Document (01) Data Collection & process

15:30-17:00 April 17 (Sunday)

Outline and general views

Progress on actual data collection and process
for longtail tuna and kawakawa
will be presented in sessions on
CPUE standardization and ASPIC

Contents

- Data collection (general)
- Data collection (ASPIC)
- Progress of the data collection
- Data processing skill
- Data message and massage?
- Sample size
- Data catalogue

Document (01) Data Collection & process

15:30-17:00 April 17 (Sunday)

Starting point

Most basic work

before CPUE standardization+ stock assessments

Why

Without data, you can not do (analyses) anything.

Data collection (general)

What do you need to collect ? Depending on the Objectives

Government (Management) :

➔ Catch for national statistics and planning

Research

Research	catch	nominal CPUE (C+E)	Biology	Ecology	Environment
CPUE standardization					
Biological study					
stock assessments					
Stock structure					
habitat					

Data collection

How to collect ?

- Port sampling (sub sample, interview, eye ball..)
- Observer program (human and video system) link…
- Logbook (paper and Electronic logbook)
- Catch report from fishing company

How detail we need to collect

- Ideally set-by-set data is perfect

  ➔ less bias to estimate for catch, CPUE Stock structure etc..

➔ Information by species, gear, area, year, month, day, boat and set

- But if we can not.. We collect aggregated data

For example by species, gear, area, year and month.

(normal situation)

Data collection for ASPIC

What do we need to collect for our case
CPUE standardization＋stock assessments（**ASPIC**）
by stock and species

CPUE standardization➔ Nominal CPUE (catch and Effort)

by set (gear, fishing area, year, month, day, boat and set)

ASPIC➔ Annual **global Catch** by country

+ standardized CPUE

How about our case ? Not real data collection ➔ Compilation

➔ Collect & compile available historical catch +CPUE (ASPIC)(8 countries)

- Data collection (**catch**)
  - Annual total catch by Country and Ocean (Indian and Pacific)
    - ➔ no need by gear and area
    - ➔ ASPIC needs the annual global catch

- Data collection : **CPUE** (catch and effort) (set by set)
  - by gear, Ocean, area, year, month, day and Boat
    - ➔ CPUE standardizations needs fine scale data
    - ➔ if not available, aggregated data OK

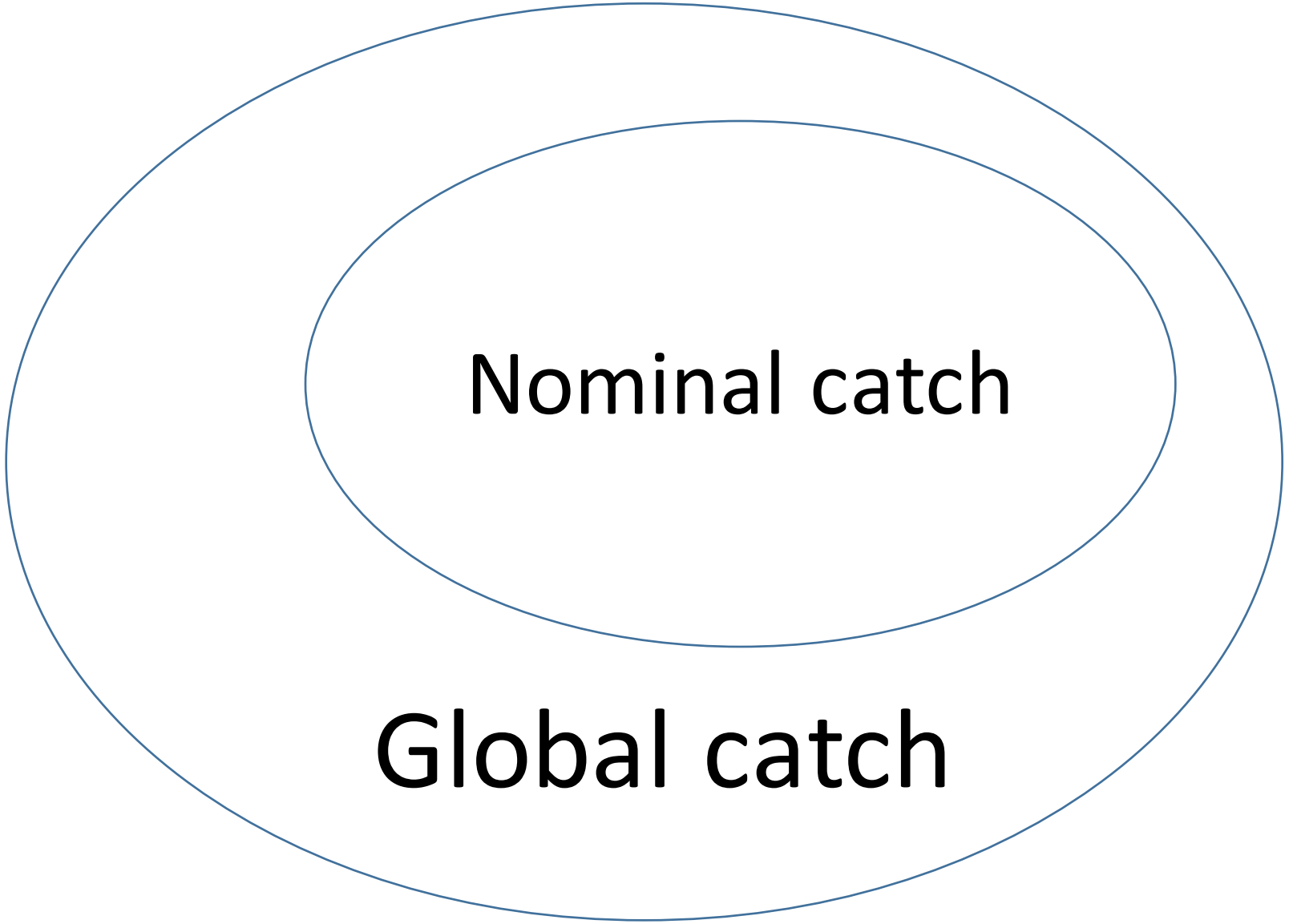
Quiz

Global catch [1/]
vs nominal CPUE (nominal **Catch** [2/] and effort)

What are differences between

Global Catch [1/] and nominal **Catch** [2/]

Same ?


Nominal catch
Global catch

(sometimes)

Global catch

=

Nominal catch

Progress of the data collection

Data collection by through coordinators in 8 member countries (completed in December, 2015)

ONLY Thailand and Philippines long time series of CPUE

(key parameters for ASPIC)

Special mission (further investigations) (Jan/Feb, 2016)

Nishida+ Sweetness (SEAFDEC)

Data coordinators (8 member countries) (Completed in Dec., 2015)
What is the stock (1 or 2)

No	Country	stock	Member country	Coordinators	Post	Agency	e-mail
1	Brunei	1	Brunei	Ms Noorizan Karim	Head	Capture Fisheries Industry Division	noor6263@gmail.com
				Mr Matzaini Juna	Head		matzaini.juna@fisheries.gov.bn
2	Cambodia	1	Cambodia	Mr. Suy Serywath	Director	Fisheries Research and Development Institute, Fisheries Administration (FiA)	serywath@gmail.com
				Mr. Kao Monirith	Deputy Director		kaomonirith@yahoo.com
3	Indonesia	2	Indonesia	Dr Khairul Amri	Chief scientist	Research Institute for Marine Fisheries	Kh_amri@yahoo.com
				Mr Thomas Hidayat	Researcher		hidayatthomas245@gmail.com
4	Malaysia	2	Malaysia	Mr Samsudin Bin Basir	Chief scientist	Department of Fisheries	s_basir@yahoo.com
				Mr Sallehudin Jamon	Scientist		dinjamon@rocketmail.com
							sallehudin_jamon@dof.gov.my
5	Myanmar	1	Myanmar	Dr Htun Thein	Assistant Director	Marine Resources Survey & Research Unit, Department of Fisheries	htunthein.akyab@gmail.com
				Mr Nay Myo Aye	Assistant Fisheries Officer		nvnghia@rimf.org.vn
6	Philippines	1	Philippines	Mr Noel Barut	Director	Bureau of Fisheries and Aquatic Resources (BFAR)	noel_c_barut@yahoo.com
				Ms Grace Lopetz	Aquaculturist II		gmvlopez@yahoo.com
7	Thailand	2	Thailand	Ms. Suwantana Tossapornpitakkul	Fishery Biologist	Marine Fisheries Technology Research and Development Institute, Department of Fisheries	tsuwantana@yahoo.com
				Mr. Chalit Sangangam	Fishery Biologist		chalitster@gmail.com
8	Viet Nam	1	Viet Nam	Mr Nguyen Viet Nghia	Deputy Director	Research Institute for Marine Fisheries	nvnghia@rimf.org.vn
				Mr Pham Hung	Officer	Research Institute for Marine Fisheries	hungfam83@gmail.com

Stock structure

IOTC
LOT

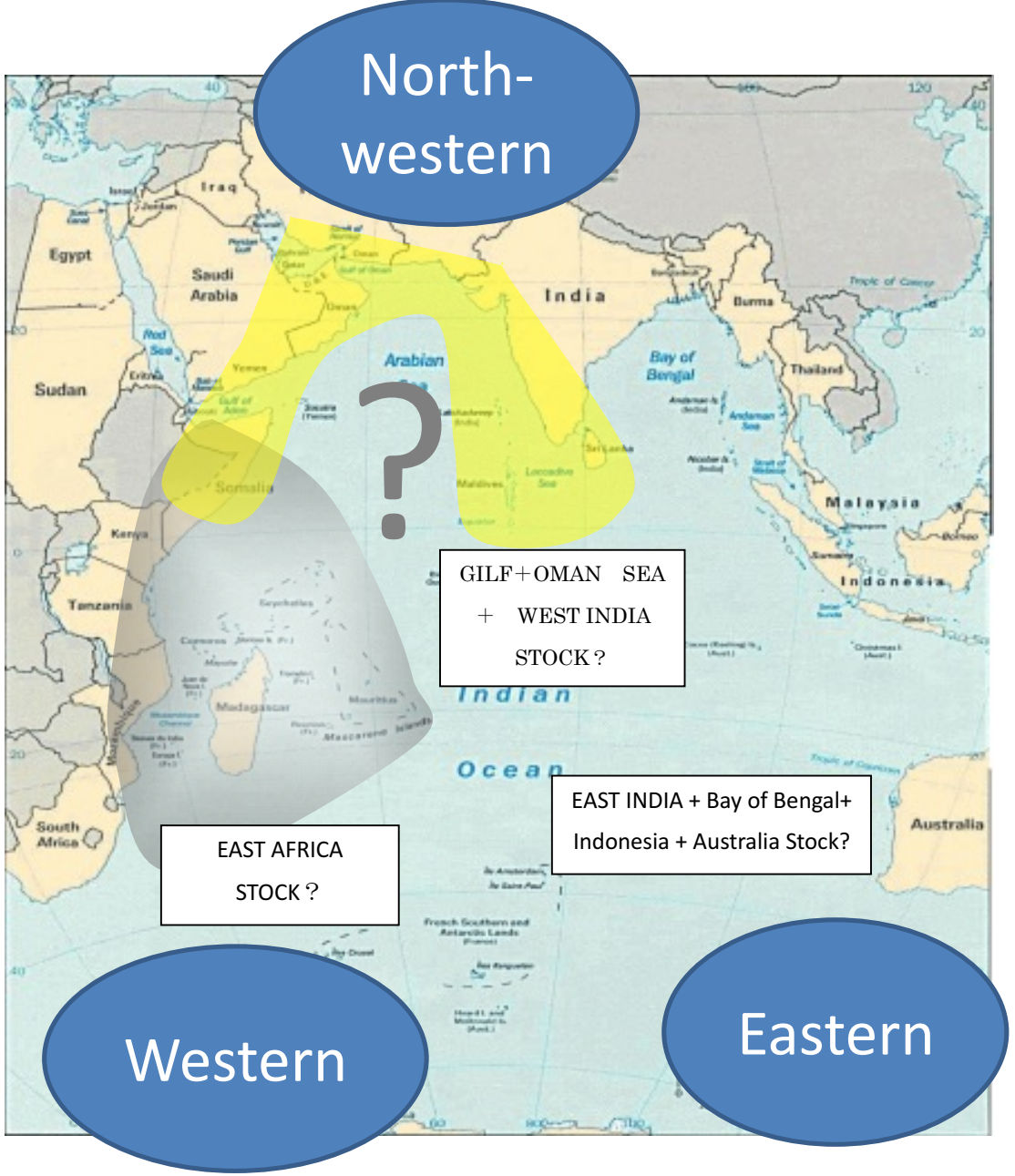
Unknown

But possible
Stock structure

LONGTAIL

Hypothesis
Stock
Structure

Northwestern
Stock
[Oman+Gulf]
Oman, Iran,
Pakistan, UAE,
Yemen, Somalia,
Qatar
[West India]
India, Sri Lanka
and Maldives

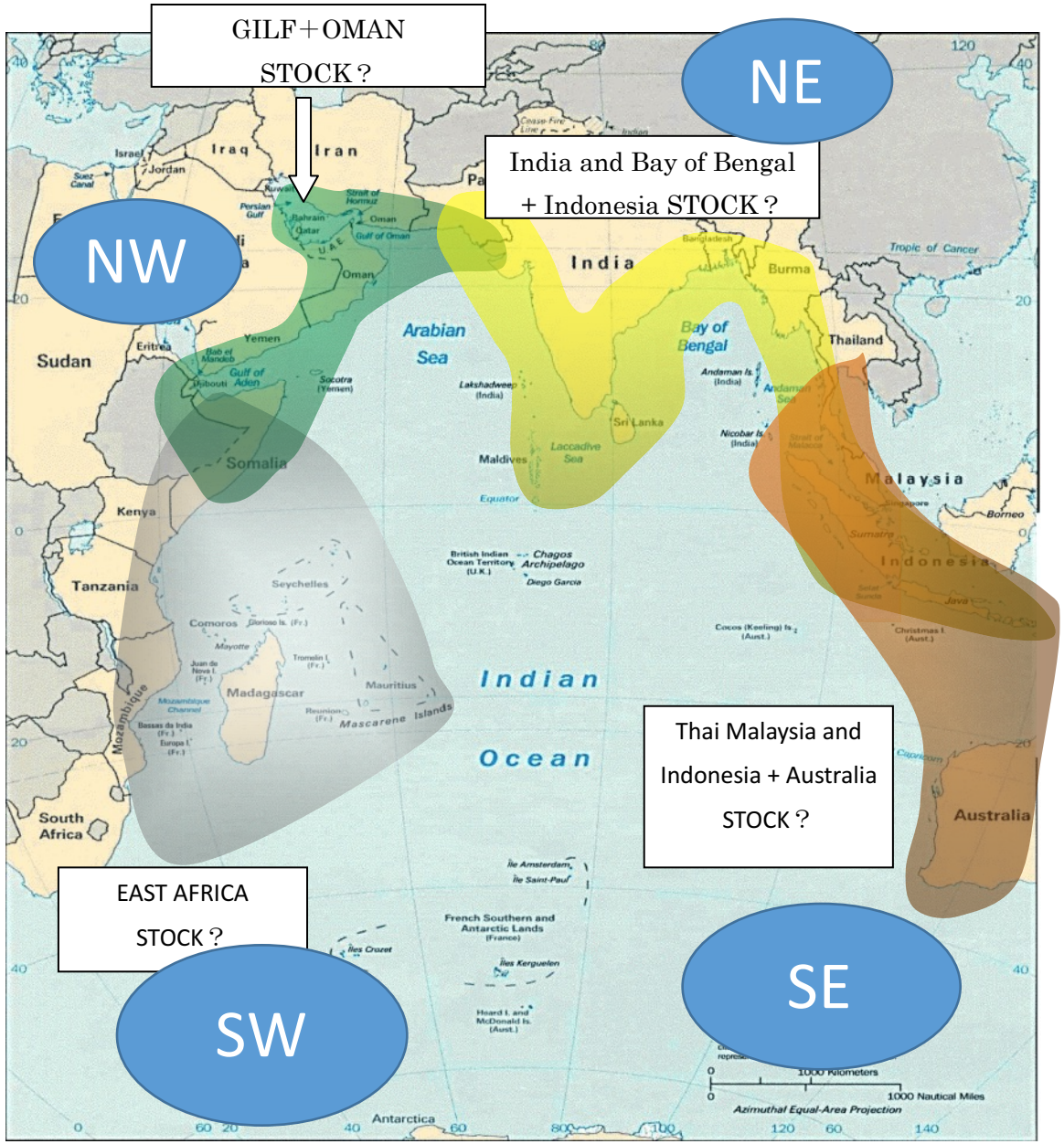


Frigate

Kawakawa

Hypothesis
Stock structure

Northwestern Stock
[Oman+Gulf]
Oman, Iran, Pakistan, UAE, Yemen, Somalia, Qatar



IOTC
KAW

Unknown

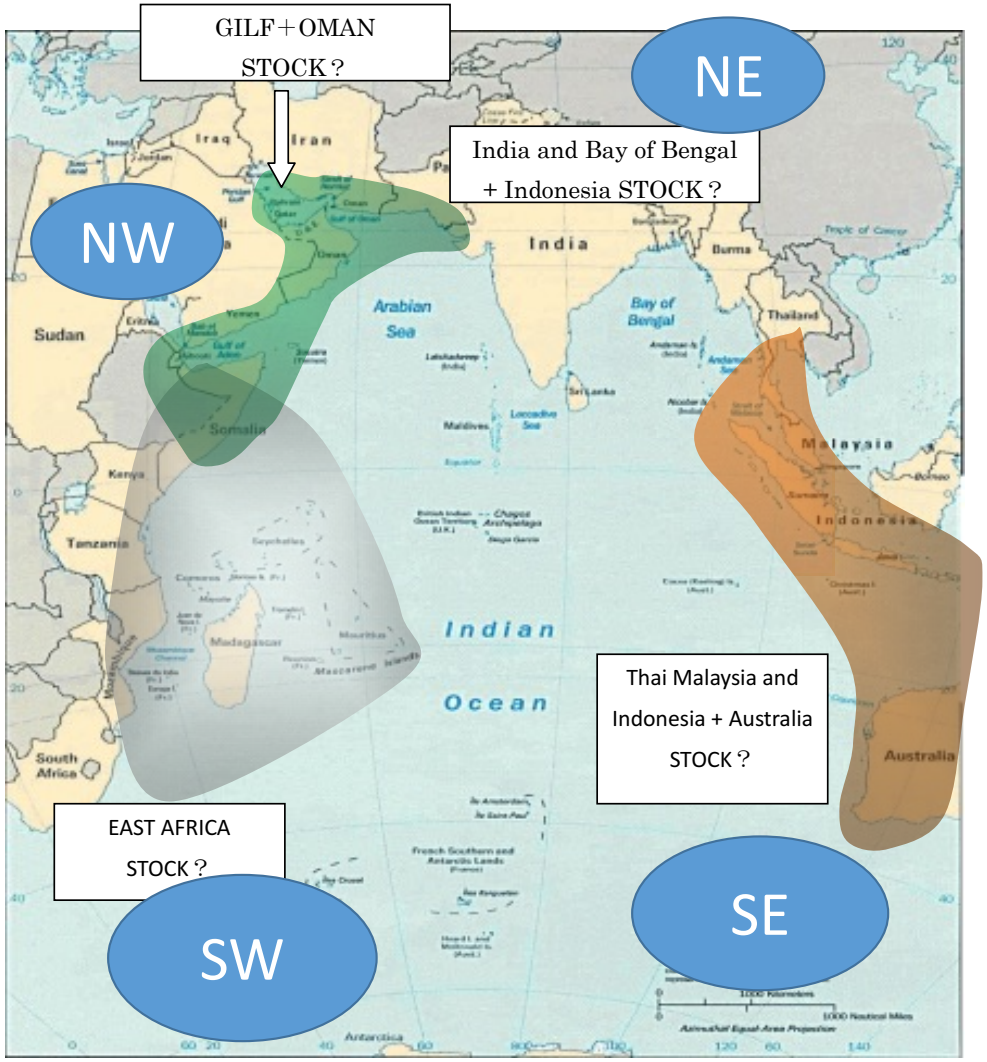
But possible
Stock structure

Frigate

Kawakawa

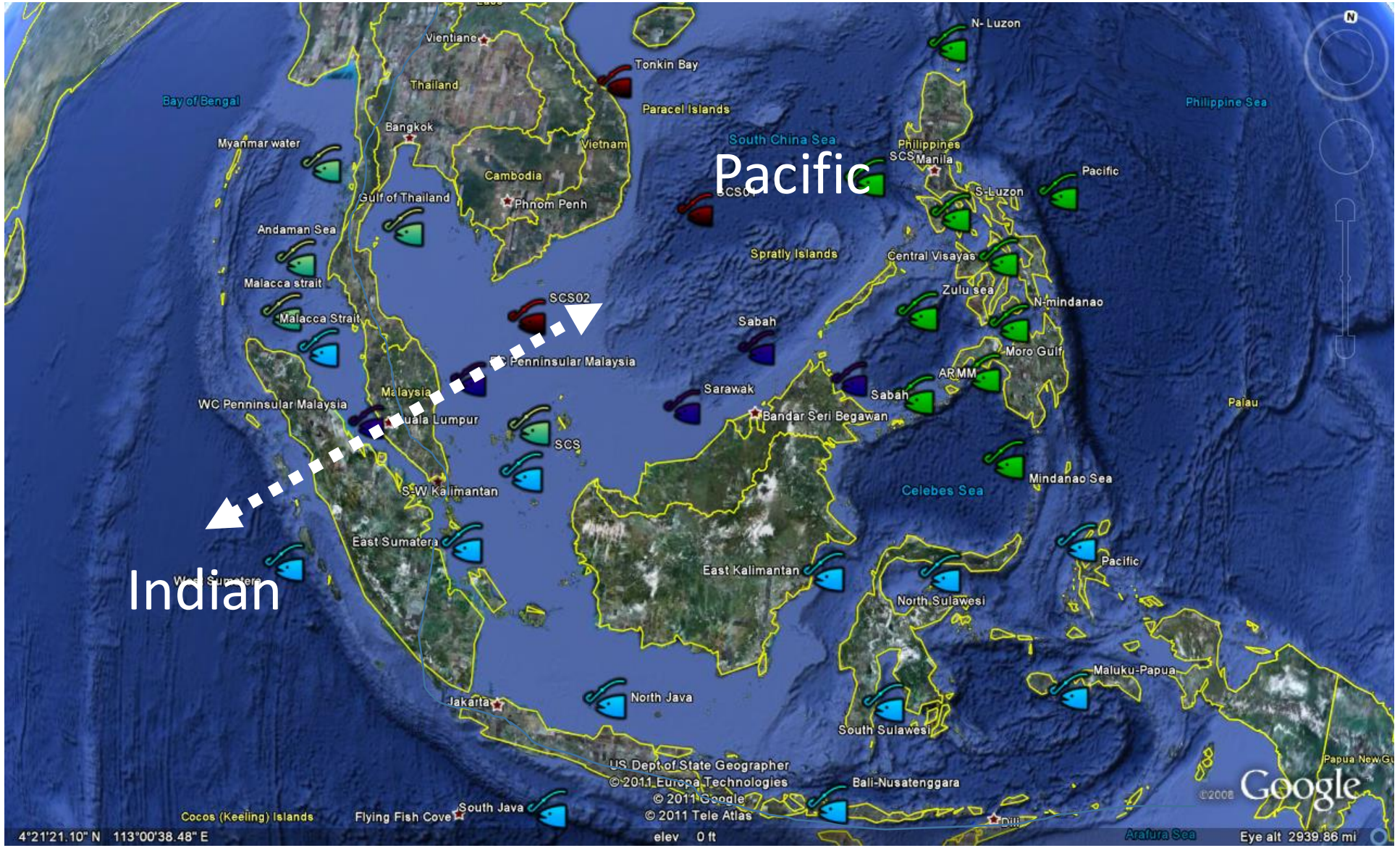
Hypothesis
Stock structure

Northwestern
Stock
[Oman+Gulf]
Oman, Iran,
Pakistan, UAE,
Yemen, Somalia,
Qatar



Stock structure: 2 stocks hypothesis (Indian and Pacific)

assuming very low levels of gene flows (connectivities)..



Stock strcture➔Management unit

We need to do the stock assessments by stock

In our case we don't know the Stock structure

Simple hypothesis (Indian vs. Pacific)

Geographical feature

Gene flow (connectivity)

Data collection by stock (some country covers 2 stocks)

	(a) Pacific stock FAO 71 area	(b) Indian stock FAO 57 area
(1) Brunei		
(2) Cambodia		
(3) Indonesia (2 stocks)		(Malacca)
(4) Malaysia (2 stocks)		(Malacca)
(5) Myanmar		(Bay of Bengal and Andaman)
(6) Philippines		
(7) Thailand (2 stocks)		(Andaman)
(8) Viet Nam		

Template to collect catch data (1)

(1) Annual total catch by species, stock, country, year and gear

Example 1

Stock Pacific
Country Thailand (Gulf of Thailand)
Species longtail tuna

Template (Excel file)

year	Catch by Gear (tons)			
	Troll line	Gillnet	Purse seine	Total
1976	235	678	123	1,036
1977	367	89	239	695
				
2014	110	44	333	487

Note: if species are aggregated, you can indicate as follows (example):
Longtail + Kawakawa (combined), all neritic tuna (combined) etc

Template to collect catch data (2)

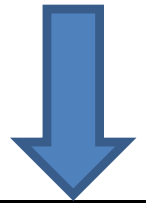
Example 2

Stock Indian

Country Myanmar (Bay of Bengal + Andaman Sea)

Species Longtail tuna + Kawakawa (combined)

Template (Excel file)

year	Catch by Gear (tons)		
	Troll line	Other gears	Total
1990	1,235	234	1,469
1991	367	1,234	1,601
			
2014	110	744	854

Note: if species are aggregated, you can indicate as follows (example):

Longtail + Kawakawa (combined), all neritic tuna (combined) etc

Template to collect CPUE (Catch+Effort) data (1)

By gear, area, year, month, day and boats including 0 catch(important)

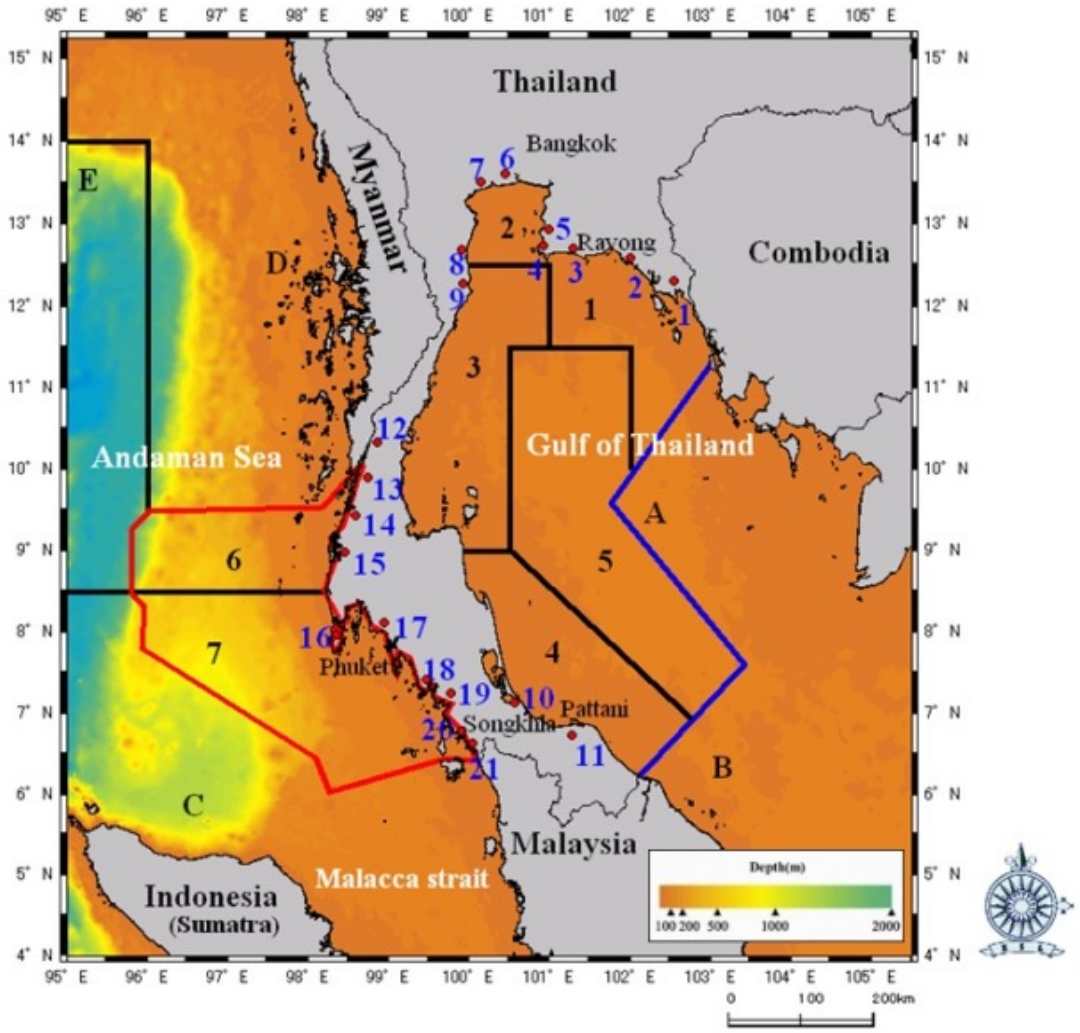
For example.. for GILL and Moro Gulf (**Statistical area 10**) …

				← catch →			← effort →		
Year	**Month**	**Day**	**Boat**	**KAW**	**LOT**	**OTH**	**hour**	**haul**	**day**
1999	12	25	Val-maru	0	0	2,345	23	5	2
2012	1	19	Grace maru	33	0	568	12		

(no data) (blank)

Template to collect CPUE (Catch+Effort) data (2)

Example of fishing Grounds



Fishing ground (statistical area) in Thai waters

(Example of the fishing ground)

If you don't have fishing ground information, then you can use the landing site.

Important : 0 (zero catch)

If you don’t report 0 catch , we will have overestimated nominal CPUE then Standardized CPUE

← catch → ← effort →

Year	Month	Day	Boat	KAW	LOT	OTH	hour	haul	day
1999	12	25	Val-maru	0	0	2,345	23	5	2
2012	1	19	Grace maru	33	0	568	12		

(no data) (blank)

Need the Data processing skill using..

- Excel/ACCES : good for data process for **small + medium size data**

Various functions (sort, filter, pivot tables….) ➔ useful and Powerful

- Users friendly

If you need to process **large data (e.g. million data set),** you need special programs such as Java, C++, Script, SQL-like Query…

Don’t try data process of large data by excel..

Otherwise you will get stiff shoulder
and may get troubles with your partners.

Another important points data process

- Don’t forget data MASAAGE (QC)
  (Former SEAFDEC SG loves this word. Maybe the new SG, too?)

Range Check: year, month, day, latitude, longitude..

GIS check locations

GIS mapping (important)

Visualization of the data ➔ Important

(one of data massage techniques)

After mapping your data

You may find Kawakawa at Mt. Pinatubo??

(barbequed kawakawa)

Barbecued Kawakawa (location errors)

(location errors)
you may find Pacific Yellowfin tuna on Mt Fuji

Data need massage to release stiff shoulder (errors) then you (data) will be happy for accurate analyses such as CPUE standardization + ASPIC

CPUE: Sample size is also important

If you have catch and effort data for many years,
but some might be from a very few sampling or bycatch

Then you can not use such CPUE

We need to investigate sample size
before CPUE standardization

Some example : Philippines CPUE data west Philippines Sea investigation on sample size by gear type PS looks OK

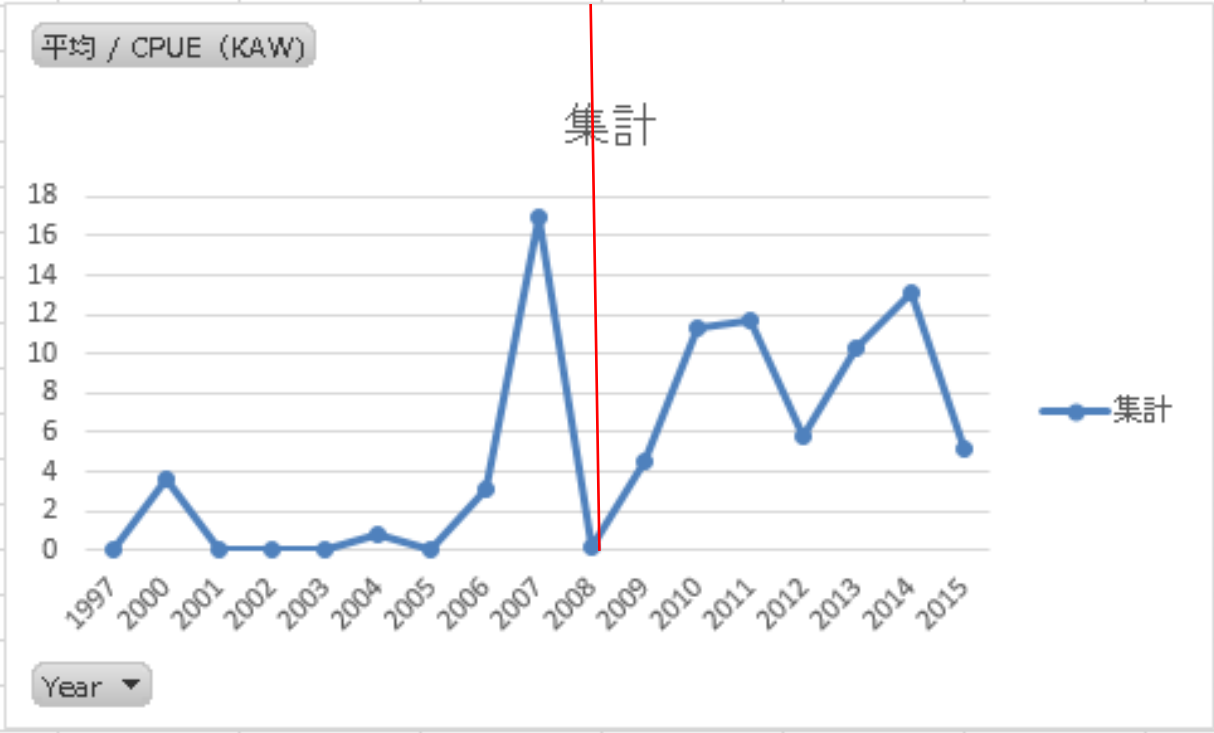
	A	B	C	D	E	F	G	H	I	J	K
4	行ラベル	Bagnet	Bottom gillnet	Danish seine	Drift gillnet	Handline	Jigger	Multiple Handline	Multiple hook and line	Otter trawl	Purse seine
5	1997										23
6	1998	174									
7	1999	250									
8	2000	191				4					88
9	2001	201				10					80
10	2002										104
11	2003	87								117	105
12	2004	218		28						67	118
13	2005	278	15	78	6	7	3		25	26	176
14	2006	31		19							86
15	2007	169		71				6			227
16	2008	66		71						1	309
17	2009	47		83							241
18	2010	42		77							256
19	2011			59							285
20	2012			58							121
21	2013			42							108
22	2014			51		146		168			153
23	2015			44		194	1	301			211

Processed nominal CPUE data
PS data n=2,692 16 years data (1997 and 2001-2015) : LOOKS fine

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	ground_des	landing_center	gear_des	Year	Month	Day	vesselnam e	Effort	UnitEffo	KAW	LOT	OTHER	CPUE (KAW)	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	01	E.C CLAIMS	1	days	0	0	850	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	01	E.C CLAIMS	2	days	0	0	637.5	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	01	E.C CLAIMS	3	days	0	0	818.125	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	01	E.C CLAIMS	4	days	0	0	244.375	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	04	E.C CLAIMS	1	days	0	0	1736.429	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	04	E.C CLAIMS	2	days	0	0	2163.571	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	07	E.C CLAIMS	1	days	0	0	2571.431	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	07	E.C CLAIMS	2	days	0	0	149.9992	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	07	E.C CLAIMS	3	days	0	0	1778.574	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	07	Saint Joseph I	1	days	0	0	931.8073	0	
	West Philippine Sea	Balogo_Matalvis, Masinloc, Zambales	Purse seine	1997	04	07	Saint Joseph I	2	days	0	0	621.2049	0	
	West	Balogo_Matalvis,	Purse				Saint							

After data process.. It was found that….
many 0 or low catches in first 12 years (1997-2008) : **unstable (bycatch)**
last 7 years (2009-2015) **target stable** : by 6 years are too short

行ラベル	平均 / CPUE(KAW)
1997	0
2000	3.672895887
2001	0
2002	0
2003	0
2004	0.796610169
2005	0
2006	3.100775194
2007	16.989721
2008	0.194174757
2009	4.531689864
2010	11.32827522
2011	11.73946142
2012	5.845590491
2013	10.26545181
2014	13.07477878
2015	5.104615368
総計	6.254912907



Another important point
After data collection
Data catalogue is important to know the situation

- From the catalogue we can learn

➔What gear types of CPUE are available

➔Sample size

➔available years

After data collection

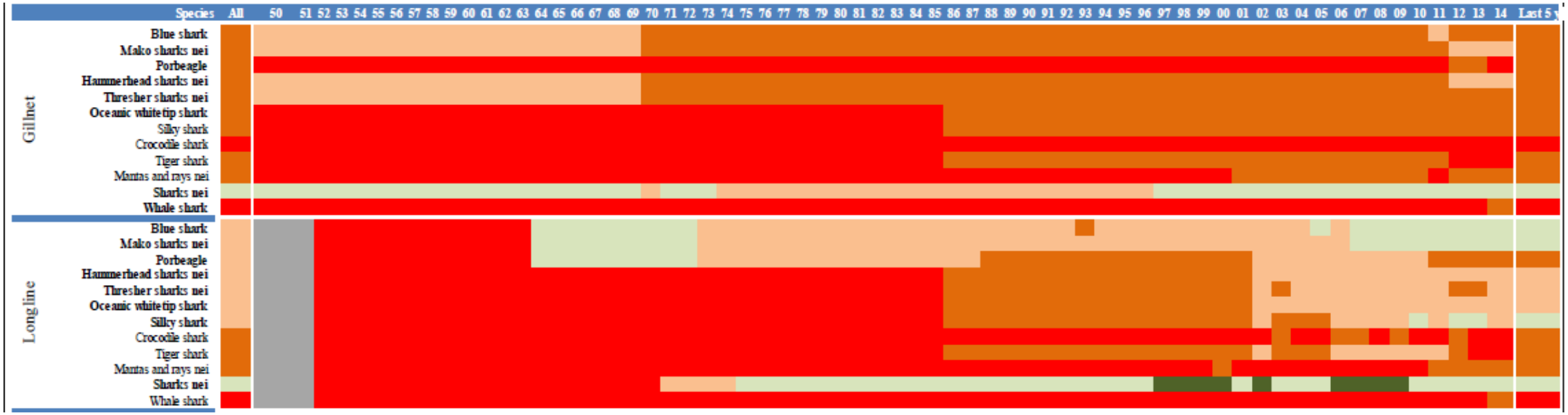
Data catalogue is important to know the situation

Some crazy example Guess what is this ????????

Some crazy example : AVAILABILITY OF CATCH DATA FOR SHARKS BY GEAR(IOTC) BLOOD TABLE

Key

- No catch data available at all
- Catch data available from less than 10% of the fleets for which nominal catches of IOTC species are available
- Catch data available from 10% to 30% of the fleets for which nominal catches of IOTC species are available
- Catch data available from 30% to 75% of the fleets for which nominal catches of IOTC species are available
- Catch data available from more than 75% of the fleets for which nominal catches of IOTC species are available



Data process TWO other Important issues

(1) Data massage or QC (Quality Control) (Massage)

Always Check errors, outliers etc.

(2) Effective managements (Message: where)

to NAME folders, files and Excel sheets

keep original file : don't use you will loose information (and cry)

Copy work file and use the work files

Summary

- Data collection (general and ASPIC)
- Progress of the data collection
- Data processing skills
- Data massage(QC)? And message (folders) management
- Sample size for analyses (example: CPUE standardization）
- Data catalogue (to see the global situation)