Chapter 4
AIS Marine Traffic Database

4.1. Introduction

For the purpose of this study it is primordial to build a robust marine traffic data management system to manage the big amount of data provided by the tracked ships, and to facilitate the access and management of the data itself. Furthermore, looking deep into the subject it is found that building a marine traffic data management system is not original in itself, since many studies and research have been conducted in this matter. The notion of marine traffic data management system is almost limited to building marine traffic databases that can handle the data and ease its access and management.

The previously built marine traffic databases [17], mainly indexed by time, are based on direct observation, Radar/ARPA images, and/or automatic wake extraction form radar signals. Those databases give real image of the traffic and offer unlimited options while reconstructing the traffic scenarios since it includes all environmental parameters that can influence the behavior of the ships. Nevertheless, they present many limitations related to the ship's static data since the main particulars of the ship such as the length and breadth of the ship have to be deducted by regression formulas and the like. However, the other static particulars such as draft, name, IMO number, etc… and the voyage related data remain unknown because there is no way to acquire them from the data itself.

Lately, many commercially built marine databases, mainly indexed by ship's MMSI/IMO numbers [20], emerged in the market offering the user unlimited access to the ships' static data. These databases provide a complete listing of almost the whole world shipping fleet including fishing vessels and most of the service vessels operating around the world. The data is indexed and classified to ease the access and retrieve of the required fields. Using the advance in the database and information technologies, which are already proven safe and secure in other fields of work and research areas such as banking, trade and human resource management, has offered the marine data users unlimited opportunities and access to the world fleet's static data.

Nevertheless, one of the biggest challenges that faced the development and spread of the previously mentioned databases was the connectivity and adaptability of the commercially build databases indexed by ships' particulars, to the marine traffic databases mainly indexed by time. This latter issue was not the only reason that stopped the progress and the wide spread of the databases among the concerned users. The high installment fee and the running cost of the databases also played a major role in limiting the number of users to certain entities. Actually, the high cost of the databases was justified by the fact that collecting ships static data and combining it with the dynamic data was a long process and a peculiar task where many people are involved and many facilities are needed.

Finally by the implementation of the AIS by 2004, the task of collecting and storing the data is simplified. Furthermore, the effort and the cost needed to run and update a database is greatly decreased. The AIS provided the Ships' data both static, dynamic and voyage related in an electronic format. Therefore, the time and resources needed for processing the data and its storage was reduced and so it did the cost.

As it can be noted in the **Fig 4.1** the introduction of the AIS technology has provided a merge between two different blocs that used to be separated by their nature. The AIS, by its VDL technology, has permitted to transfer the ships' related data in an electronic format at a very low cost if not to say free.

To build the intended AIS Marine Traffic Database (hereafter called AIS Database), we needed to acquire the actual ships' AIS data from the tracked ships, and also to provide the acquired data with static data from a commercially built database that covers the designated area. So the built AIS Database will include the ships' actual dynamic and voyage related data provided by ships, and the ships' static data provided by the commercially built database for more reliability.

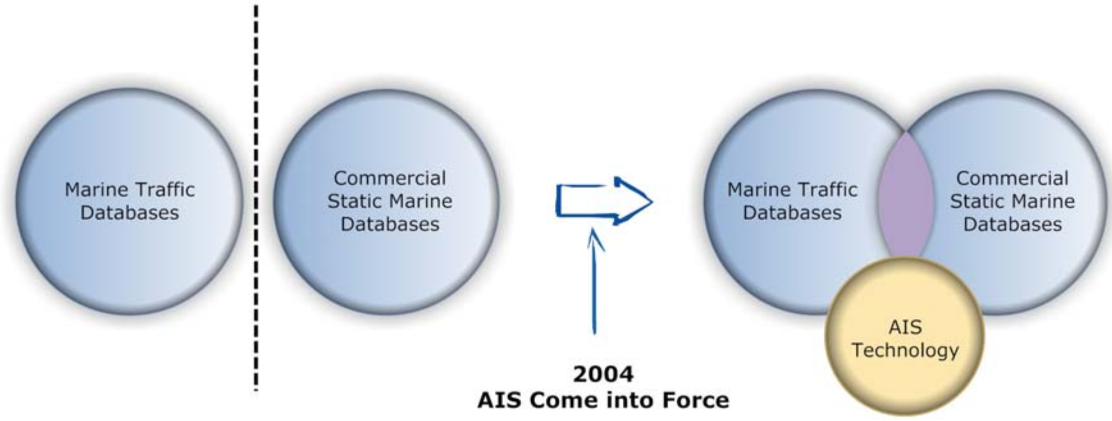


Fig 4.1 Marine Traffic Data Before and After AIS Technology

4.2. AIS Data Acquisition Network

4.2.1. Introduction

In order to carry out long term and long range vessel traffic observations in Tokyo bay[(4)(6)(7)], a fully automated remote Radar/AIS network system had been implemented and fully operated since September 2003. The remote Radar/AIS network is composed of two radar stations (Yokosuka radar station and Kawasaki radar station), AIS receiver installed at Kawasaki radar station, and a remote monitoring station at TUMSAT. Moreover the Radar/AIS network is provided with a website with a visualization tool that can be remotely accessed through an internet web browser, where the maritime traffic in Tokyo Bay can be observed. The architecture of the Radar/AIS network system is shown in **Fig 4.2.**

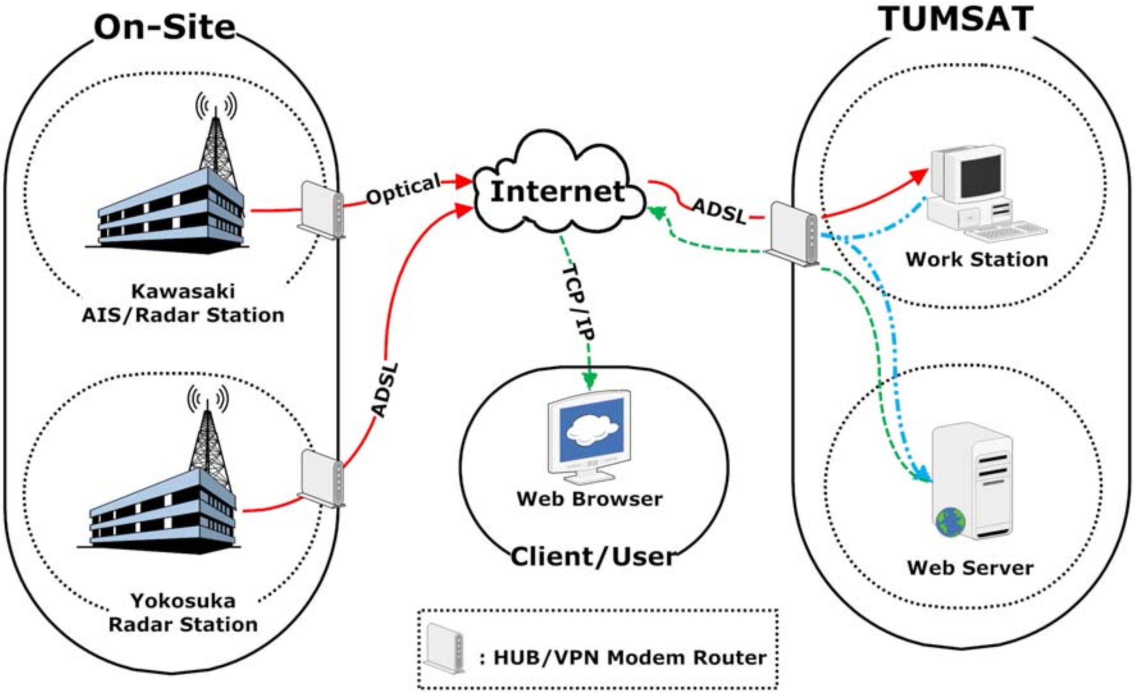


Fig 4.2 TUMSAT Remote Radar/AIS Network System Architecture

The operating concept and the hardware architecture of the network system (Radar and AIS) are shown in **Fig 4.3.** To the left is the block diagram of the radar station and AIS installation (Kawasaki station), and to the right is the block diagram of the communication controller and AIS data recorder.

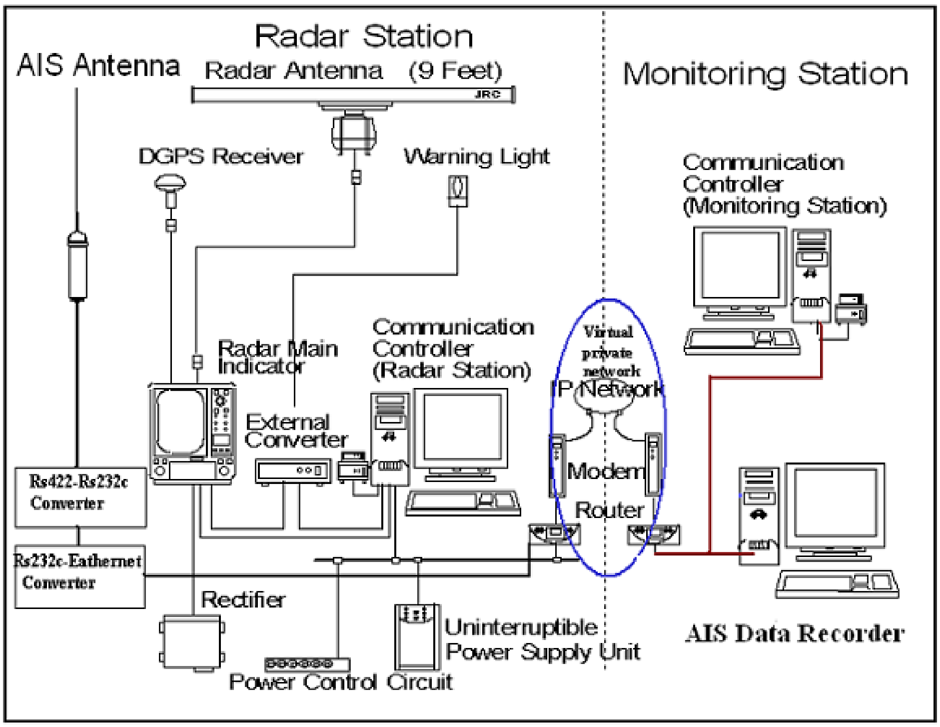


Fig 4.3 TUMSAT Remote Radar/AIS Network System Block Diagram

Both stations (Radar station and Monitoring station) are connected by a Virtual Private Network (VPN) via an IP-based network and a modem router using an ADSL (Asymmetric Digital Subscribe Line) for Yokosuka station and an Optical Communication Line for Kawasaki station. Radar signals and AIS data provided by the radar stations and AIS transponder are converted and transmitted to the monitoring station, where the data can be either displayed on the monitor screen or recorded for future use. Further details and information regarding the line connections between the previously mentioned hardware components are provided in **Fig 4.4.**

The AIS receiver implemented in Kawasaki station covers the whole area of Tokyo Bay, while the radars installed in both Kawasaki and Yokosuka stations covers only a part of it.

4.2.2. AIS Data Transfer Process

The Radar/AIS data, collected from the Remote Radar/AIS Network System for observing Vessel Traffic in Tokyo Bay hereafter AIS network, transfer process to the TUMSAT work station is ensured by the hardware equipment shown in **Fig 4.3**. However, as for the AIS data, **Fig 4.4** details the hardware equipment and line connections used in transferring the data from the AIS transponder to the server at Kawasaki station through several converters. Then it is uploaded through an IP based VPN network to the TUMSAT work station where it is processed and dispatched for other uses.

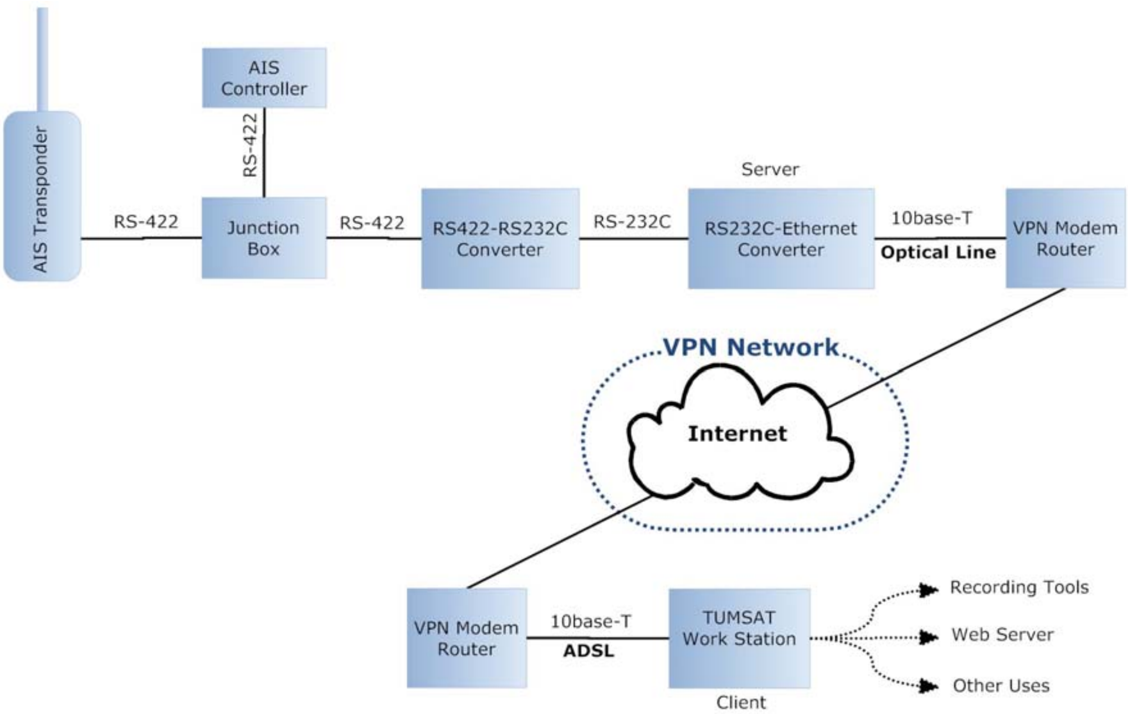


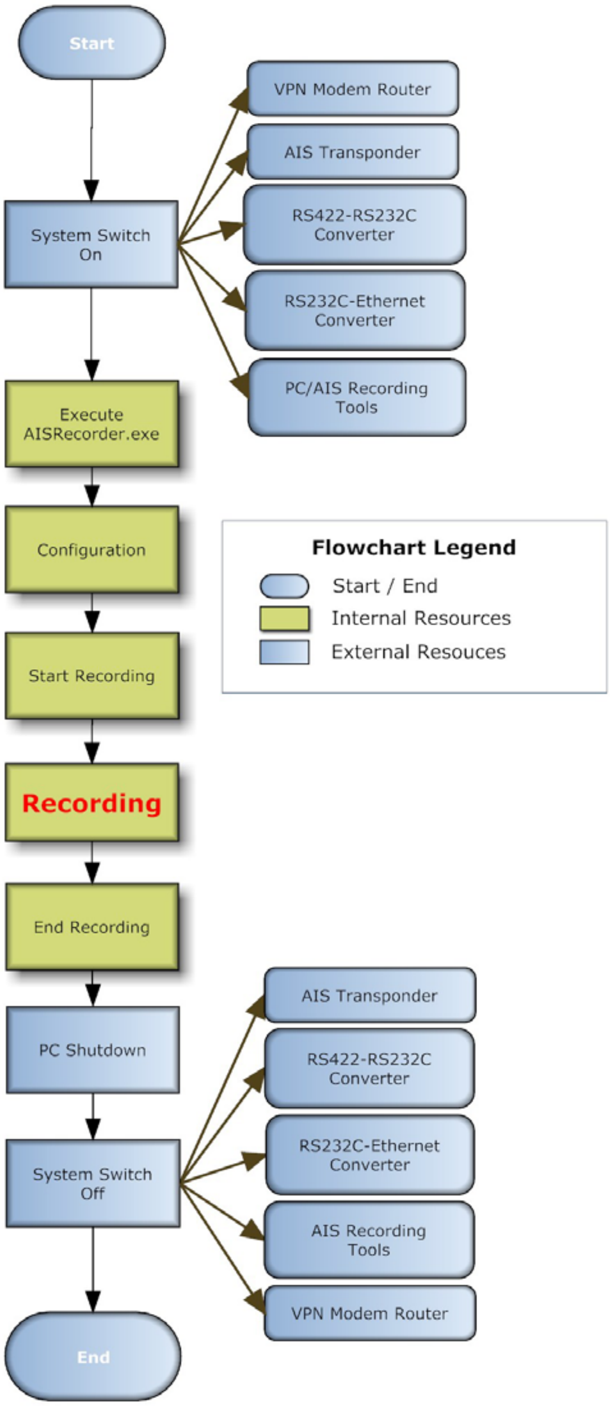
Fig 4.4 AIS data acquisition hardware

4.2.3. AIS Data Recording Tools

The AIS network is provided with a set of tools to remotely record both radar images and AIS data. However, hereafter only the tools regarding the AIS data will be described.

To safely operate the network system and successfully record the desired AIS data the below procedure has to be implemented as shown in **Flowchart 4.1**

- Switch-On all the external systems as by the following order; VPN modem router, AIS transponder, RS422-Rs232C converter, RS232C-Ethernet converter, and the PC AIS recording tools.
- Execute recording program named "AISRecorder.exe"
- Configure the recording parameters
- Start recording
- Finish recording
- PC shutdown
- Switch Off the external systems as by the following order; AIS transponder, RS422-Rs232C converter, RS232C-Ethernet converter, AIS recording tools, and the VPN modem router.



Flowchart 4.1 AIS Data Recording Process

Furthermore, the recording tools are provided with a Selective and Non-Selective recording options as described below.

4.2.3.1. Non-Selective AIS Data Recording

All the acquired data is recorded without any exception or condition as shown in **Flowchart 4.2**

After starting the recording tools, the system listens to intercept data. When new data is intercepted the system sends it immediately to the buffer memory. Then another process is started automatically as follows:

- The data in the buffer is analyzed for integrity
- The VDM (VHF Data Link Message) legend is checked
- If the check is successful, the Message ID (MSG ID) is recorded. Otherwise, another loop is generated
- Only after the two previous steps are successful, the data is exported/saved to file.

The record procedure is executed at every loop.

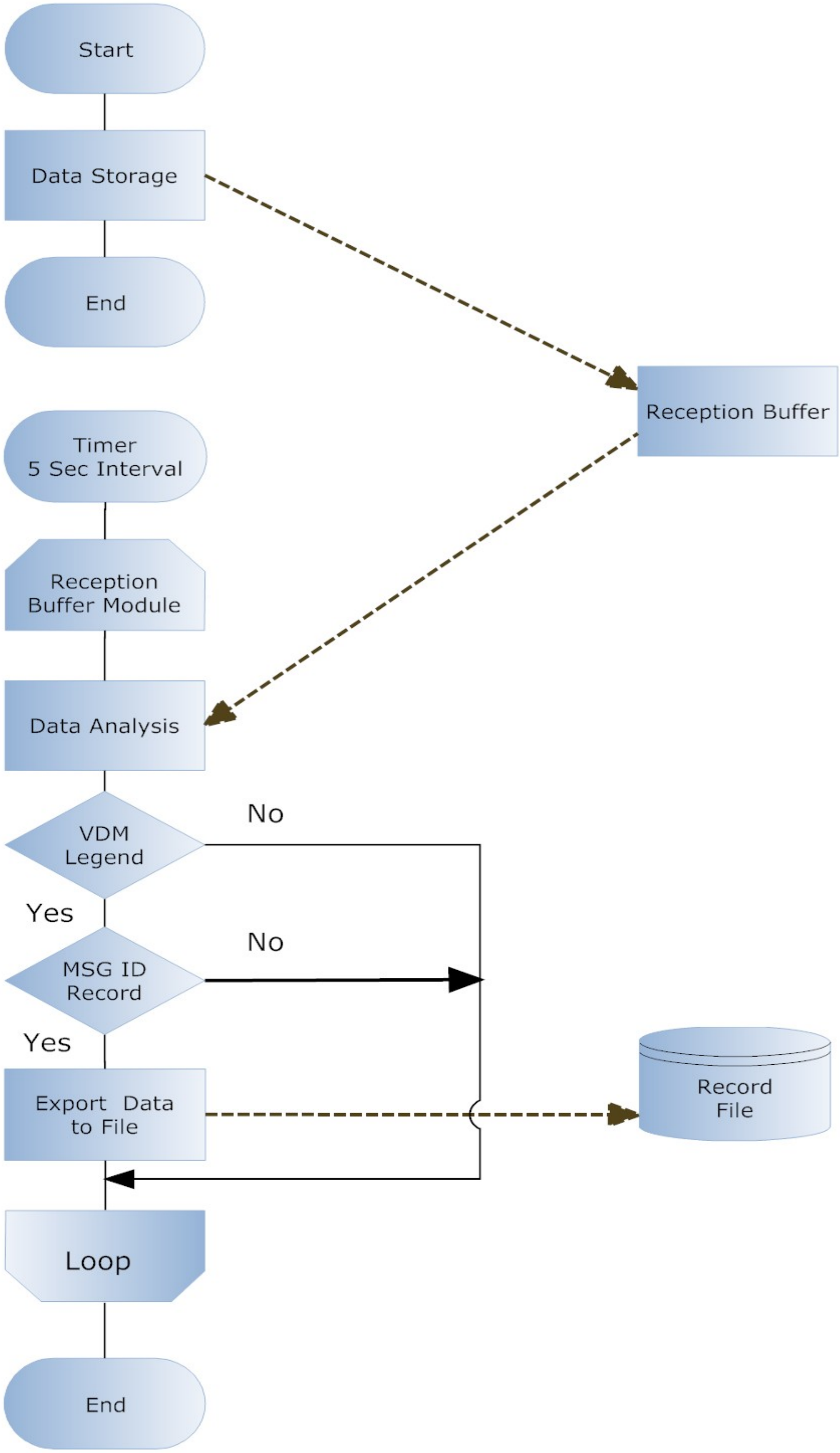
4.2.3.2. Selective AIS Data Recording

The acquired data are processed and subjected to selective conditions where only desired targets are recorded as shown in **Flowchart 4.3**

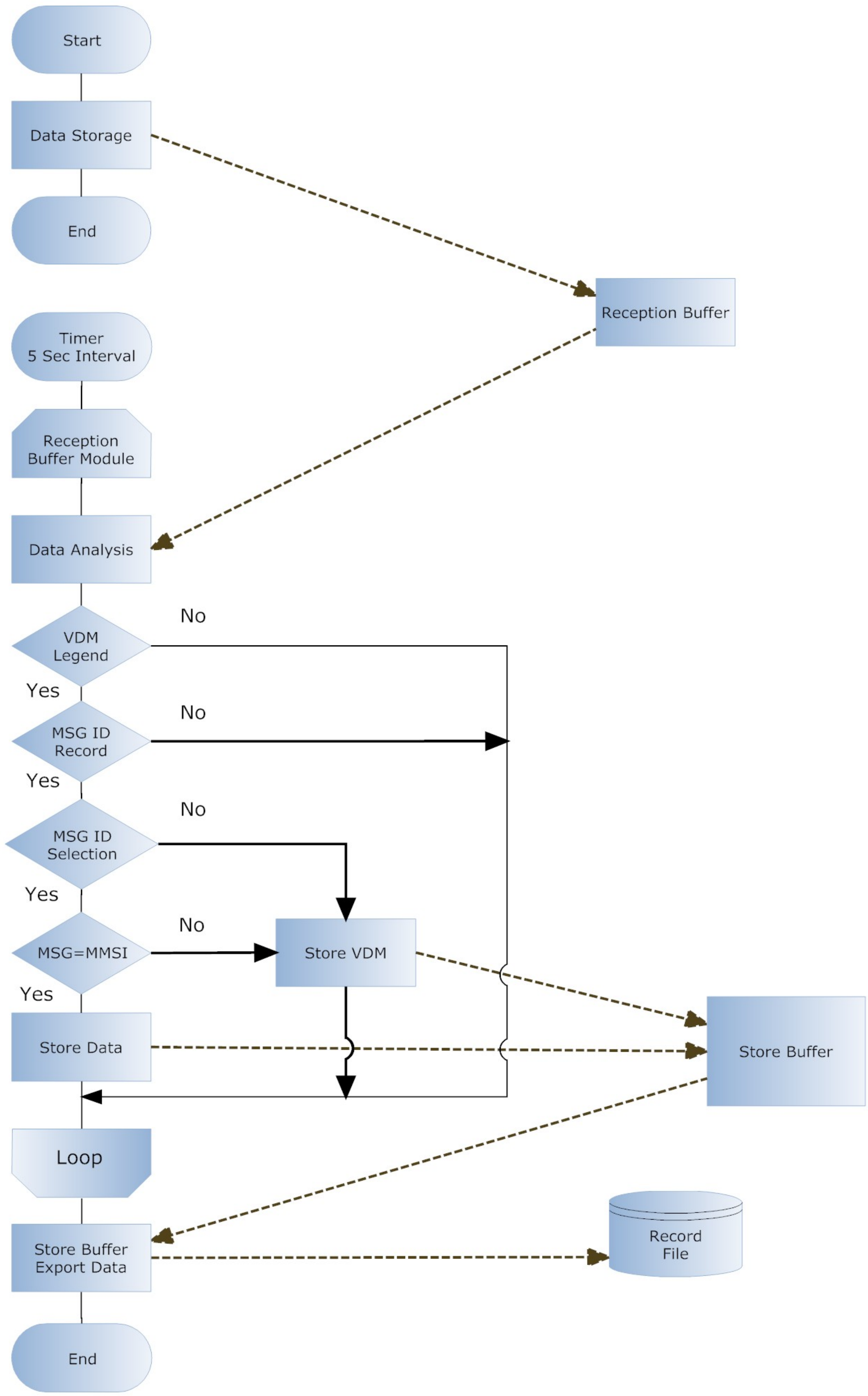
After starting the recording tools, the system listens to intercept data. When new data is intercepted the system sends it immediately to the buffer memory. Then another process is started automatically as follows:

- The data in the buffer is analyzed for integrity
- The VDM (VHF Data Link Message) legend is checked
- If the check is successful, the Message ID (MSG ID) is recorded. Otherwise, another loop is generated
- If the MSG ID is recorded successfully, the MSG ID is compared to the MMSI of the desired target(s).
- If the ID matches the MMSI, the data is sent to the buffer memory
- Otherwise, the VDM legend is stored for later use and another loop is generated
- After all the MSG IDs are checked, the stored buffer memory is emptied and exported/saved to file.

The record procedure is executed after the end of the whole loop.



Flowchart 4.2 Non-Selective AIS Data Recording



Flowchart 4.3 Selective AIS Data Recording

4.2.4. AIS Data Acquisition Process

To build the AIS database, Raw AIS data is collected from the previously introduced AIS network system. The Raw AIS data acquisition process architecture is shown in **Fig 4.5.** The AIS data acquisition process follows the below described steps:

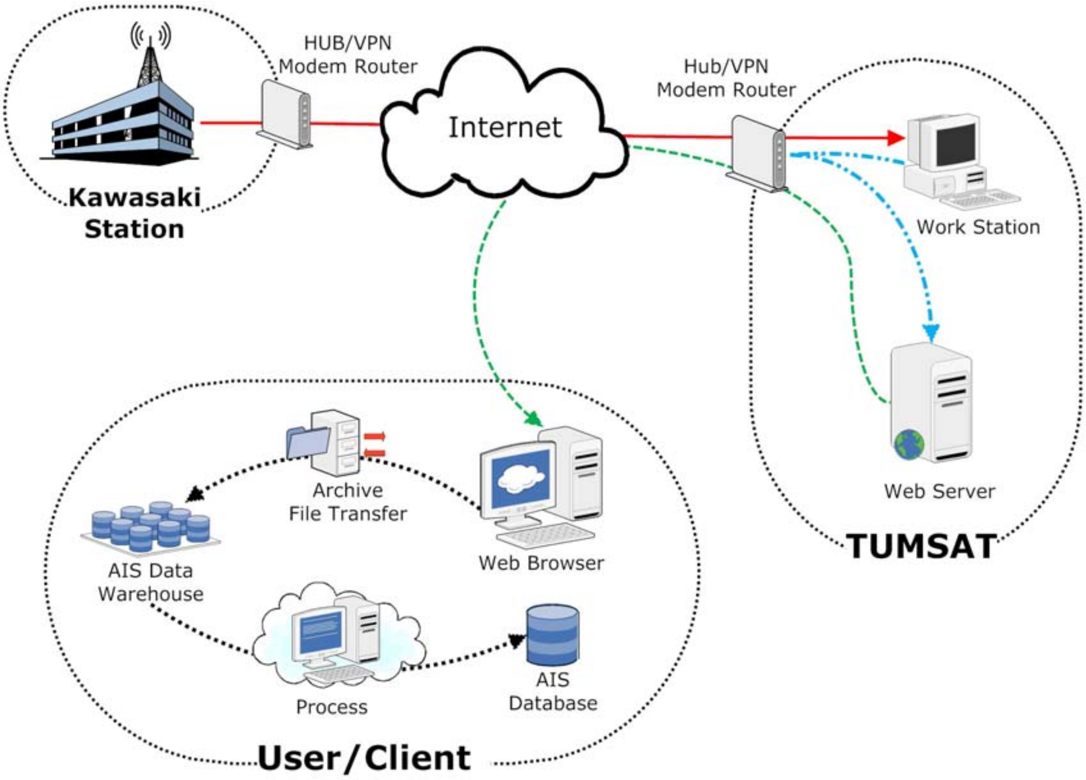


Fig 4.5 AIS Data Acquisition Process

4.2.4.1 Step1: Raw AIS Data Acquisition Program Design

A User PC based program is made to acquire the data from the TUMSAT web server through internet. The DLL that guaranteed the exclusive secured access to the TUMSAT web server is granted by the person in charge of the AIS network. The GUI of the AIS Data Acquisition Program, hereafter program, is shown in **Fig 4.6.**

The program is provided with

- 'Recording Frequency' input space that allows the user to fix the recording frequency in seconds (the default value is set to 5 seconds).
- 'Start Saving Data' button to start the recording process
- 'Stop Saving Data' button to end the recording process.
- 'Exit' button to end the program.

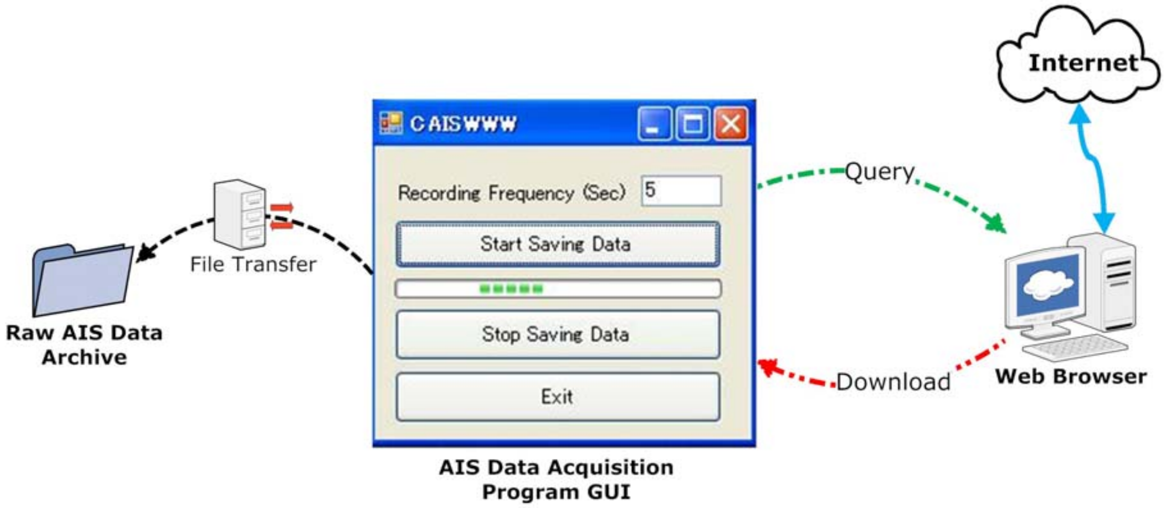


Fig 4.6 Raw AIS Data Acquisition Process

4.2.4.2 Step2: Raw AIS Data Recording

The AIS data recording process is as follows:

- The program is deployed on the User computer to the Archive location (where the data is going to be saved).
- The program is executed to start the AIS data recording.
- A query is sent from the program to the TUMSAT web server through an internet web browser.
- If AIS data are intercepted in the TUMSAT web server, another process is started to download the data and save it to the chosen location. Otherwise the program will return empty query.
- The downloaded data is saved to the AIS data archive.
- When the recording is finished the program is first stopped to save the last downloaded data, then ended by clicking on 'Exit'

4.2.5. Raw AIS Data Archive

The Raw AIS data is saved using CSV storage engine that stores data in text files using comma-separated values format (CSV files), according to the following architecture as shown in **Fig 4.7**

- For a fixed time, the Raw AIS data is gathered and saved to a CSV file as follows "ais-yyyymmddhhnnss.CSV".

  (yyyy=year, mm=month, dd=day, hh=hours, nn=minutes, ss=second)
- The CSV files of fixed day are saved to a folder named "ais-yyyymmdd"
- The month folders are saved to a folder named "Month Year".

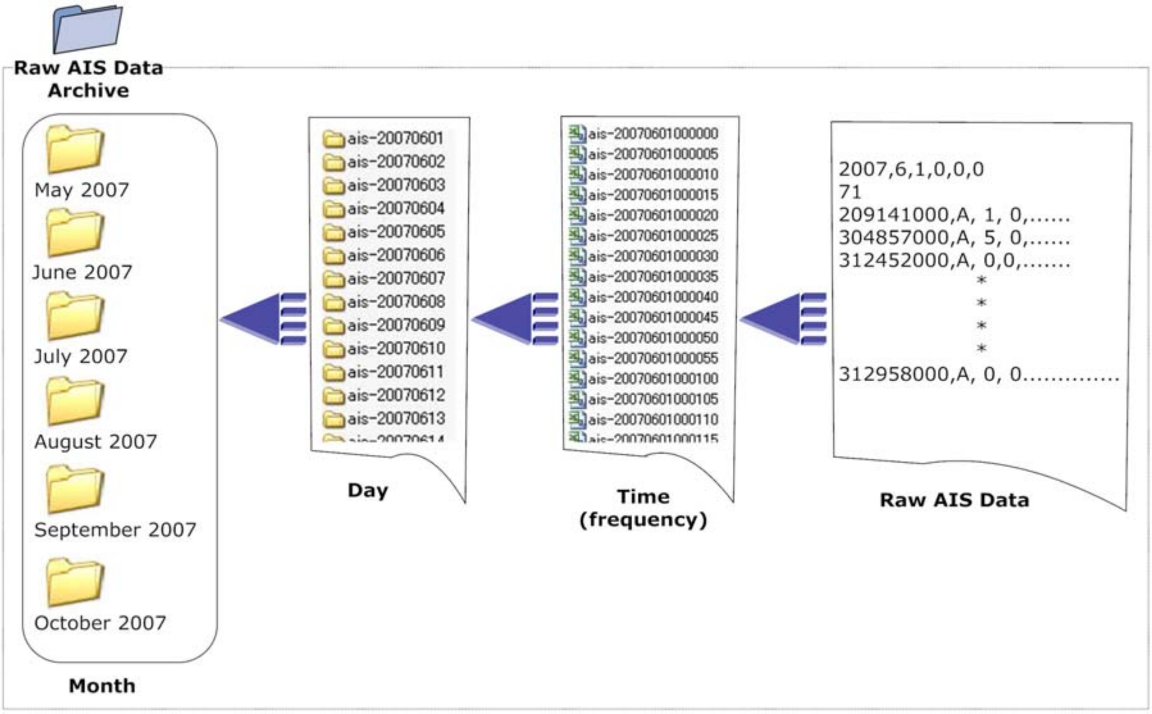


Fig 4.7 Raw AIS Data Archive Architecture

4.2.6. Raw AIS Data Format

The format of the collect Raw AIS Data format is as shown in **Fig 4.8.**

2007,9,1,0,0,0

82

211703000,A, 5, 0, .1, 200.1, 1, 139.782611666667, 35.60652, 327, 39, 0, 48161540, 211, 449, 135, 21, 6, 3, 15, 4, 30, 19, 212,"TOKYO", 0

……. (Next ship)

Legend:

yyyy,mm,dd,hh,nn,ss

Nbr of Ships

MMSI, AIS class, Navsat, ROT , SOG, COG, Acceleration, Longitude, Latitude, Heading, Time Stamp, Version of AIS, IMO Nbr, AIS code, Antenna position1, Antenna position2, Antenna position3, Antenna position4, Deviation, ETA, Draft, Destination, DTE.

…… (Next ship)

Fig 4.8 Raw AIS Data Format

4.2.7. Raw AIS Data Storage Size

The storage size of the Raw AIS data is as shown in **Table 4.1**

Table 4.1 Storage Size of Raw AIS Data

	May	June	July	August	September	October
Nbr of Days	16	27	25	23	30	7
Size	3.2 GB	5 GB	3.4 GB	4 GB	4.9 GB	1.4 GB
					Total Size	**22 GB**

4.3. AIS Database

4.3.1. Introduction

The CSV storage engine has many limitations that can be summarized on the fact that it does not support indexing and does not support table partitioning, which is very important to improve the speed of managing and querying data.

However as listed in Appendix VI there are many database storage engines that can be used. To choose the adequate storage engine a comparison is made between the main storage engines provided by MySQL [(5)] based on the requirements that the AIS Database should support (the requirements are ordered by priority):

- Transaction type database
- Foreign keys
- Backup / point in time recovery
- Geospatial data type (for future research)

After comparing all the database storage engines supported by MySQL, InnoDB is found to be the most suitable engine for our research. Therefore all the databases and tables built hereafter use InnoDB storage engine.

To ease the design and creation of the AIS Database, WAMPSERVER (WAMP) is used. WAMP is a form of mini-server that can run on almost any Windows Operating System, WAMP includes Apache 2, PHP 5, and MySQL is preinstalled (phpMyAdmin and SQLitemanager are installed to manage the databases). WAMP offers a graphic interface where the database can be created, populated, and managed easily.

Find more information in Appendix VI and Appendix VII**.**

The Architecture of the AIS database is as shown in **Fig 4.9**

As it can be seen in **Fig 4.9** the data from the AIS Data Warehouse together with the one of Static Data Database are processed, then provided with MMSI Index to build the AIS database.

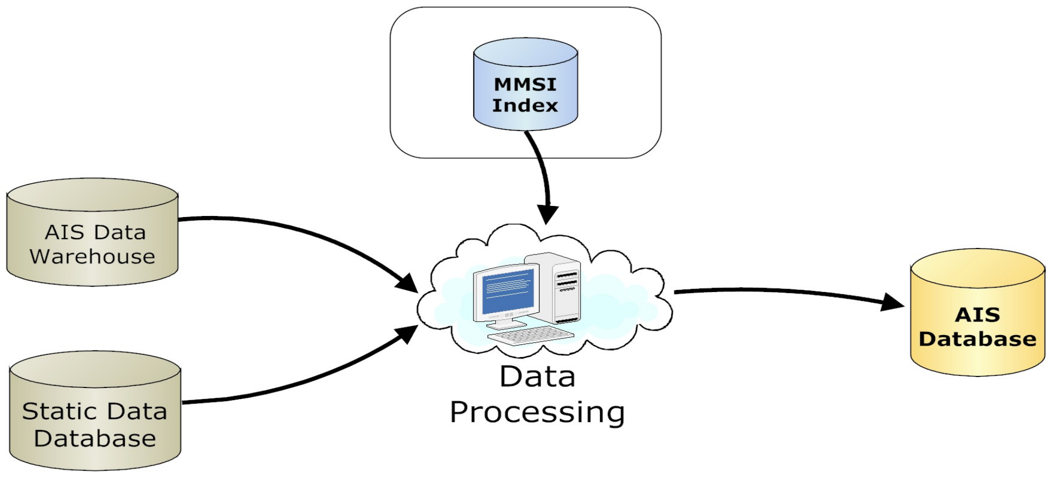


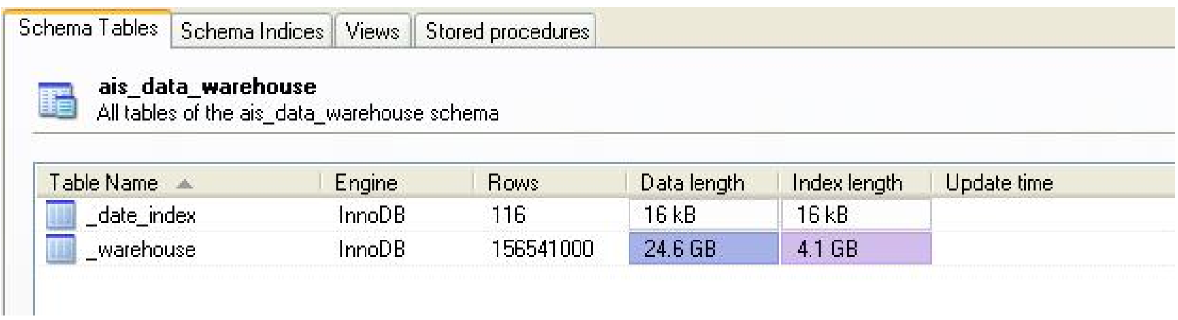
Fig 4.9 AIS Database Architecture

4.3.1. AIS Data Warehouse

WAMP server is also used to build the AIS Data Warehouse (hereafter Warehouse). The Raw AIS Data Archive is inputted to a database designed to store the Raw AIS data without any processing just for the purpose of easing data management.

4.3.2.1. AIS Data Warehouse Structure and Format

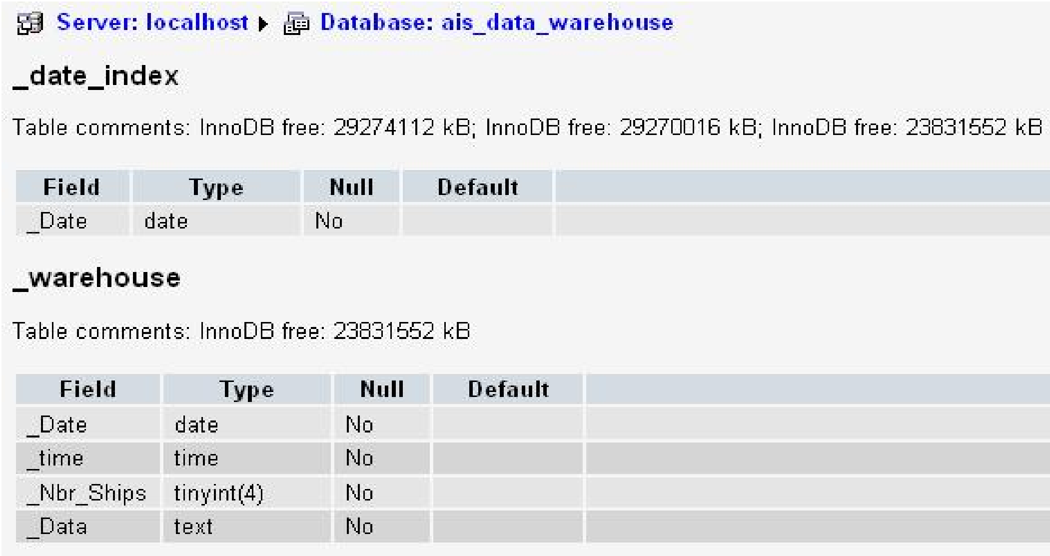
The Warehouse is composed of two tables (Date Index and Warehouse). The Warehouse tables Structures, and legends are shown in **Fig 4.10** and **Fig 4.11**


Schema Tables | Schema Indices | Views | Stored procedures

ais_data_warehouse
All tables of the ais_data_warehouse schema

Table Name	Engine	Rows	Data length	Index length	Update time
_date_index	InnoDB	116	16 kB	16 kB	
_warehouse	InnoDB	156541000	24.6 GB	4.1 GB	

Fig 4.10 AIS Data Warehouse Tables



Server: localhost ▸ Database: ais_data_warehouse

_date_index

Table comments: InnoDB free: 29274112 kB; InnoDB free: 29270016 kB; InnoDB free: 23831552 kB

Field	Type	Null	Default
_Date	date	No	

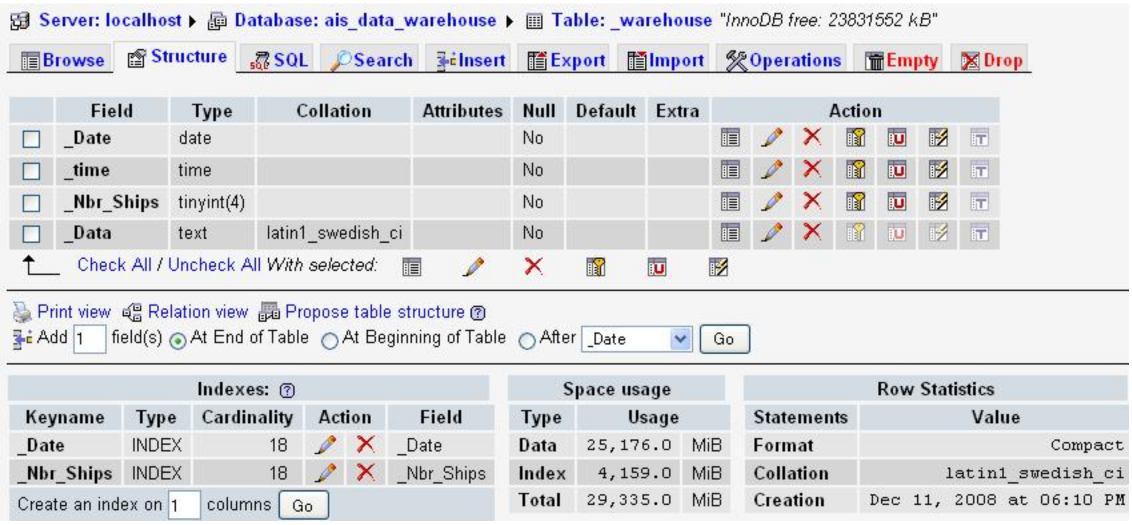
_warehouse

Table comments: InnoDB free: 23831552 kB

Field	Type	Null	Default
_Date	date	No	
_time	time	No	
_Nbr_Ships	tinyint(4)	No	
_Data	text	No	

Fig 4.11 AIS Data Warehouse Data Dictionary

4.3.2.1.1 Warehouse table contains four fields of data as follows: Date, Time, Number of Ships, and AIS data as shown in **Fig 4.12 and Table4.2**



Server: localhost ▸ Database: ais_data_warehouse ▸ Table: _warehouse "InnoDB free: 23831552 kB"

Browse | Structure | SQL | Search | Insert | Export | Import | Operations | Empty | Drop

Field	Type	Collation	Attributes	Null	Default	Extra	Action
_Date	date			No			
_time	time			No			
_Nbr_Ships	tinyint(4)			No			
_Data	text	latin1_swedish_ci		No			

Check All / Uncheck All *With selected:*

Print view | Relation view | Propose table structure

Add 1 field(s) ◉ At End of Table ○ At Beginning of Table ○ After _Date Go

Indexes:

Keyname	Type	Cardinality	Action	Field
_Date	INDEX	18		_Date
_Nbr_Ships	INDEX	18		_Nbr_Ships

Create an index on 1 columns Go

Space usage

Type	Usage	
Data	25,176.0	MiB
Index	4,159.0	MiB
Total	29,335.0	MiB

Row Statistics

Statements	Value
Format	Compact
Collation	latin1_swedish_ci
Creation	Dec 11, 2008 at 06:10 PM

Fig 4.12 Warehouse Table Structure

Table 4.2 Warehouse Table Legend

_Date	**_time**	**_Nbr_Ships**	**_Data**
yyyy-mm-dd	hh:nn:ss	Nbr of Ships	MMSI, AIS class, Navsat, ROT , SOG, COG, Acceleration, Longitude, Latitude, Heading, Time Stamp, Version of AIS, IMO Nbr, AIS code, Antenna position1, Antenna position2, Antenna position3, Antenna position4, Deviation, ETA, Draft, Destination, DTE

2007-09-01	00:00:00	82	211703000,A,5,0,0.1,200.1,1,139.78261166667, 35.60652, 327, 39, 0, 48161540, 211, 449, 135, 21, 6, 3, 15, 4, 30, 19, 212,"TOKYO", 0

4.3.2.1.2 Date Index table contains only one data field (date: yyyy-mm-nn) used as index to ease querying and management. As shown in **Fig 4.13.**

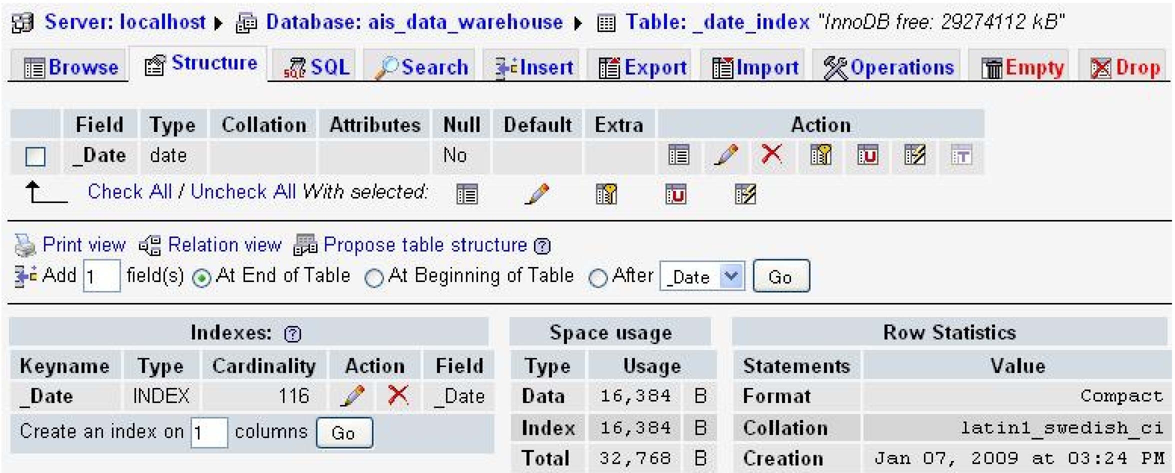


Fig 4.13 Date Index Table Structure

4.3.2. Static AIS Database

4.3.3.1. Overview

By examining the data of the Warehouse, it is found that the ships' data contain many errors as far as the static data is concerned. The errors may be caused either by wrong data input, or by corrupted transmission protocols and/or decoding software. Thus to overcome this problem and provide accurate ships' static data, the static data provided by the World Shipping Encyclopaedia (Encyclopaedia) of the Lloyd's Register Fairplay .Ltd is used.

The Encyclopaedia database provides expanded fleet criteria (all ships of 100 GT and above) as well as companies and ports data to give a complete marine information system. The database includes information regarding:

Ships: With details of all vessels of 100 GT and above including new buildings, the Encyclopaedia is a powerful database giving access to more than 300 fields of information. However the database is indexed by IMO number, and not by MMSI number.

Companies: Covering every aspect of the industry, the database allows quick identifying of the company, contact details, personnel, and industry sector, products and services provided of over 124,000 maritime companies

Ports and terminals: Full details including port descriptions, restrictions, location, port plans, and atlas and service providers are given on over 9,600 ports and terminals worldwide.

Distance tables: With 3,000 locations and route selections, the distance tables provide the fastest and most accurate method to calculate the distances and voyage times between major ports and terminals

User notes: Users can add notes to each ship, company and port record, which are searchable and automatically updated with each new release of the Encyclopaedia

Reports and exports: A comprehensive range of reporting and export functions is provided. You can export up to 2,500 records, up to 12 fields at a time.

The following **Fig 4.14** and **Fig 4.15** provide snapshots of the Encyclopaedia with regards to the ships category.

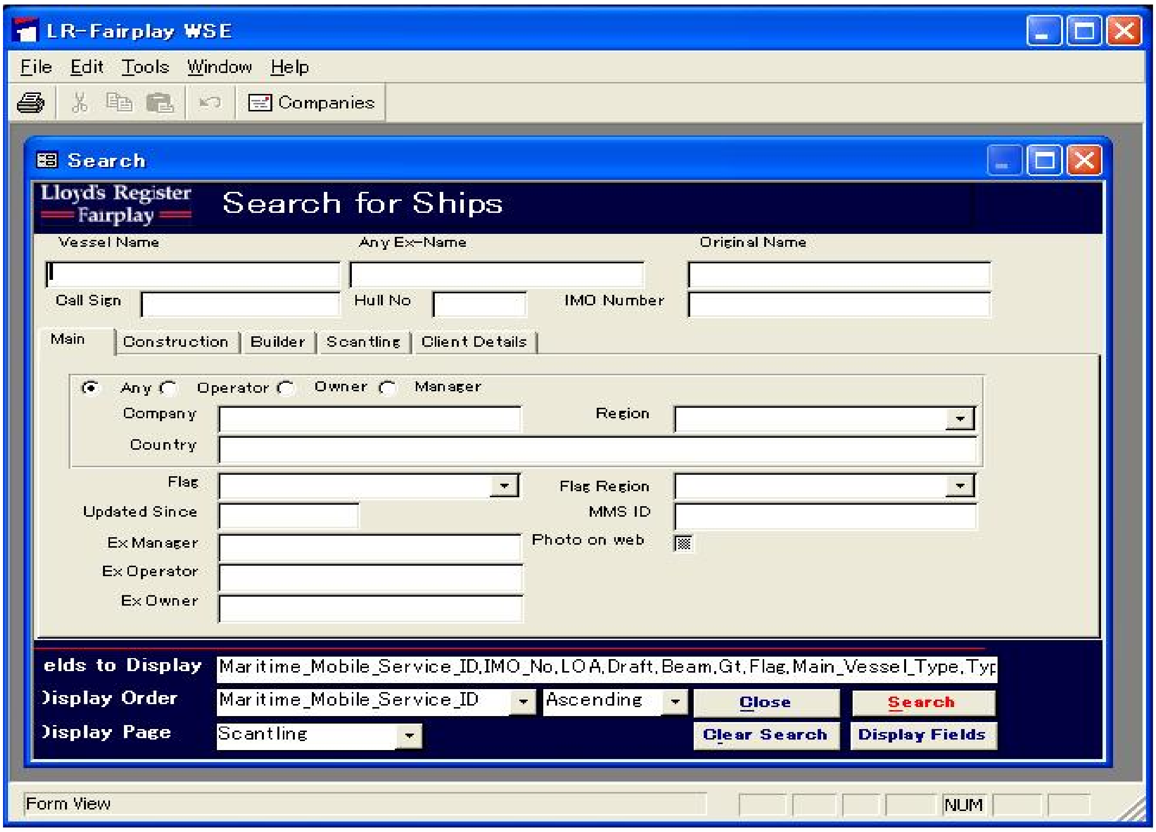
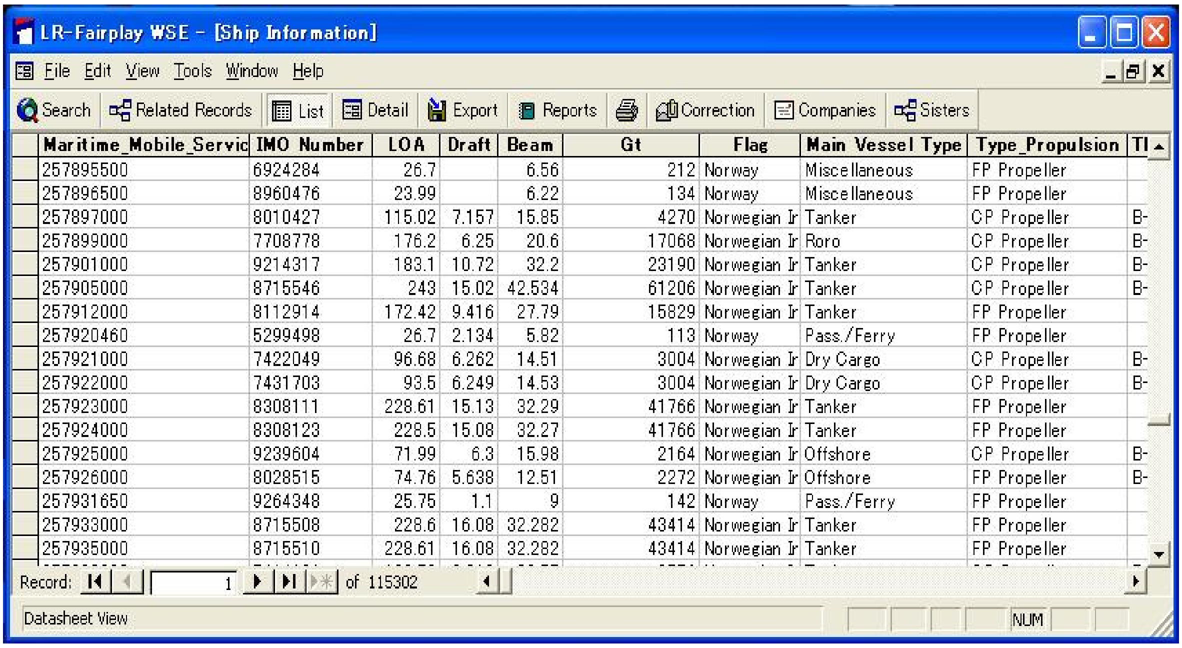


Fig 4.14 Search Window of the World Shipping Encyclopaedia



Maritime_Mobile_Servic	IMO Number	LOA	Draft	Beam	Gt	Flag	Main Vessel Type	Type_Propulsion	Tl
257895500	6924284	26.7		6.56	212	Norway	Miscellaneous	FP Propeller	
257896500	8960476	23.99		6.22	134	Norway	Miscellaneous	FP Propeller	
257897000	8010427	115.02	7.157	15.85	4270	Norwegian Ir	Tanker	CP Propeller	B-
257899000	7708778	176.2	6.25	20.6	17068	Norwegian Ir	Roro	CP Propeller	B-
257901000	9214317	183.1	10.72	32.2	23190	Norwegian Ir	Tanker	CP Propeller	B-
257905000	8715546	243	15.02	42.534	61206	Norwegian Ir	Tanker	CP Propeller	B-
257912000	8112914	172.42	9.416	27.79	15829	Norwegian Ir	Tanker	FP Propeller	
257920460	5299498	26.7	2.134	5.82	113	Norway	Pass./Ferry	FP Propeller	
257921000	7422049	96.68	6.262	14.51	3004	Norwegian Ir	Dry Cargo	CP Propeller	B-
257922000	7431703	93.5	6.249	14.53	3004	Norwegian Ir	Dry Cargo	CP Propeller	B-
257923000	8308111	228.61	15.13	32.29	41766	Norwegian Ir	Tanker	FP Propeller	
257924000	8308123	228.5	15.08	32.27	41766	Norwegian Ir	Tanker	FP Propeller	
257925000	9239604	71.99	6.3	15.98	2164	Norwegian Ir	Offshore	CP Propeller	B-
257926000	8028515	74.76	5.638	12.51	2272	Norwegian Ir	Offshore	FP Propeller	B-
257931650	9264348	25.75	1.1	9	142	Norway	Pass./Ferry	FP Propeller	
257933000	8715508	228.6	16.08	32.282	43414	Norwegian Ir	Tanker	FP Propeller	
257935000	8715510	228.61	16.08	32.282	43414	Norwegian Ir	Tanker	FP Propeller	

Fig 4.15 Queried Results of the World Shipping Encyclopaedia

4.3.3.2. Static AIS Database Structure and Format

However, for the purpose of this study only 12 fields that represent a vital importance to assess the maritime traffic in Tokyo Bay are used. The names of the ships are not extracted to not give any room for commercial interpretation or any other consequential conflict arising therefrom.

The final extracted data are shown in **Fig 4.16**

IMO Nbr, MMSI Nbr, LOA, Draft, Beam, Gross Tonnage, Flag, Vessel Type, Propulsion Type, Thruster Type, Hull Type, Classification Society.

Fig 4.16 Static AIS Database Data Format

After examining the final extracted data, it is found that many ships (around 40%) don't have an MMSI number included. So it is decided to not include MMSI number in the Static AIS Database to prevent any conflict that may arise while manipulating the data transactions inside the database.

The Static AIS Database structure and format are shown **Fig 4.17** and **Fig 4.18**

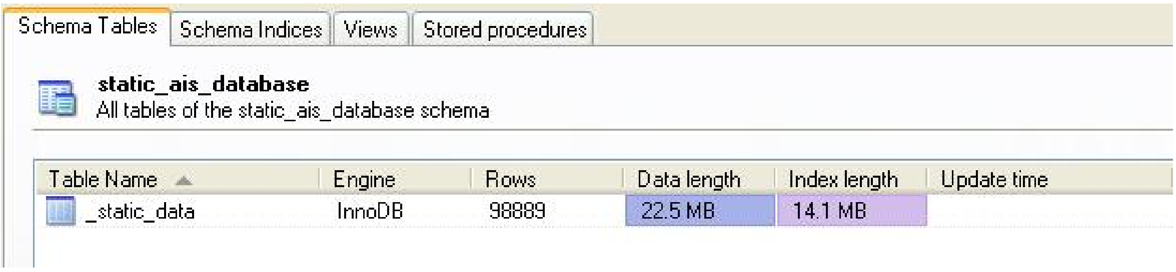


Fig 4.17 Static AIS Database Table

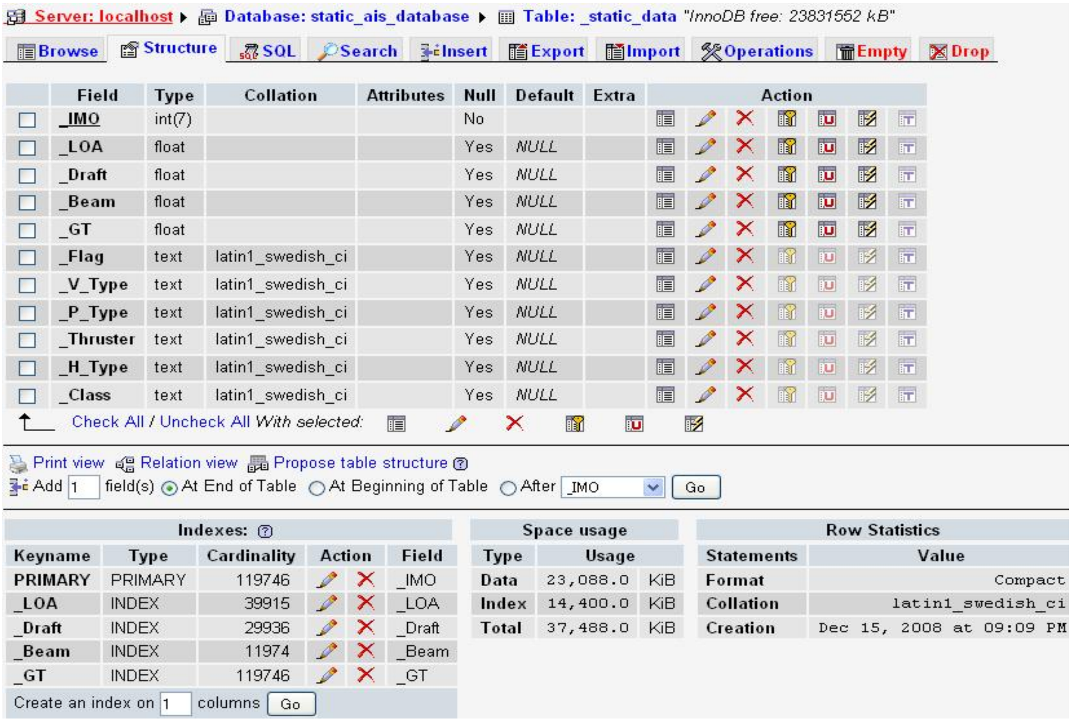


Fig 4.18 Static AIS Database Structure

4.3.3. MMSI Index

A Maritime Mobile Service Identity (MMSI) is a series of nine digits that are assigned to ship stations, ship earth stations, coast stations, coast earth stations, and group calls. The initial digit of an MMSI categorizes the category of the station, where the MID (Maritime Identification Digit) reveals the region and the country to where the ship is affiliated. Refer to (Appendix VIII) for more information.

The MMSI Index database is built based on the MMSI categories and MID identification properties. The EER (Extended Entity Relationship) Diagram of the database is shown in **Fig 4.19**.

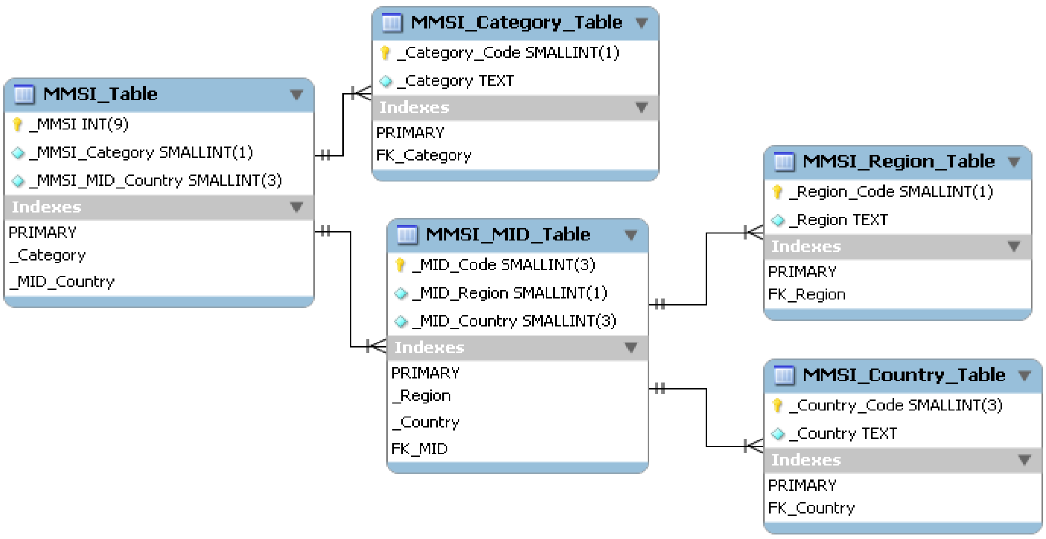


Fig 4.19 MMSI Index Database EER Diagram

4.3.4.1. MMSI Index Structure and Format

The MMSI Index database is composed of five tables as follows: MMSI List table, MMSI Category table, MID table, MID Region table, and MID Country table. The MMSI Index database tables' structures are shown hereafter in **Fig 4.20 and Fig 4.21.**

Schema Tables | Schema Indices | Views | Stored procedures

mmsi_index
All tables of the mmsi_index schema

Table Name	Engine	Rows	Data length	Index length	Update time
_category_code	InnoDB	10	16 kB	0 B	
_country_code	InnoDB	272	16 kB	0 B	
_mid_code	InnoDB	272	16 kB	32 kB	
_mmsi_list	InnoDB	48792	1.5 MB	3 MB	
_region_code	InnoDB	6	16 kB	0 B	

Fig 4.20 MMSI Index Database Tables

Server: localhost ▸ Database: mmsi_index

_category_code

Table comments: InnoDB free: 23831552 kB; (`_Cat_code`) REFER `mmsi_index/_mmsi_list`(`_MMSI_Cat

Field	Type	Null	Default	Comments
_Cat_code	smallint(1)	No		
_Category	text	No		

_country_code

Table comments: InnoDB free: 23831552 kB; (`_Country_code`) REFER `mmsi_index/_mid_code`(`_MID_C

Field	Type	Null	Default	Comments
_Country_code	smallint(3)	No		
_Country	text	No		

_mid_code

Table comments: InnoDB free: 23831552 kB; (`_MID_Code`) REFER `mmsi_index/_mmsi_list`(`_MMSI_MID

Field	Type	Null	Default	Comments
_MID_Code	smallint(3)	No		
_MID_Region	smallint(1)	No		
_MID_Country	smallint(3)	No		

_mmsi_list

Table comments: InnoDB free: 23831552 kB

Field	Type	Null	Default	Comments
_MMSI	int(9)	No		
_MMSI_Category	smallint(1)	No		
_MMSI_MID	smallint(3)	No		

_region_code

Table comments: InnoDB free: 23831552 kB; (`_Region_Code`) REFER `mmsi_index/_mid_code`(`_MID_Re

Field	Type	Null	Default	Comments
_Region_Code	smallint(1)	No		
_Region	text	No		

Fig 4.21 MMSI Index Database Data Dictionary

4.3.4.1.1. MMSI List table contains three data fields as follow: MMSI, MMSI category (MMSI initial digit), and MID as shown in **Fig 4.22**.

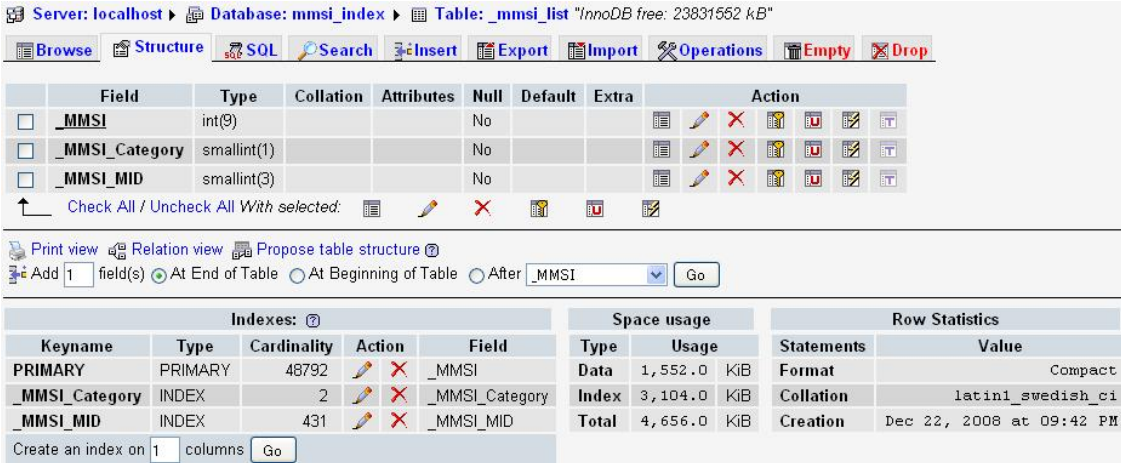


Fig 4.22 MMSI List Table Structure

4.3.4.1.2. MMSI Category table is composed of a dictionary where every Category code is assigned a value in text format as shown in **Fig .4.23**.

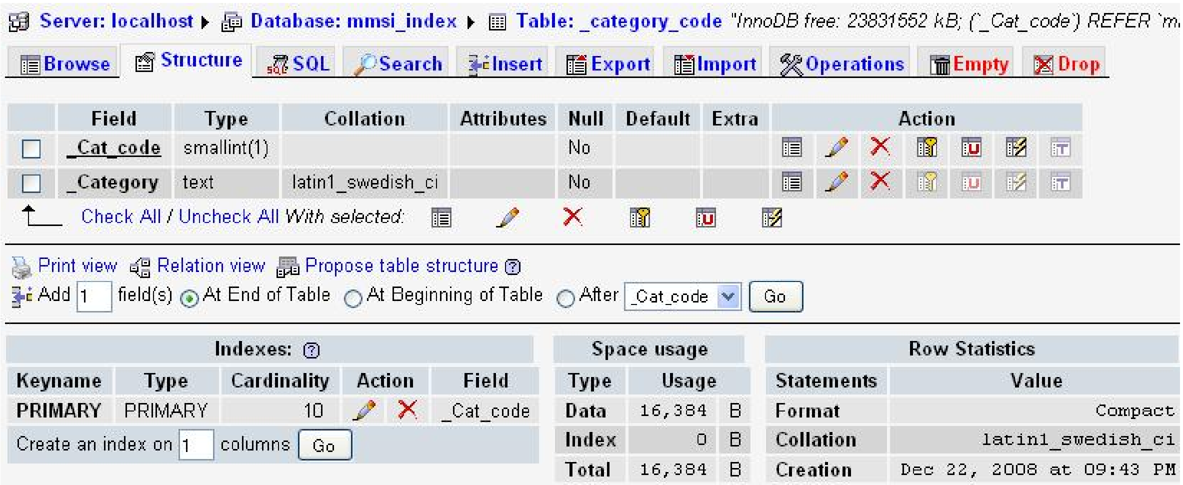


Fig 4.23 MMSI Category Table Structure

4.3.4.1.4. MID table contains three data fields as follow: MID, Region MID (MID initial digit), and Country MID as shown in **Fig 4.24**.

4.3.4.1.5. MID Region table is composed of a dictionary where every Region code is assigned a value in text format as shown in **Fig 4.25**.

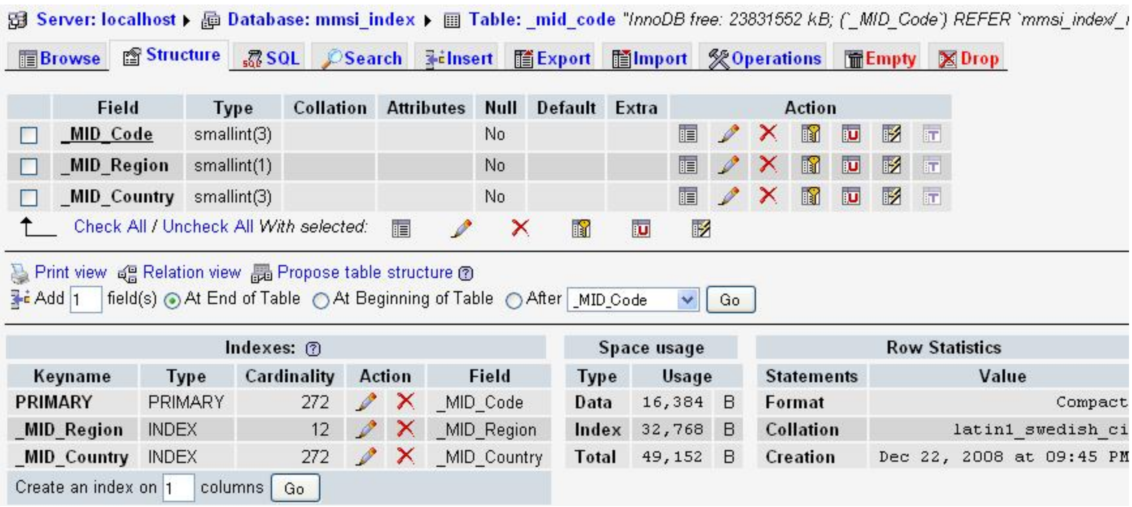


Fig 4.24 MID Table Structure

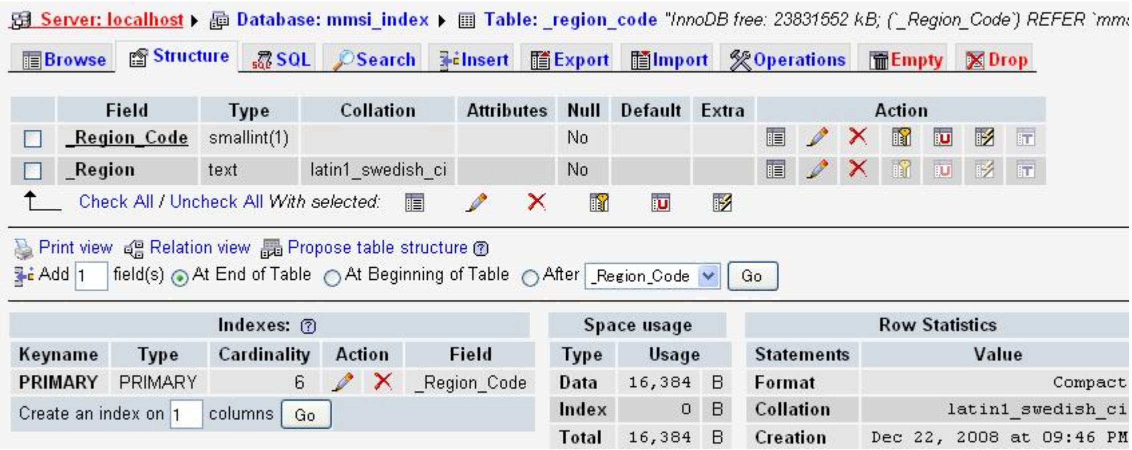


Fig 4.25 Region Table Structure

4.3.4.1.5. MID Country table is composed of a dictionary where every Country code is assigned a value in text format as shown in **Fig 4.26**.

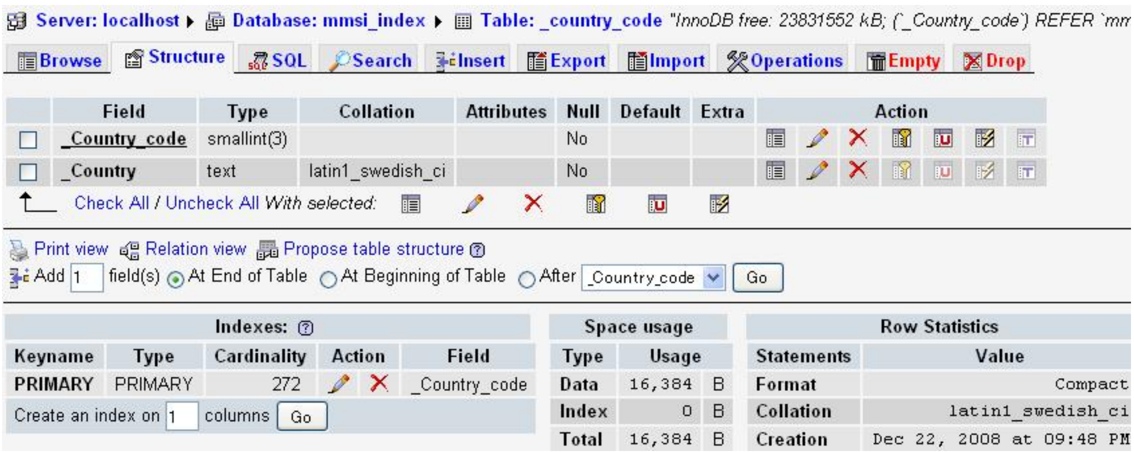


Fig 4.26 Country Table Structure

An example of an MMSI index (**MMSI=** 431xxxxxx) and its legend is shown hereafter in **Fig 3.27.**

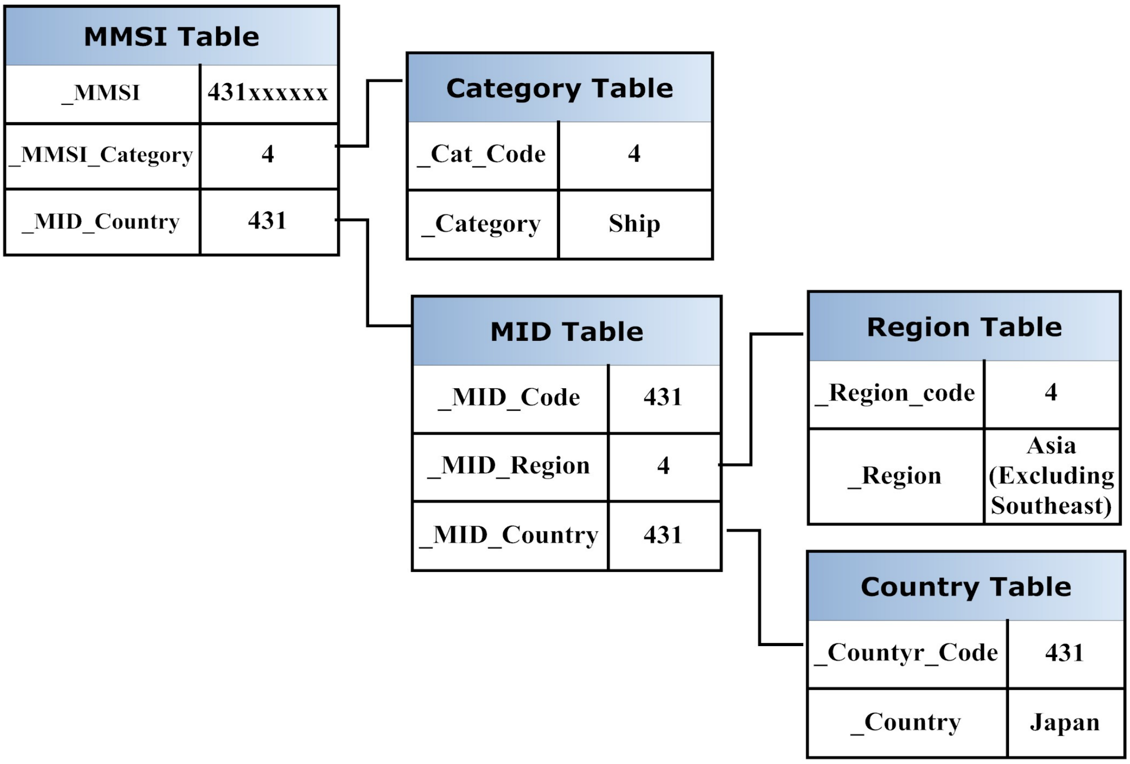


Fig 4.27 MMSI Index Legend

4.3.4. AIS Data Processing

The data stored in the AIS Data Warehouse presents many issues that can be summarized in the following:

- Discontinuity and non stability of interval of plotting.

To maximize the probability of intercepting data, the interval of recording has been fixed to 5 seconds. However, because of the AIS reporting intervals (2 seconds -10seconds), and knowing that most of the ships sailing in Tokyo Bay don't exceed 15 knots, the recorded data suffered duplication phenomena.

- Sometimes data is corrupted.

Some of the recorded AIS data is corrupted and full of errors (Example: MMSI=10).

In order to solve these issues and to increase the integrity of the data and thus the integrity of the database, an offline data processing is carried out as follows: Data Filtering, Data Interpolation, and Data Recompilation as shown in **Fig 4.28.**

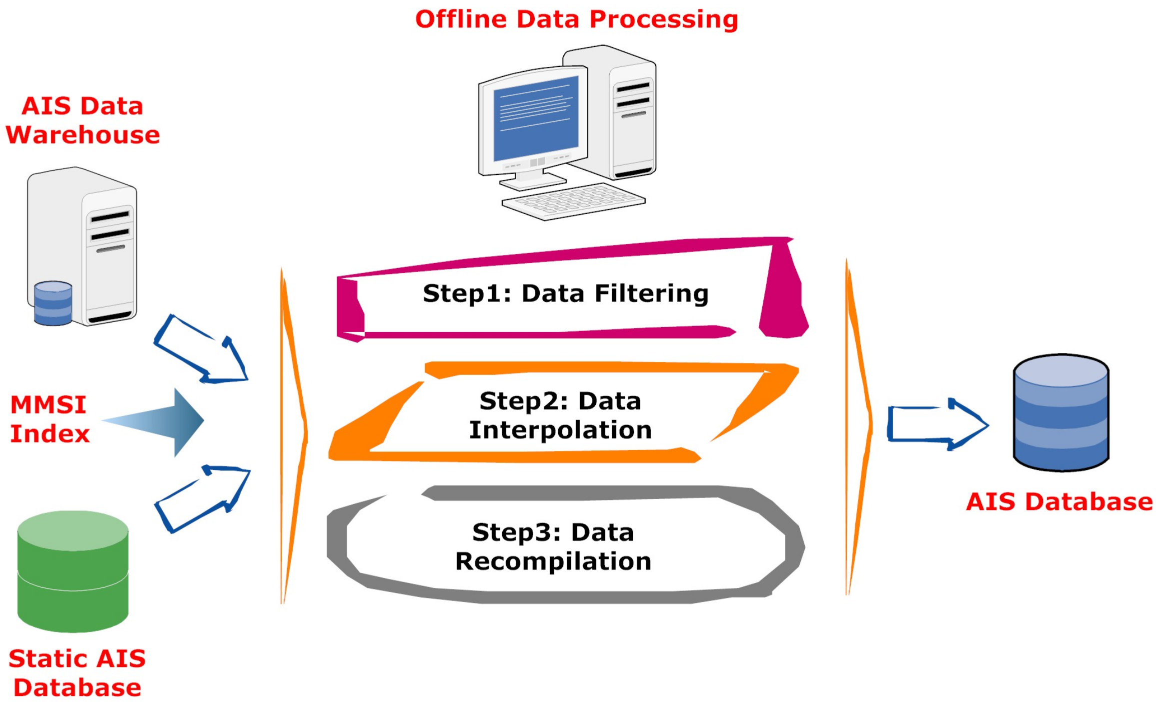


Fig 4.28 AIS Data Processing

4.3.5.1 Data Filtering

As mentioned before the recorded data suffered some discontinuity and interval instability, therefore to overcome these issues a data filtering has to be carried out as follows:

- Arrange data MMSI-wise for each moth.

The data exported to new MMSI named files contain only the following data: Date, Time, Longitude, Latitude, Heading, COG (Course Over Ground), SOG (Speed Over Ground), ROT (Rate Of Turn), and Destination.

- Build MMSI/IMO dictionary

The MMSI/IMO dictionary is to be used while recompiling the data and when matching static data form the Static Database (indexed by IMO number) with the dynamic data (indexed by MMSI number)

- Delete Obsolete MMSI files (MMSI numbers that are composed of less than 9 digits).

The data filtering process is shown in **Fig 4.29.**

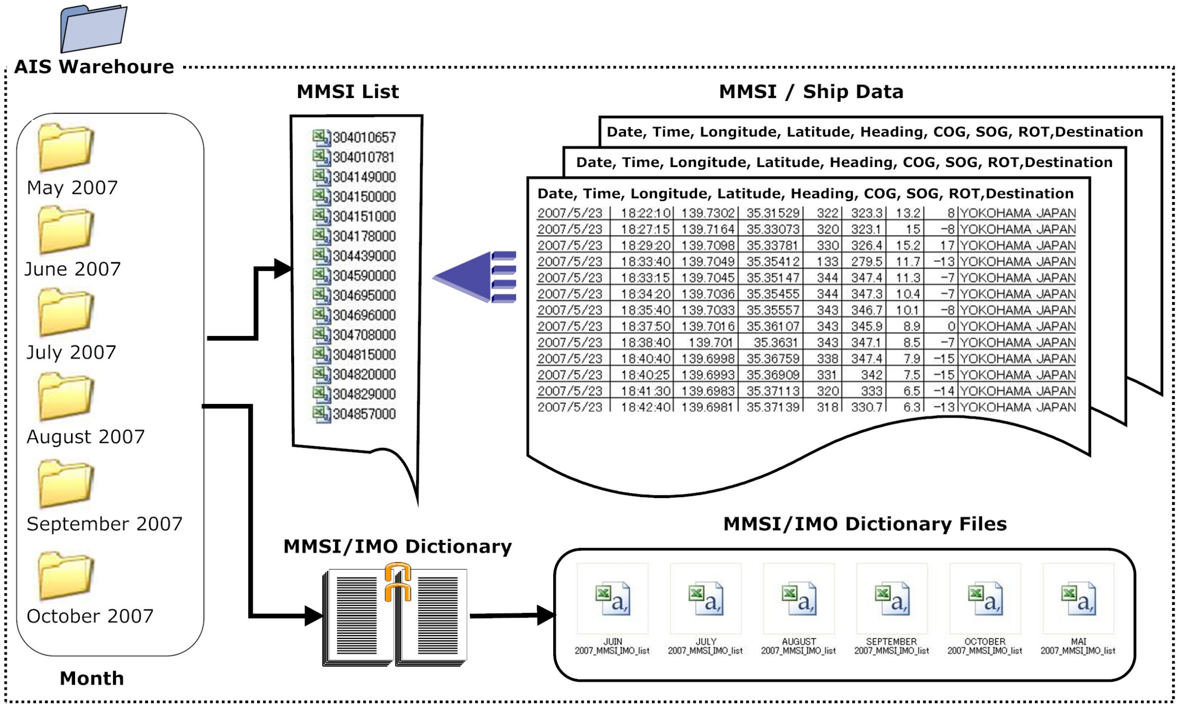


Fig 4.29 AIS Data Filtering

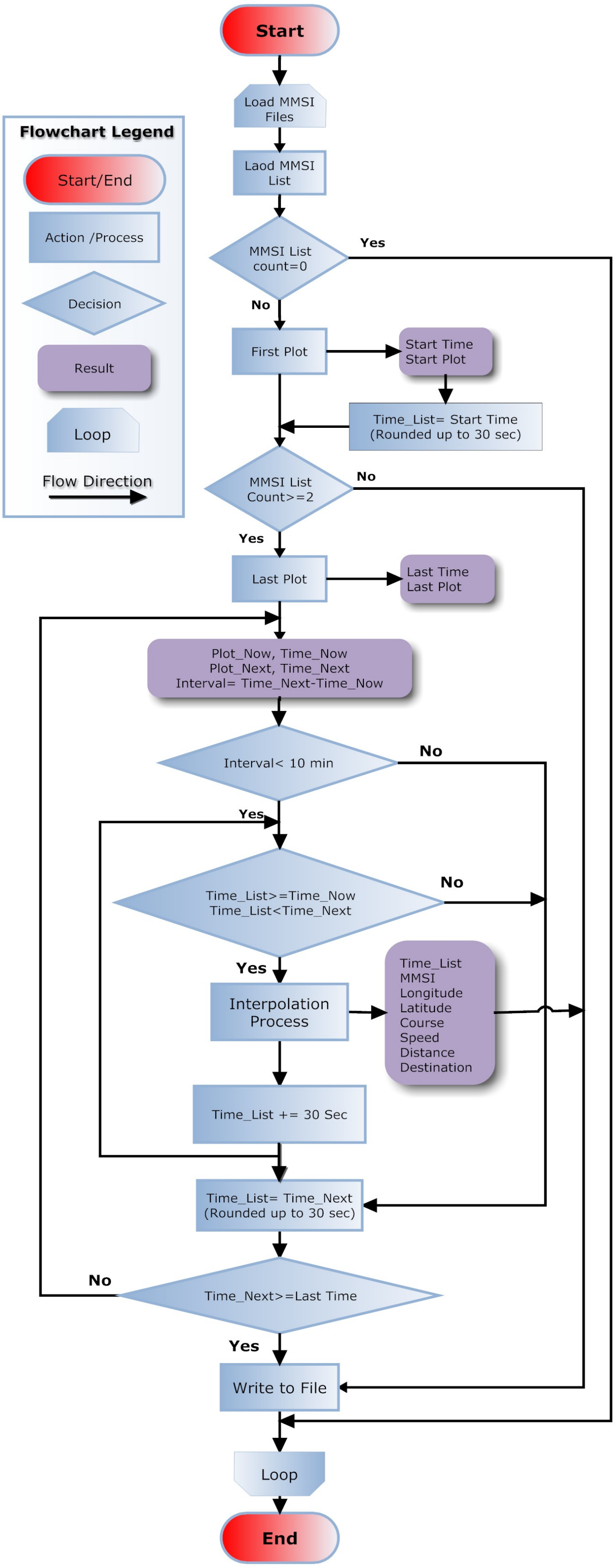
4.3.5.2 Data Interpolation

In this part of processing the data is interpolated to fit the gaps generated by unstable time intervals or discontinuity in plotting. Linear interpolation is carried out as shown in **Flowchart 4.4.**

The interpolation process can be summarized as follows:

- The MMSI file is loaded and read
- The First Plot is determined (Start Time and Start Plot (Other data))
- A Time List is generated based the Start Time that is rounded up to 30 seconds
- The Last Plot is determined (Last Time and Last Plot(Other data))
- The MMSI List is read plot by plot and the positions in between are interpolated
- If the interval between two plots is more than 10 minutes* the interpolation is suspended and the First Plot is reset to the next plot
- The interpolated data is exported to a file

* The interval of 10 minutes mentioned above is directly determined based on observation of recorded traffic, on the ROT, and on the fact that the ships sailing with a speed limitation (Uraga Suido Traffic Route (USTR) <= 12knots) can travel at maximum speed around 2 nm, which is less than the short segment of USTR.



Flowchart 4.4 AIS Data Interpolation Flowchart

As for the interpolation between two successive plots (Plot1 (time1, data1), Plot2 (time2, data2)), the linear interpolation formula **(1)** has been used as illustrated in **Fig 4.30**:

Int_Value= data1 + (data2 - data1) * ((time_interval1 / time_interval2) ………………………… **(1)**

Where:

Int_Value: Interpolated Value

data1 / data2: Data to be interpolated such as Longitude and Latitude.

time_interval1= Time_List – time1

time_interval2= time2 - time1

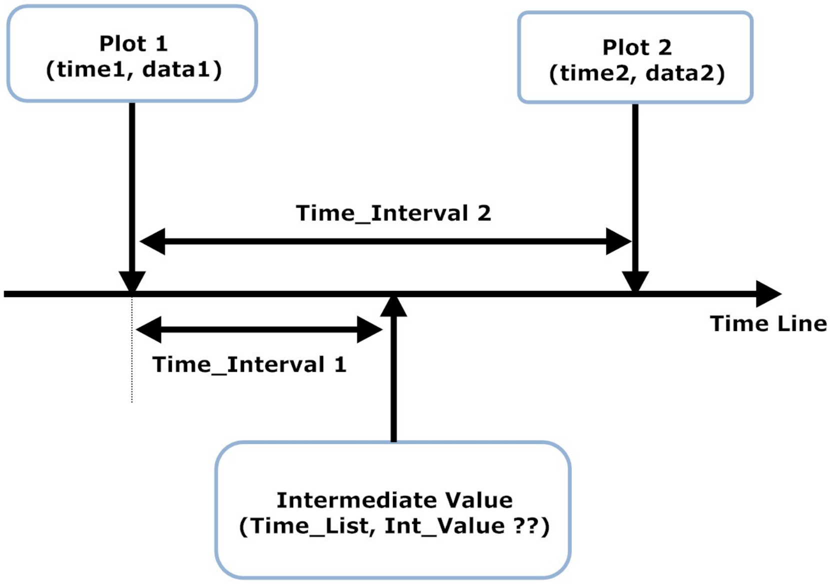


Fig 4.30 AIS Data Interpolation Concept

The calculation of the Course and Distance between two successive positions is carried out based on the navigational mathematics. However the Speed is calculated based on the motion speed calculation formula (Average Speed= Distance traveled / Time of travel).

The interpolated data is saved into files with the same format as for the data filtering.

4.3.5.3 Data Recompilation

At this stage of processing, the data is recompiled into time files as follows:

- The MMSI file is loaded and read
- The MMSI List is read plot by plot
- The MMSI/IMO matching is checked and updated by the MMSI/IMO dictionary built at the first stage of the data processing (Data Filtering)
- The recompiled data is exported to the time corresponding file

The final format of the recompiled data is as shown in **Fig 4.31**

Date, Time, IMO Nbr, MMSI, Longitude, Latitude, Course, Speed, Distance.

Where:

Distance: distance between two successive positions.

Fig 4.31 AIS Recompiled Data Format

Note: Destination is withdrawn from the final data because most of the time it is corrupted or not correct.

4.3.5. AIS Database Design

After processing the data, we have come to the final stage where the AIS Database is to be built and the previously processed data to be inputted into it.

4.3.6.1. AIS Database Structure and Format

The AIS Database is composed of 8 tables as follows:

- Dynamic data table
- Static data table
- MMSI table
- MMSI category table
- MMSI MID table
- MMSI region table
- MMSI country table
- MMSI/IMO dictionary table

The AIS Database EER Diagram and architecture are shown in **Fig 4.32 and Fig 4.33.**

Schema Tables | Schema Indices | Views | Stored procedures

ais_database
All tables of the ais_database schema

Table Name	Engine	Rows	Data length	Index length	Update time
dynamic_data_table	InnoDB	34855921	2.3 GB	6.4 GB	
mmsi_category_table	InnoDB	10	16 kB	0 B	
mmsi_country_table	InnoDB	272	16 kB	0 B	
mmsi_imo_dic_table	InnoDB	46496	1.5 MB	1.5 MB	
mmsi_mid_table	InnoDB	272	16 kB	32 kB	
mmsi_region_table	InnoDB	6	16 kB	0 B	
mmsi_table	InnoDB	48792	1.5 MB	3 MB	
static_data_table	InnoDB	144517	21.5 MB	14.1 MB	

Fig 4.32 AIS Database Tables

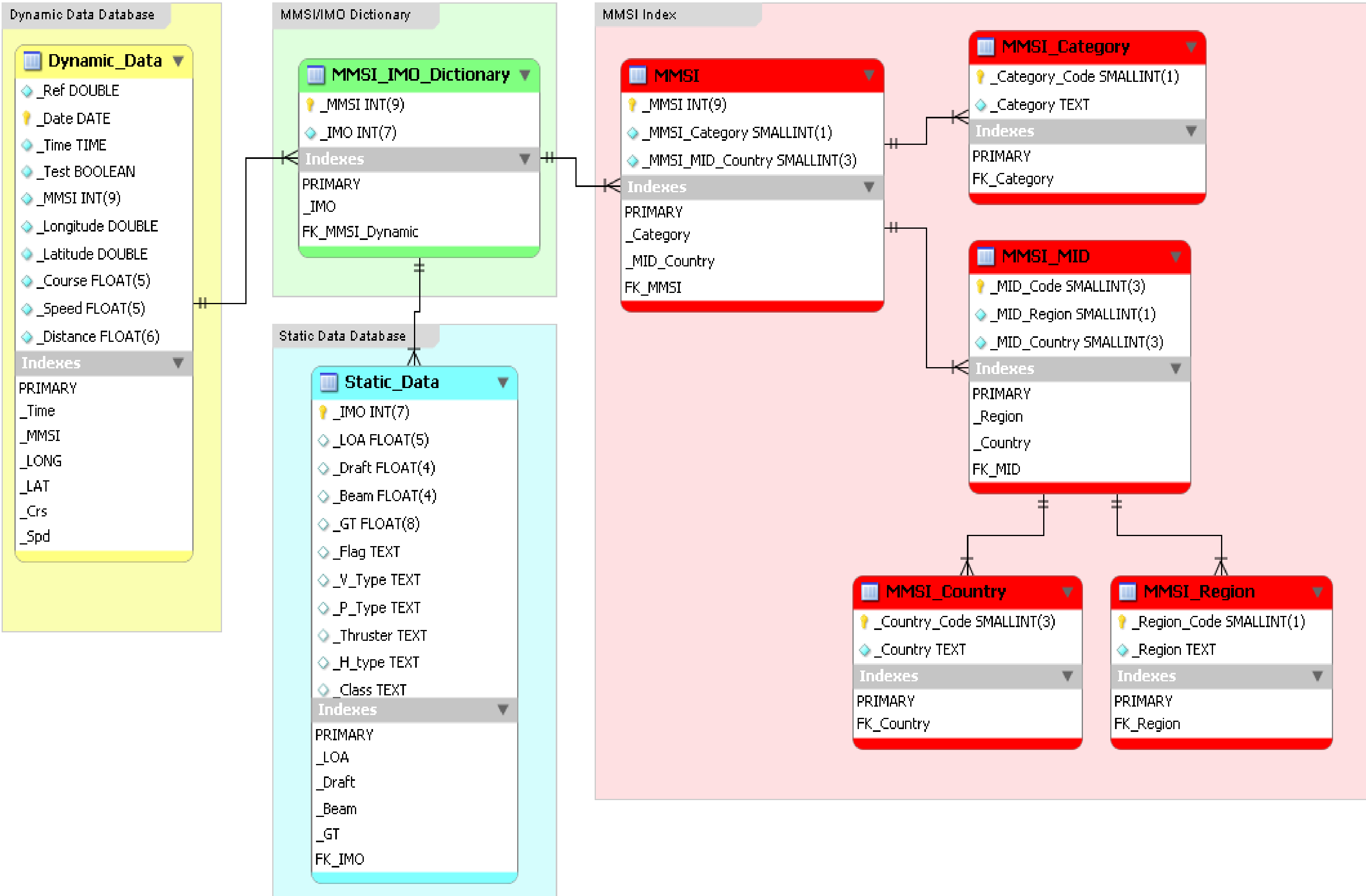
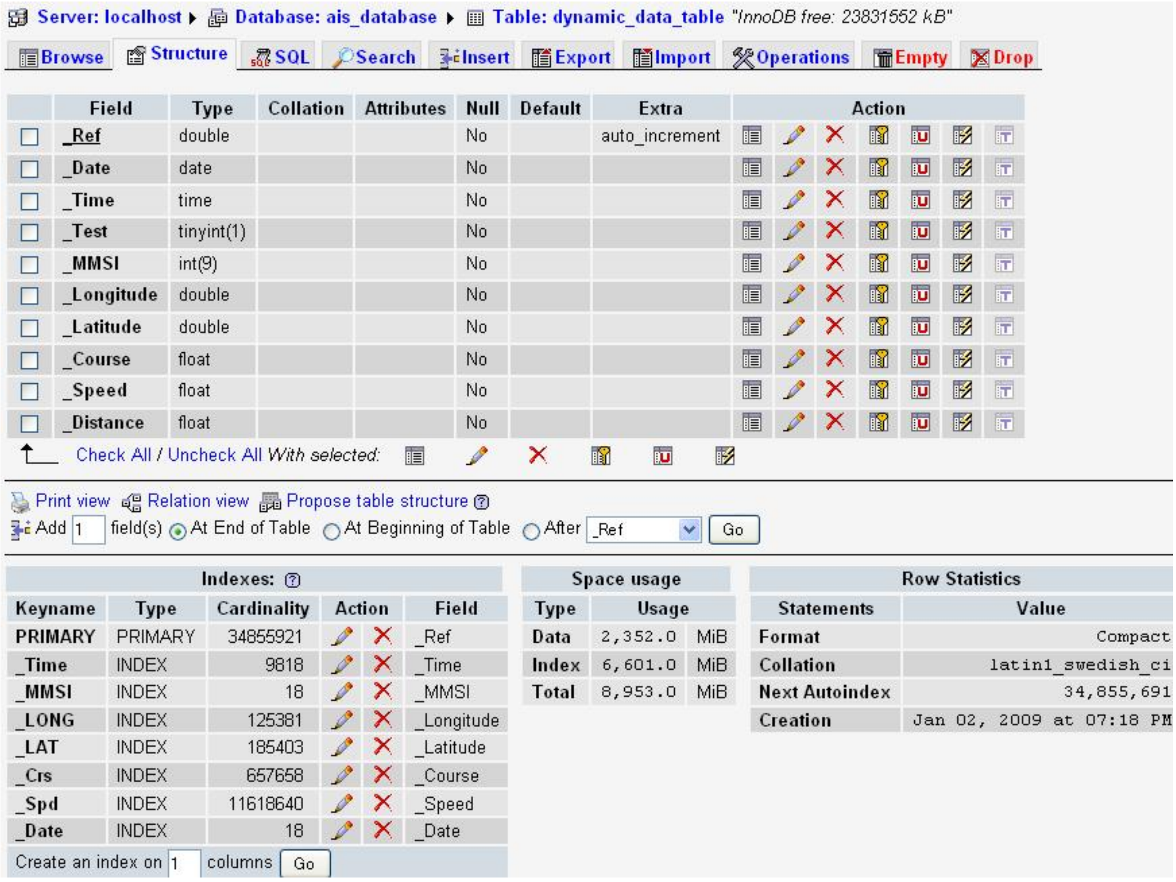


Fig 4.33 AIS Database EER Diagram

Since the other tables have already been fully explained in the Static AIS Database for the Static data table, and in the MMSI Index for the other tables (MMSI table, MMSI MID table, MMSI region table, and MMSI country table), only Dynamic data table and MMSI/IMO dictionary will be discussed herebelow.

The Dynamic data table structure is shown in **Fig 4.34**, whereas the data format is shown in **Fig 4.35**



Server: localhost ▸ Database: ais_database ▸ Table: dynamic_data_table "InnoDB free: 23831552 kB"

Browse | Structure | SQL | Search | Insert | Export | Import | Operations | Empty | Drop

Field	Type	Collation	Attributes	Null	Default	Extra	Action
_Ref	double			No		auto_increment	
_Date	date			No			
_Time	time			No			
_Test	tinyint(1)			No			
_MMSI	int(9)			No			
_Longitude	double			No			
_Latitude	double			No			
_Course	float			No			
_Speed	float			No			
_Distance	float			No			

Check All / Uncheck All *With selected:*

Print view   Relation view   Propose table structure

Add 1 field(s) At End of Table   At Beginning of Table   After _Ref   Go

Indexes:

Keyname	Type	Cardinality	Action	Field
PRIMARY	PRIMARY	34855921		_Ref
_Time	INDEX	9818		_Time
_MMSI	INDEX	18		_MMSI
_LONG	INDEX	125381		_Longitude
_LAT	INDEX	185403		_Latitude
_Crs	INDEX	657658		_Course
_Spd	INDEX	11618640		_Speed
_Date	INDEX	18		_Date

Create an index on 1 columns Go

Space usage

Type	Usage	
Data	2,352.0	MiB
Index	6,601.0	MiB
Total	8,953.0	MiB

Row Statistics

Statements	Value
Format	Compact
Collation	latin1_swedish_ci
Next Autoindex	34,855,691
Creation	Jan 02, 2009 at 07:18 PM

Fig 4.34 Dynamic Data Table Structure

Ref, Date, Time, Test, IMO Nbr, MMSI, Longitude, Latitude, Course, Speed, Distance.

Where:

- ✓ Ref: Reference ID (to make the line input unique)
- ✓ Test: Boolean (True/False)
  - True (1): Static data available
  - False (0): Static data not available
- ✓ Distance: Distance between two successive positions

Fig 4.35 Dynamic Data Format

The MMSI/IMO dictionary is intended to solve the issue stated beforehand; where the static data is indexed by IMO number while the dynamic AIS collected data is indexed by MMSI. So to say, the dictionary will be used as a bridge to connect the dynamic data to the static data. The MMSI/IMO dictionary table structure is shown in **Fig 4.36**

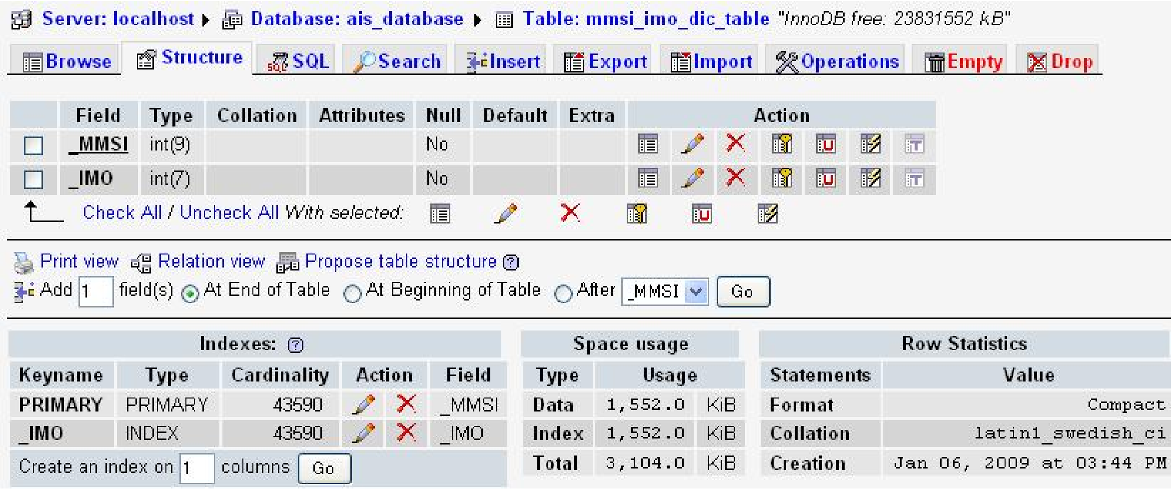


Fig 4.36 MMSI/IMO dictionary Table Structure

4.3.6. AIS Database Query Browsers

To connect to the server, a MySQL username should be provided and, most likely, a password. If the server runs on a machine other than the one where you log in, you will also need to specify a host name. Once the proper parameters are known, the connection can be made either through a Command User Interface (CUI), or through Graphical User Interface (GUI).

All statistics could be generated by querying the database using a single query, multiple queries and/or transactions. The users can build their queries and generate views with a combination of any attributes, and filter the data in a way to display only the desired information. The AIS database is provided with a Mapping GUI where the user can visualize the query results. The database can be accessed both locally and remotely through any secured network provided that required access parameters are known. The database data flow is shown in **Fig 4.37.**

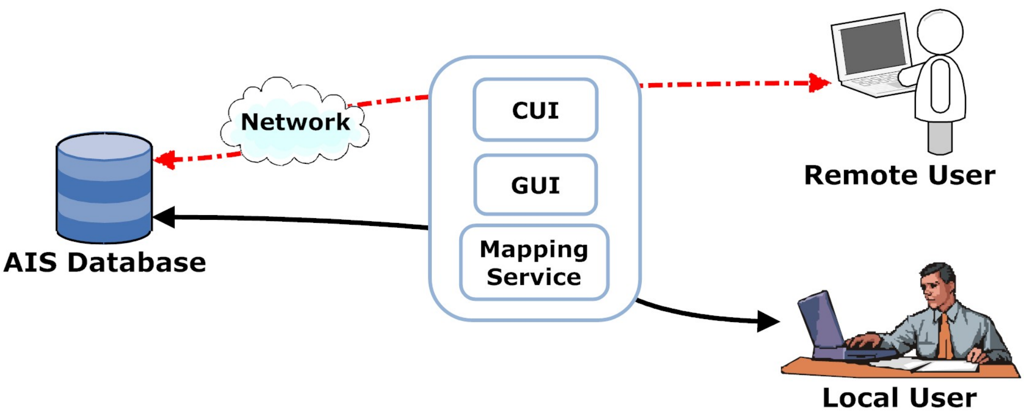


Fig 4.37 AIS Database Data Flow

4.3.7.1. CUI Query Browser

To connect to the server, browse, view, and export the transparent data of any of the built databases through the CUI; MySQL console shown in **Fig 4.38** is used and the required query is inputted in SQL language to extract the desired data as shown in **Fig 4.39**.

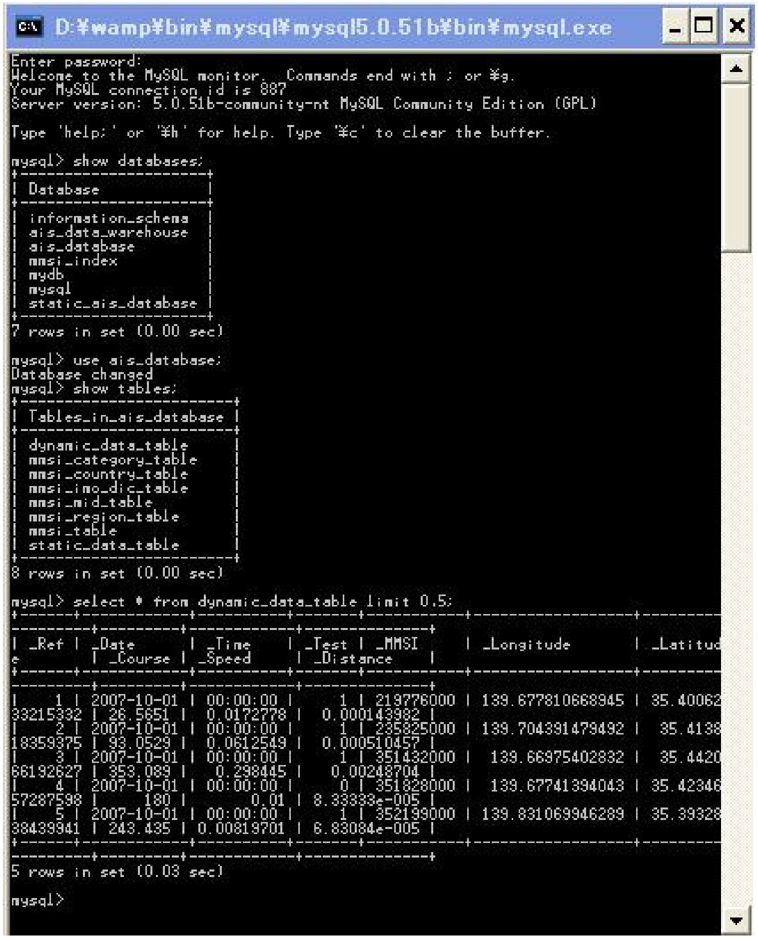


Fig 4.38 MySQL CUI Console

4.3.7.2. GUI Query Browser

To connect to the server, browse, view, and export the transparent data of any of the built databases through the GUI; MySQL Query Browser and WAMP phpMyAdm Query Browser (Refer to Appendix VII) are used as shown in **Fig 4.40** and **Fig 4.41**.

4.3.7.3. Mapping GUI

The AIS Database is provided with a mapping GUI, where the user can visualize the extracted data as shown in **Fig 4.42.** The Mapping GUI is also provided with a set of tools to carry out simulations and assessment of the marine traffic inside Tokyo Bay.

```
Shell> mysql -h host -u user -p
Enter password: ********
Welcome to the MySQL monitor. Commands end with; or ¥g.
Your MySQL connection id is 853
Server version: 5.0.51b-community-nt MySQL Community Edition (GPL)
Type 'help;' or '¥h' for help. Type '¥c' to clear the buffer.
mysql> use ais_database;
Database changed
mysql> show tables;
+-------------------------+
| Tables_in_ais_database  |
+-------------------------+
| dynamic_data_table      |
| mmsi_category_table     |
| mmsi_country_table      |
| mmsi_imo_dic_table      |
| mmsi_mid_table          |
| mmsi_region_table       |
| mmsi_table              |
| static_data_table       |
+-------------------------+
8 rows in set (0.00 sec)
mysql> select * from dynamic_data_table limit 0,3;
+----+-------+--------+-----+------+--------------+-----------------+-------+-------+----------+
| _Ref | _Date | _Time | _Test | _MMSI | _Longitude | _Latitude | _Course | _Speed | _Distance |
+----+-------+--------+-----+------+--------------+-----------------+-------+-------+----------+
| 1 | 2007-10-01 | 00:00:00 | 1 | 219776000 | 139.677810668945 | 35.4006233215332 | 26.5651 | 0.0172778 | 0.000143982 |
| 2 | 2007-10-01 | 00:00:00 | 1 | 235825000 | 139.704391479492 | 35.4138183593750 | 93.0529 | 0.0612549 | 0.000510457 |
| 3 | 2007-10-01 | 00:00:00 | 1 | 351432000 | 139.669754028320 | 35.4420661926270 | 353.089 | 0.2984450 | 0.002487040 |
+----+-------+--------+-----+------+--------------+-----------------+-------+-------+----------+
3 rows in set (0.08 sec)
mysql>…
Input Query/Transaction …
 mysql> quit;
```

Fig 4.39 Data Browsing with MySQL CUI Console

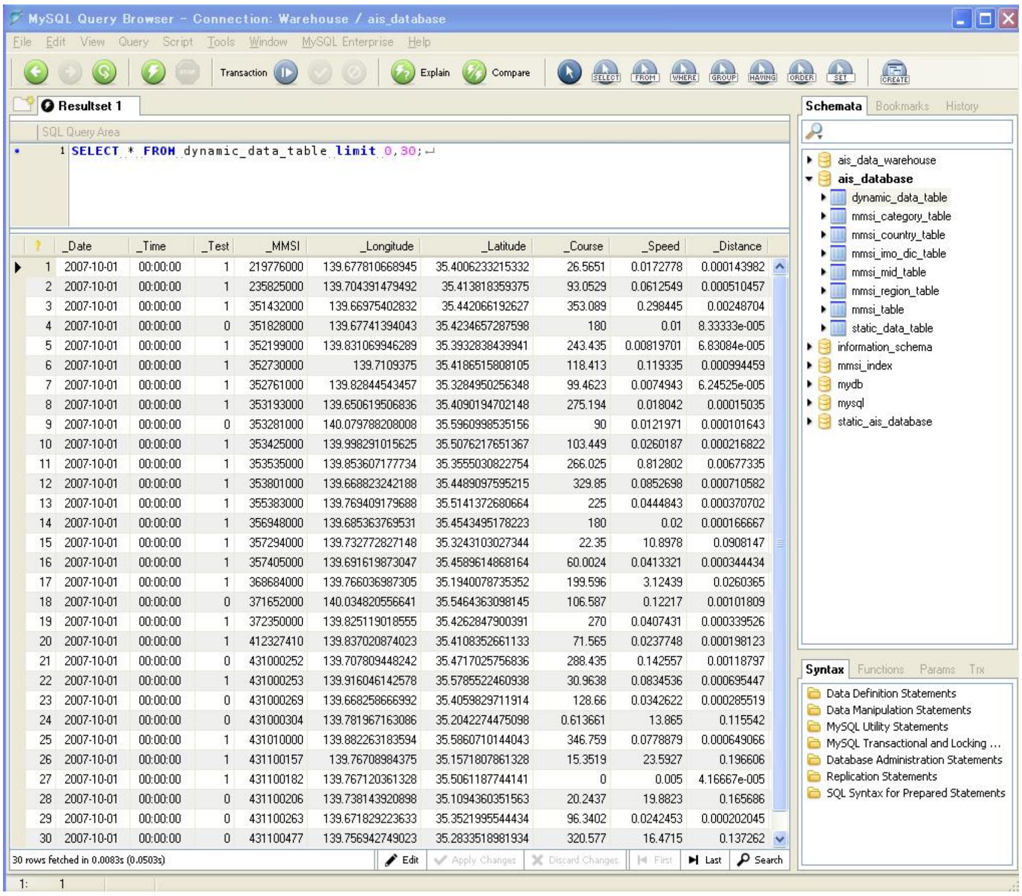


Fig 4.40 MySQL Query Browser

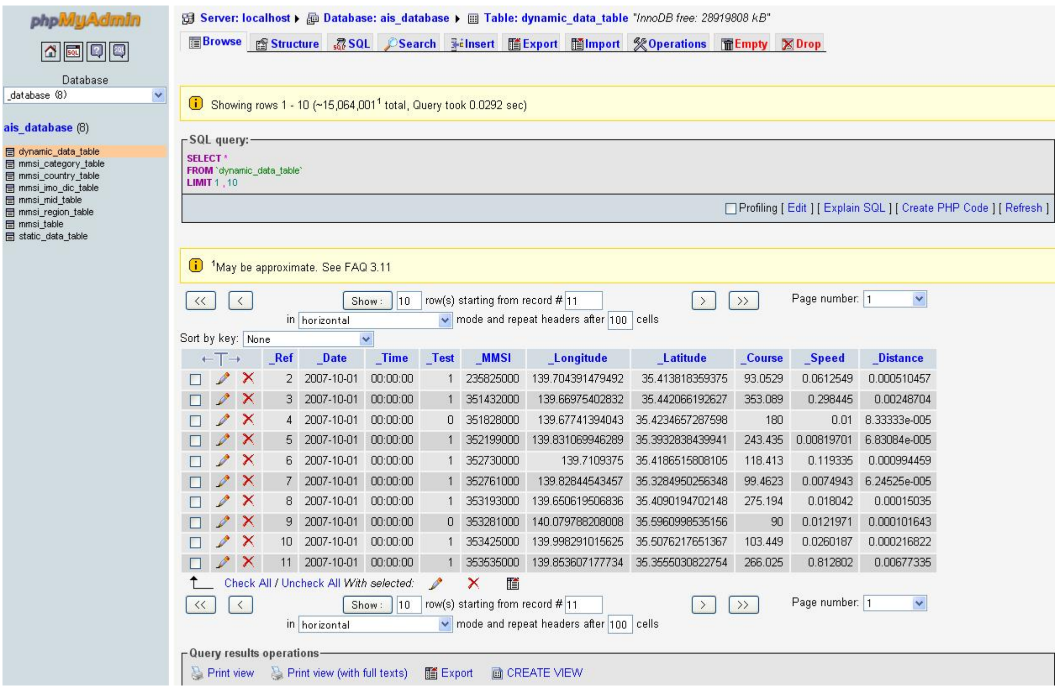


Fig 4.41 WAMP phpMyAdm Query Browser

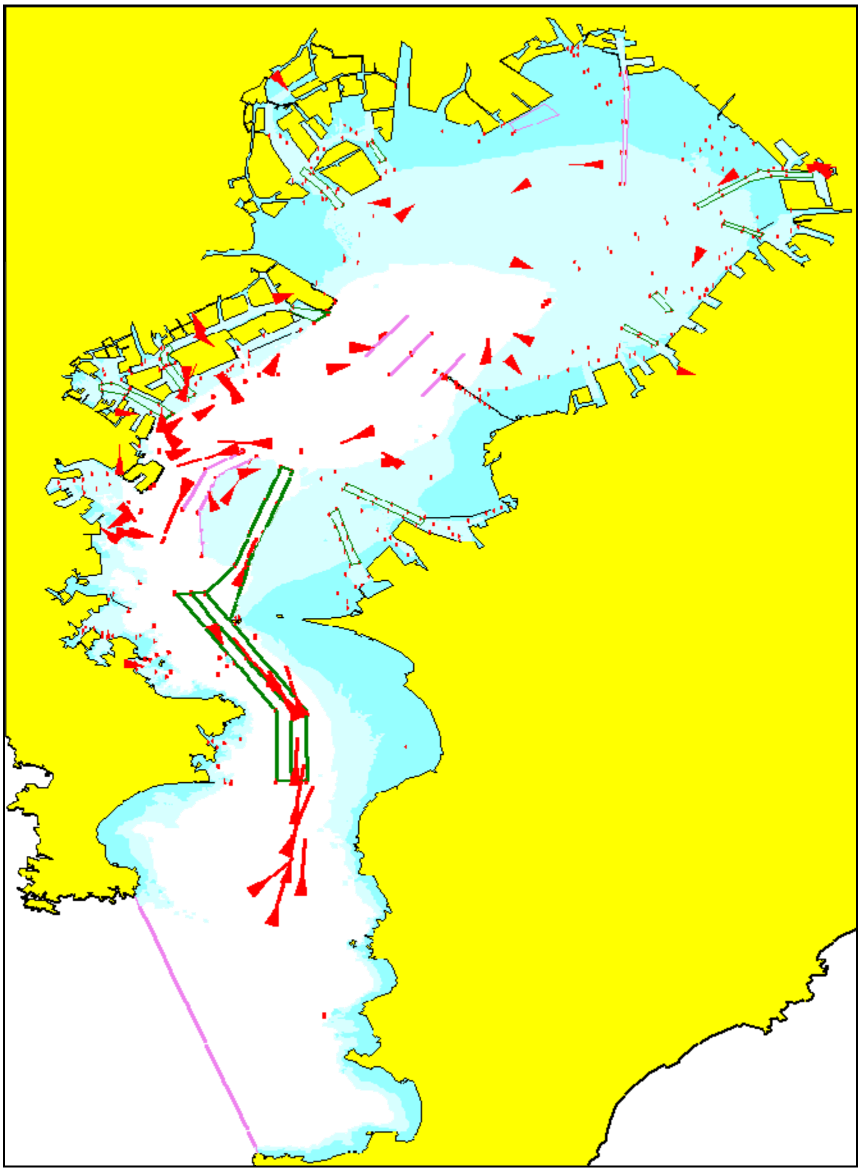


Fig 4.42 Mapping GUI .

4.3.7. Querying AIS Database

Taking advantage of the structure of the AIS Database, the AIS data can be used and/or visualized in a variety of ways. The desired data can be generated either in a single query or in a transaction where the display interface is decided by the final user. There are three basic groups of information that can be derived from the AIS data: Static information, Dynamic positioning information, and Voyage related information. Other related information can be combined between these main categories. In the following examples some typical queries are illustrated to give a glimpse of what can be derived from the built AIS Database.

4.3.8.1 MMSI Based Queries

The MMSI information can be queried in several ways such as:

- Category based queries: Only ships belonging to a specific category will be displayed.
- Region based queries: Only ships belonging to a specific region will be displayed.
- Country based queries: Only ships belonging to a specific country will be displayed.

An illustration of a region and country based query is shown in **Fig 4.43,** where only ships belonging to Asia countries (Excluding the Southeast) (region 4) are displayed. And specifically in red color Japanese MMSI registered ships (Country Code: 431- 432) and in bleu color ships other than Japanese but belonging to region 4.

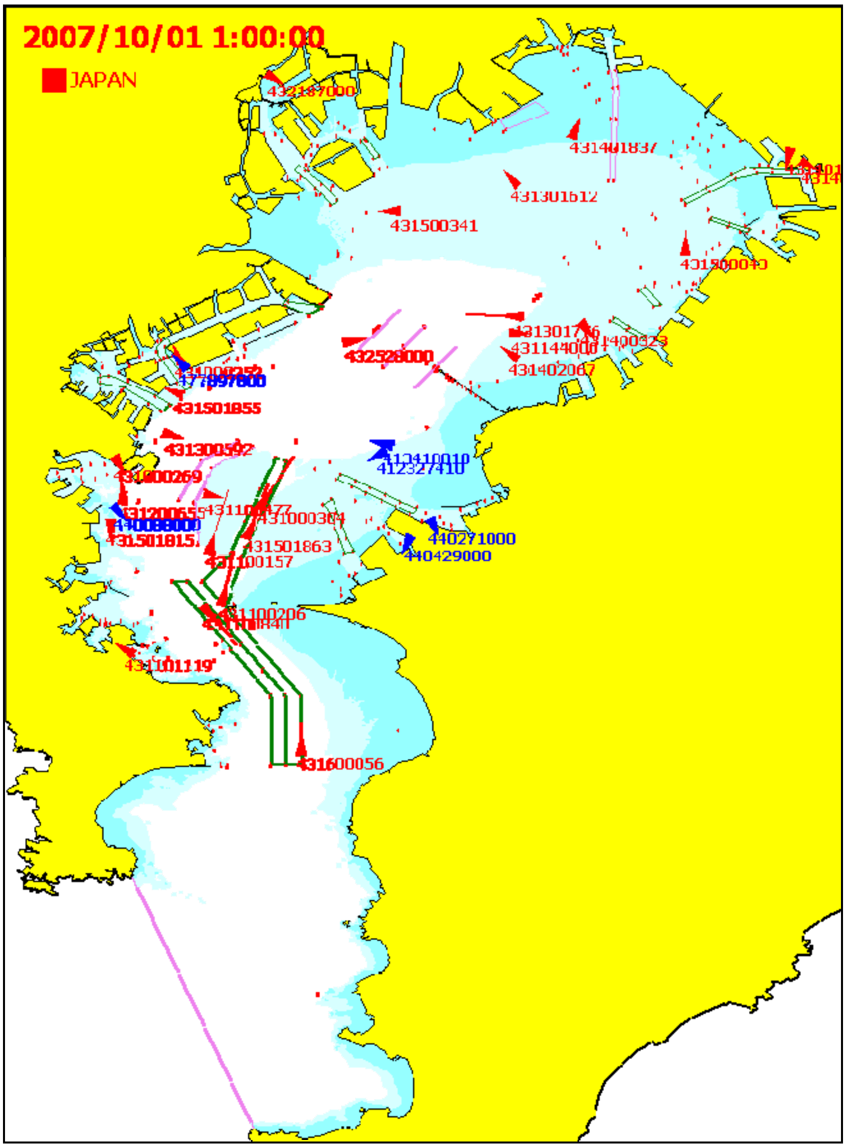


Fig 4.43 Region/Country Based Query (Asia countries excluding the Southeast))

4.3.8.2 Location Based Queries

To track and visualize the activity of ships, category of ships, and /or group of ships in a specified area, a location based query can be generated, where only the ships navigating in the selected area are visualized. An illustration of a location based query is shown in **Fig 4.44,** where only ships navigating inside the bleu rectangle are displayed.

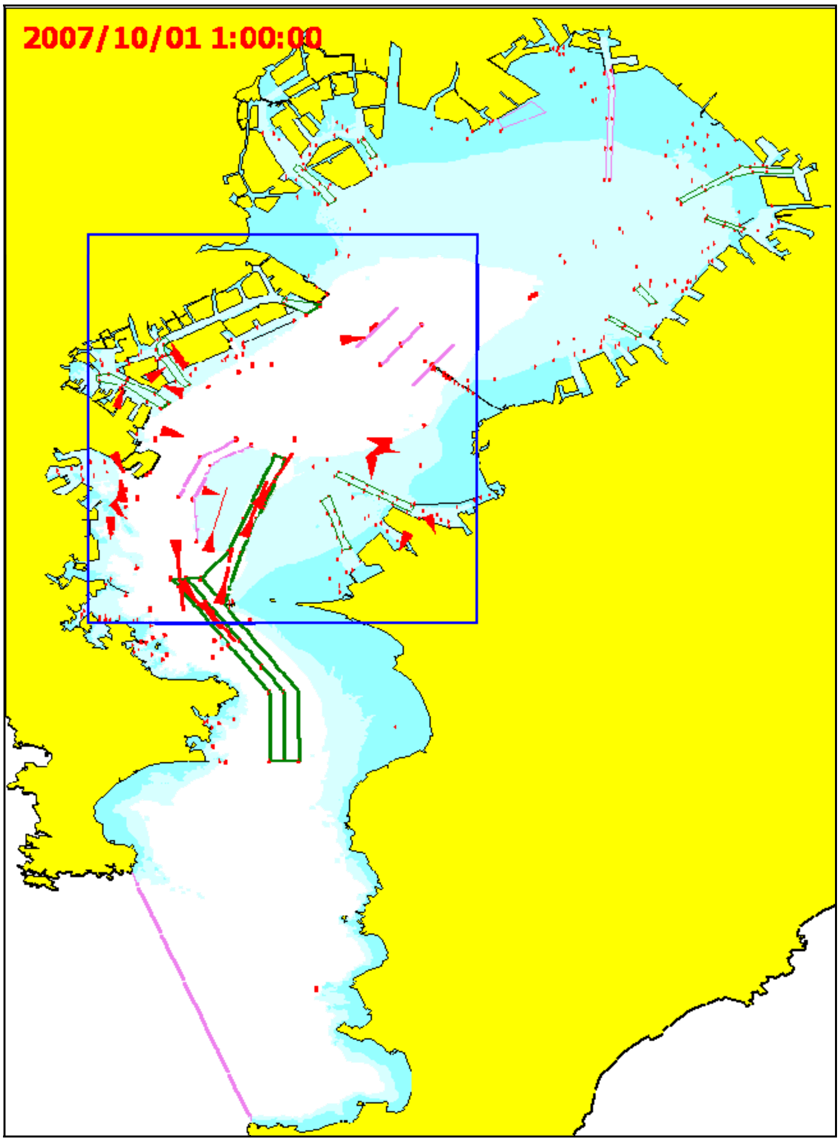


Fig 4.44 Location Based Query

4.3.8.3 Static Data Based Queries

The Static Data information can be queried and visualized in several ways such as:

- Length based queries (LOA and Beam)
- GT based queries

- Draft based queries
- Other parameters based queries (Flag, Class, Type of vessel…etc)

An illustration of a Static Data based query is shown in **Fig 4.45,** where the ships having a LOA between 100m and 150m are displayed in red color and the other ships in bleu color.

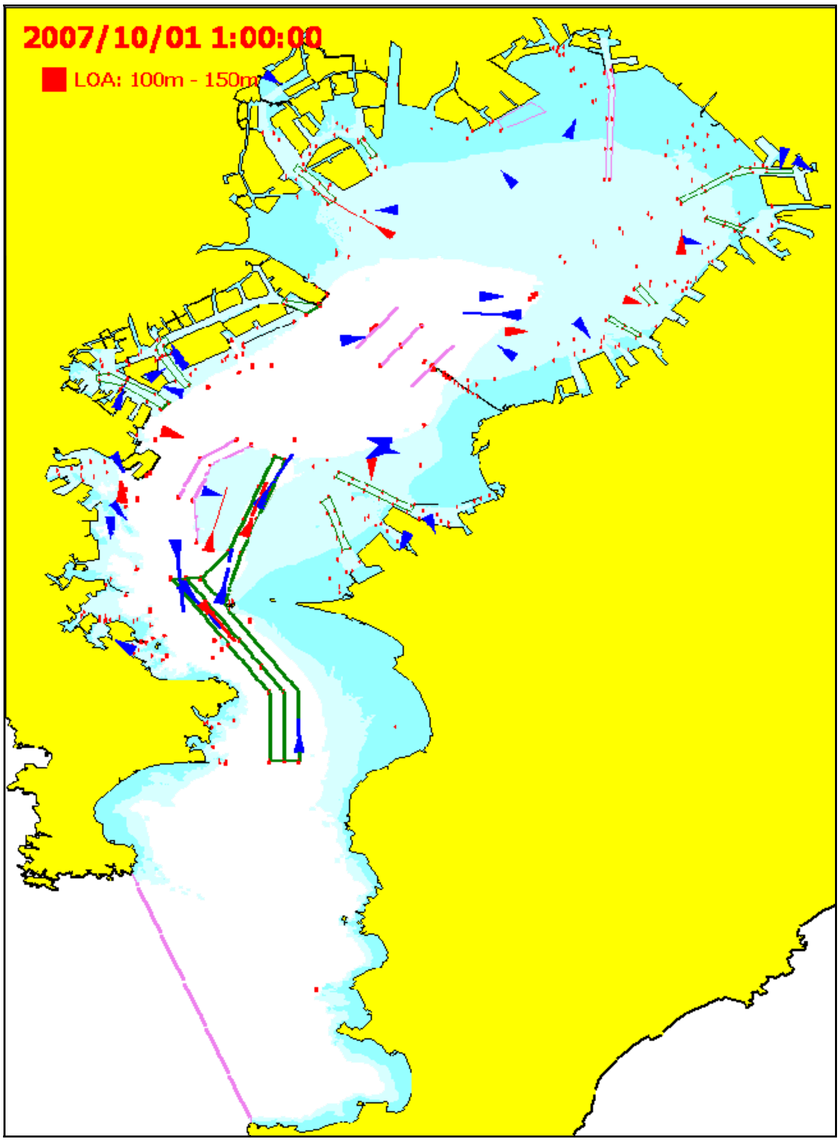


Fig 4.45 Static Data Based Query

4.3.8.4 Time Based Queries

The user can generate queries based on time interval in order to assess the presence and patterns of marine traffic or for any other purpose by a certain category of ships or the like.

An illustration of a time based query is shown in **Fig 4.46,** where the ships navigating in Tokyo Bay during 1 hour interval are displayed (from 00:00:00 to 01:00:00).

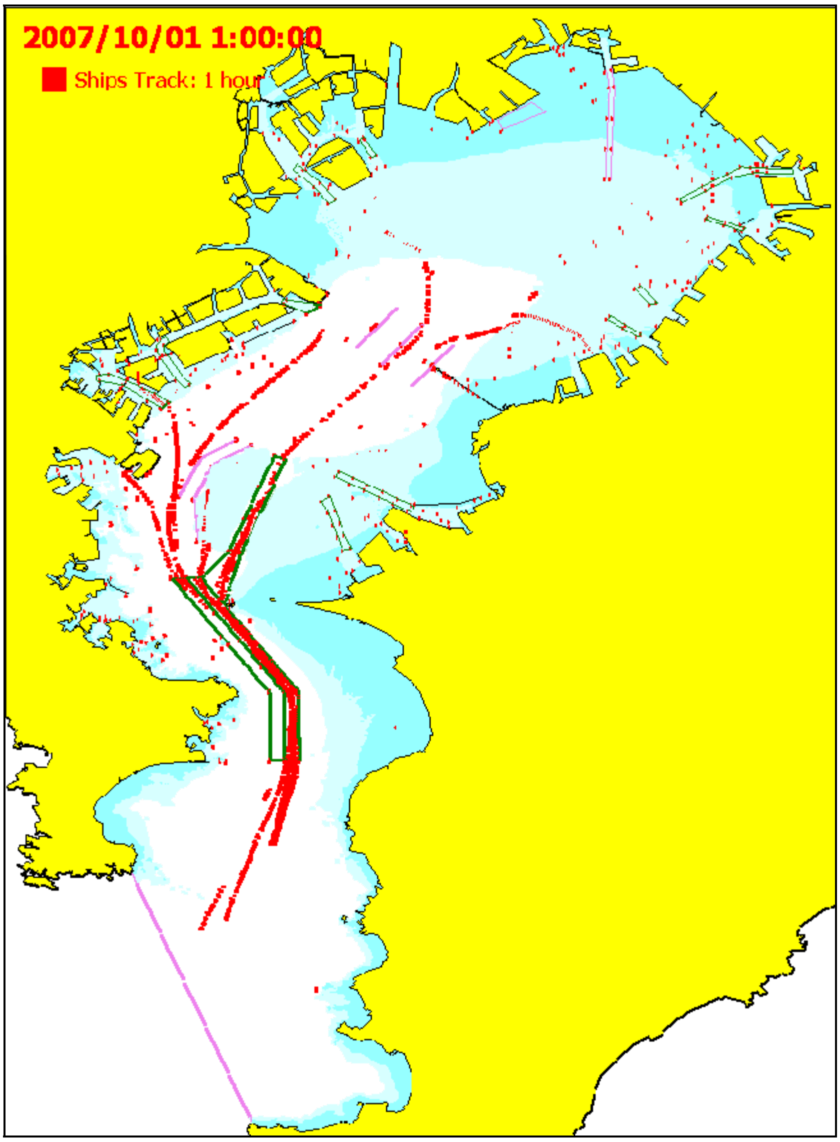


Fig 4.46 Time Based Query (Ships' tracks during 1 hour interval)

4.3.8.5 Queries for Statistics Purposes

Other queries, either single attribute or combined attributes, can be generated for statistics purposes, the parameters can be filtered and arranged in such away to give specific results. Hereafter some examples of the statistics that can be generated from the AIS Database:

- Number of ships belonging to a certain category, region, country…etc
- Number of ship that entered or left Tokyo Bay or a specific area in Tokyo Bay
- Number of ships with a certain speed, course, draft, beam …etc
- Average number of ships through a period of time
- Average speed of ships, category of ships, and type of ships…etc that transits a specific area.

Other statistics can be generated by combining the above mentioned queries to extract the desired information. Some statistics will be generated for the purpose of this study in the next chapter.

4.1. Conclusions

The AIS data presents vital source of information for a variety of purposes and administrations. By building the AIS Database based on MySQL technology, the AIS data information is easily accessed and visualized on either a CUI or GUI. Furthermore, the AIS Database offers many potential uses of the AIS derived information for statistic purposes and the like. However, the AIS Database is only as good as the AIS data received from ships; therefore the cooperation of the mariners is highly required to maintain the integrity of the information.