>
> Starting on the next run of DMIS, fish that are authorized to be landed on EFP trips but are not sold thu a dealer will be added to the BHC_ fish.

Chad's email April 11, 2018.

Versions

There is a "_current" versioning of tables. As of April 2020, current contains a little bit more data.
> I now use _trip for everything, whereas a yearly ago I used _current. There were some discrepancies for earlier years (2013?) that _trip had correct and _trip_current did not. Otherwise I believe they are the same. (Chad)

I have been using current since there is more recent data included. Hopefully for older data the two match up, but I haven't looked into that in a while. (Greg)

AMS, DAS, DAS2 Databases

This is a partially complete. Let's call it 6 out of 10.

Tables: AMS, DAS, DAS2

Location: Sole

Schema: AMS, DAS, DAS2

Overview

These are the three schema that track Allocations. DAS begat DAS2. DAS2 begat AMS.

Changes to Collections Methods

Tips 'n Tricks

General Caveats

- You can't simply "stack" data from these three datasets because there are some duplicate entries. I think this happened because GARFO transitioned from one to the next and continued to record in both.
  - Use DAS for multispecies fishing years 2004 and 2005
  - Use DAS2 for FYs 2006-2008 inclusive. Scallop and multispecies allocations are stored here.
  - use AMS for 2009 and after. Scallop and multispecies allocations are stored here. There may be other stuff too.
  - I think these guidelines are okay for other species, but I'm not positive.

- In general, there are (BASE) Allocations, base allocation adjustments, carryover, Leases (in and out), transfers, and sanctions.

AMS stores transfers in a funky way:

> The allocation_transfer table has a different structure than the ams.allocation_tx table.
>
> The column is called allocation number,and the value may be an MRI. That is because GC scallop has MRI's but tilefish does not, It uses allocation numbers. They serve the same purpose, which is to uniquely identify an access privilege, whether it is an MRI or allocation number.
>
> The from/to tells what allocation number the pounds came from or to. If the amount is positive, the from/to column contains the allocation number that bought the pounds. If the amount is negative the from/to column contains the allocation number that sold them.
>
> In AMS, there are two rows, one for the seller and one for the buyer. . They each have columns called "root_mri" and "charge_mri" - which is misleading because the tilfish entries are not MRI's [Steve Cohen @ GARFO. April, 2016]

- I've embedded some code to extract initial allocations, usage, leases, and transfers. I think it works for multispecies, but you might want to double check before just running it.

- Initial Allocations
  To extract initial allocations from DAS:

```
select * from das.das_allocation
  where fishery='MUL' and das_category='A'
    and fishing_year between 2004 and 2005;
rename fishery fmp;
rename das_category das_type;
rename right_to_days_id mri;
keep if transfer_id==.;
keep mri das_net_allocation fishing_year;
rename das_net categoryA_DAS;
```

To extract initial allocations of A-Days from DAS2:

```
SQL: "select * from das2.allocation where plan='MUL' and category_name='A';"
STATA:
rename plan fmp;
rename category_name das_type;
destring, replace;
keep if fishing_year>=2006 & fishing_year<=2008;
collapse (sum) quantity, by(right_id credit_type fishing_year);
 drop if inlist(credit_type,"LEASE IN", "LEASE OUT");
rename right_id mri;
 rename quantity categoryA_DAS;
```

To extract initial allocations from AMS:

```
SQL: select *  from ams.allocation_tx where FMP='MULT' and das_type='A-DAYS';
STATA:
destring, replace;
keep if fishing_year>=2009;
collapse (sum) quantity, by(fishing_year allocation_type root_mri);
rename allocation_type credit_type;
rename root_mri right_id;
 drop if inlist(credit_type,"LEASE IN", "LEASE OUT");
 rename quantity categoryA_DAS;
```

Then, stack them all together.

- Transfers and leases.

To get leases from DAS

```
SQL: select * from das.das_transfer_lease where fishery='MUL' and das_category='A'
  and TRANSACTION_TYPE='L' order by nmfs_approval_date desc;")

Stata:
keep if inlist(fishing_year, 2004, 2005);
```

```
rename grantor_right_to_days_id right_id_seller;
rename grantee_right_to_days_id right_id_buyer;

rename grantor_permit_number permit_seller;
rename grantee_permit_number permit_buyer;
drop user_changed date_changed user_entered transaction_type;
replace nmfs_approval_date=dofc(nmfs_approval_date);
replace date_entered=dofc(date_entered);

format nmfs_approval_date date_entered %td;
rename das_leased quantity;
rename das_price dollar_value;
rename nmfs_approval date_of_trade;
drop fishery date_entered das_category;
```

To get leases from DAS2

```
SQL:
select * from das2.allocation_use where category_name='A' and plan='MUL'
    and allocation_use_type='LEASE' and approval_status='APPROVED' ;

Stata:
keep if fishing_year>=2006 & fishing_year<=2008;

gen date_of_trade=dofc(au_date_time);
format date_of_trade %td;
collapse (sum) quantity (first) dollar_value, by(permit_d permit_c right_id
    right_id_c date_of_trade fishing_year);
rename permit_d permit_seller;
rename permit_c permit_buyer;
rename right_id_c right_id_buyer;
rename right_id right_id_seller;
order date fishing_year;
sort fishing_year date permit_s;
* there were a few data correction issues, mostly wrong/bad matches.
replace right_id_buyer=1807 if right_id_buyer==. & permit_buyer==121546;
replace right_id_buyer=559 if right_id_buyer==. & permit_buyer==310912;
replace right_id_buyer=455 if right_id_buyer==. & permit_buyer==251364;
replace right_id_buyer=2055 if right_id_buyer==. & permit_buyer==149334;
```

To get AMS lease data. I'm not 100% sure this is correct.

```
SQL "select lease_exch_id,from_permit, to_permit, from_right, to_right, fishing_year, quantity, price,
approval_date  from ams.lease_exch_applic
        where FMP='MULT' and from_das_type='A-DAYS'
        and approval_status='APPROVED' and fishing_year>=2009;"
```

```
Stata:
destring, replace;
rename to_permit permit_buyer;
rename from_permit permit_seller;
rename to_right right_id_buyer;
rename from_right right_id_seller;
compress;
rename approval_date date_of_trade;
rename price dollar_value;
replace date_of_trade=dofc(date_of_trade);
format date_of_trade %td;
gen schema="AMS" ;
```

You should be able to stack all three of these together.

- USAGE DAS and DAS2 keep track of usage differently.
  To extract usage from DAS:

```
Part 1
SQL Code:
select du.fishing_year, du.das_transaction_id, du.permit_number, du.das_charged_days, tr.sailing_port,
tr.sailing_state, tr.sail_date_time as date_sail, tr.landing_port, tr.landing_state,
tr.landing_date_time as date_land,tr.gillnet_vessel, tr.day_trip,
tr.observer_onboard, tr.das_charged_fixed, tr.fishery_code,
tr.vessel_name  from das.das_used du, das.trips tr
  where du.das_transaction_id=tr.das_transaction_id
  and du.permit_number=tr.permit_number
  and du.das_category='A' and du.fishery='MUL';") ;

Stata code:
keep if inlist(fishing_year, 2004, 2005);
rename permit permit;
rename das_charged_days charge;
rename fishery_code activity_code;

collapse (sum) charge (first) gillnet_vessel day_trip observer_onboard das_charged_fixed vessel_name ,
    by(fishing_year permit sailing_port sailing_state date_sail
    landing_port landing_state date_land activity_code );
gen schema="DAS";
```

To extract usage from DAS2:

```
Part 2a
SQL: " select du.das_trip_id, du.allocation_use_type, du.au_date_time_debited, du.permit_debited,
du.permit_credited, du.quantity, du.category_name, du.plan, du.right_id,du.credit_type,
du.fishing_year, du.dollar_value, activity_code, dt.permit_num, dt.sailing_port,
```

```
dt.sailing_state, dt.trip_start, dt.trip_end, dt.landing_port, dt.landing_state
  from das2.allocation_use du, das2.das_trip dt
     where du.category_name='A' and du.plan='MUL'
     and du.allocation_use_type='TRIP'
     and du.quantity<>0 and du.das_trip_id=dt.das_trip_id;"

Stata:
keep if fishing_year>=2006 & fishing_year<=2008;
tempfile t1 reg_use2006;
rename permit_debited permit_seller;
rename right_id right_id_seller;
rename permit_credited permit_buyer;
keep if allocation_use_type=="TRIP";
notes: This has ONLY TRIPS. This has no PTU in it.;
```

If DAS were leased and then subsequently used, these trips were stored in a different place. God knows why.

```
Part 2b
SQL: "select du.das_trip_id, du.allocation_use_type, du.au_date_time_debited, du.pt_permit_debited,
du.permit_debited, du.quantity, du.category_name, du.plan, du.right_id, du.credit_type,
du.fishing_year, activity_code, dt.permit_num, dt.sailing_port, dt.sailing_state,
dt.trip_start, dt.trip_end, dt.landing_port, dt.landing_state
 from das2.private_transaction_use du, das2.das_trip dt
 where du.category_name='A' and du.plan='MUL'  and du.quantity<>0
 and du.das_trip_id=dt.das_trip_id;"

 keep if fishing_year>=2006 & fishing_year<=2008;
```

Stack the results of parts 2a and 2 and then

```
rename permit_num permit;
rename quantity charge;
rename trip_start date_sail;
rename trip_end date_land;
drop plan category;
collapse (sum) charge, by (au_date_time_debited permit_debited permit right_id fishing_year
activity_code sailing_port sailing_state date_sail date_land landing_port landing_state);
gen schema="DAS2";
```

To get Days at sea from AMS:

```
SQL:
"select * from AMS.TRIP_AND_CHARGE where fmp='MULT' and DAS_TYPE='A-DAYS'
  and charge<>0 and fishing_year>=2009;"

Stata:
drop running_clock observer rsa mult_override fmp trip_de-charge_uc trip_source
```

```
fishing_area das_type das_id trip_id tc_id charge_type;

rename permit_nbr permit;
destring, replace;
compress;
 collapse (sum) charge, by(permit date_s date_l fishing activity_code);
 gen schema="AMS";
```

Charge is denominated in days. There was one entry in AMS that was negative. I don't think it's a data error.

This is not quite working properly - there seem to be some entries in AMS.TRIP_AND_CHARGE that indicate that multispecies A-days are charged to vessels that don't have any A-DAYS, a right_id for Multispecies, or even a Limited Access multispecies permit. It seems like this is from MNK. These should be excluded, probably by matching to valid per_nums in mqrs.mort_elig_criteria.

You should be able to stack the results of all of these to form a giant table of DAS usage from 2004-present. Permit, charge, fishing_year, date_sail, date_land, fishing_year and activity_code are in all. Sailing port, state, landings port state, gillnet, and observer were only databased in AMS. DAS2 contains right_ids, but DAS and AMS do not.

Sample Projects

Update Frequency and Completeness

- Not sure about AMS.
- DAS and DAS2 are "legacy" GARFO tables. GARFO doesn't really care about maintaining them.

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/11773
- NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/

Related Tables

Support Tables

Table 12: Unique fields

Column	Description
Placeholders	

Table 13: Primary Source fields - These fields are firsthand data.

Column	Description
Placeholders	

Table 14: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Description
Placeholders	

Table 15: QAQC columns. Quality Control or Auditing fields.

Column	Description
Placeholders	

MQRS Databases

This is mostly a placeholder. Let's call it 2 out of 10.

Tables: MORT_ELIG_CRITERIA, SECTOR_YEAR_ROSTER

Location: Sole

Schema: MQRS

Overview

The MQRS system is fun. It contains information on "eligibilities." The sector rosters are also stored here. It is used in conjunction with other tables, but not cross-referenced against them for logical consistency.

Current Collection Methods

Someone at GARFO puts data into the MQRS databases.

Changes to Collections Methods

Tips 'n Tricks

General Caveats

- The sector rosters are stored in "SECTOR_YEAR_ROSTERS." These contain a sector_id, year, and MRI (among other things.) There are also "views" (SECTOR_PARTICIPANTS_CPHYYYY and SECTOR_PARTICIPANTS_NO_CPHYYYY). These link the MRI to the PERMIT. Unfortunately, they are all broken as of April 18, 2018. They are broken because the underyling code does not properly check dates when it links the MRI to a permit. Here is code that will illustrate:

1. MRI 1755 should link to permit 100598 any time between april 26, 1994 and june 17, 2015

```
select per_num, vessel, owner, date_eligible, date_cancelled, right_id, auth_id
from mqrs.mort_elig_criteria where right_id=1755 order by  date_eligible;
```

2. None of these do that properly:

```
select * from mqrs.sector_participants_CPH where mri=1755;
select * from mqrs.sector_participants_CPH2011 where mri=1755;
select * from mqrs.sector_participants_CPH2011 where mri=1755;
```

This is happening in part because the underlying code forgot to screen for valid dates.

- There is an APP_NUM. This is different from the AP_NUM in the PERMIT system.
- There are business names. They do not necessarily match the business names from the permit system.
- History Retention and CPH are essentially synonyms.
- When an owner puts a right into CPH, the last known PER_NUM associated with that RIGHT_ID is stored.

we don't change the relationship between the CPH and the permit if a new owner or the same owner places a new set of permits on the vessel. The CPH records remain as is until they are moved to another vessel. We place the permits in CPH based on the last known or permitted vessel. [Ted Hawes @ GARFO, April 18, 2018]

This can create PER_NUMs with more than 1 RIGHT_ID. Ted Hawes suggests adding

```
auth_type not in ('CPH','HISTORY RETENTION')
```

and then

```
auth_type  in ('CPH','HISTORY RETENTION')
```

And stitching results together with a union.

- Sometimes an entry gets put into the MQRS database where the DATE_ELIGIBLE is equal to the DATE_CANCELLED. This may cause mis-merges. You also may want to exclude these completely from your query.

```
select * from mqrs.mort_elig_criteria
where date_eligible<>date_cancelled or date_cancelled is null;
```

- There was a "cleanup" in 2009. Perhaps in March? [Ted Hawes]
- Auth_id is really only used internally. auth_id is a unique field for each transaction, and from the original creators of this system back in the early 1990s, the auth_id authorizes the issuance of a permit to the vessel or it authorizes allocation to be issued to the vessel. The idea was that the auth_id would be used to connect the permit system with mort_elig_critieria when the permit staff processes an application for a limited access permit. MQRS will authorize the issuance of a limited access using the auth_id. [Ted Hawes]
- The auth_id and right_id can be the same when a new limited access permit is created. Alternatively, when auth_id=right_id, the right has never been transferred.[Ted Hawes]
- There are at least a handful of cases in which a different system (DAS, DAS2 mostly) links to an incorrect right id. For example:

```
select * from DAS.DAS_ALLOCATION where right_to_days_id not in (select distinct right_id from MQRS.mort
```

will extract all the entries in the DAS.DAS_ALLOCATION table that have an MRI that is not in the MQRS.mort_elig_criteria. This shouldn't be possible, because an entity needs an MRI

similarly:

```
select distinct right_id  from das2.allocation where plan='MUL' and
right_id not in (select distinct right_id from mort_elig_criteria where fishery='MULTISPECIES')
```

Here is a partial list of the MRI's affected by the cleanup. This came from Ted Hawes. But you could look at the "remark" field and see if there is a reference to a cleanup.

```
select PER_NUM AS PERMIT,
        AUTH_ID AS MRI,
        FISHERY,
```

```
        DATE_ELIGIBLE,
        DATE_CANCELLED
        from mort_elig_criteria where auth_id in(1179,1183,1187,1196,1219,1255,1261,1293,1296,1362,1374
```

The date_eligible and date_cancelled fields were broken during the cleanup. I'm not sure the best way to deal with it.

Here's some code to link right_ids to permits for just the multispecies fishery, courtesy of Dan Caless.:

```
SELECT PER_NUM AS PERMIT,
        RIGHT_ID AS MRI,
        FISHERY,
        DATE_ELIGIBLE,
        DATE_CANCELLED,
        AUTH_TYPE,
        ELIG_STATUS
      FROM MQRS.MORT_ELIG_CRITERIA
      WHERE FISHERY = 'MULTISPECIES'
        AND not ((TRUNC(DATE_ELIGIBLE) =  TRUNC(NVL(DATE_CANCELLED,SYSDATE+20000))) AND (CANCEL_REASON_C
        AND DATE_ELIGIBLE IS NOT NULL
        AND (TRUNC(DATE_CANCELLED) >= '01-MAY-03' or DATE_CANCELLED IS NULL):
```

Sample Projects

Update Frequency and Completeness

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/12987
- Missing from NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/

Related Tables

- BUS_OWN contains ownership data that is often linked to these PERMIT data. We'll put them in a separate section.
- CPH
- OPERATORS (captains) are permitted independently

Support Tables

- VALID_FISHERY translates the PLAN and CAT into words.

Table 16: Unique fields

Column	Description
Placeholders	

Table 17: Primary Source fields - These fields are firsthand data.

Column	Description
Placeholders	

Table 18: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Description
Placeholders	

Table 19: QAQC columns. Quality Control or Auditing fields.

Column	Description
Placeholders	

SFCLAM Databases

This is a placeholder. Let's call it -1 out of 10.

OBDBS Databases

This is a placeholder. Let's call it -3 out of 10.