Joint Irish Cetacean Database Scoping Study

End of contract report [NPWS contract ref: D/C/269]

Eugenie Regan, Dave Wall and Barry O'Neill

Report to National Parks and Wildlife Service

20 August 2008

1. Introduction

The EU Habitats Directive (92/43/EEC) places an obligation on member states to accord all species of cetacean strict protection. This, according to Article 17, requires reporting on the conservation status of these species every six years. Conservation status must be assessed with regard to three primary parameters: natural range (or spatial extent); population status; and available habitat area. Ireland must achieve Favourable Conservation Status for all species. Unfavourable status would be a decrease of 1% in population per annum or a decrease of 10% of the range of a species over a six year period.

It is the responsibility of the National Parks and Wildlife Service to report on the conservation status of cetaceans in Irish waters to the European Commission. However, the recording of Irish cetaceans is conducted by a number of different groups, including the Irish Whale and Dolphin Group (IWDG); the Coastal and Marine Resources Centre (CMRC); and EU surveys (e.g. SCANS & CODA). These groups have their data in different databases with different formats and the nature of the original data collected may also differ. For effective assessment of the conservation status of these animals in Irish waters, one joint cetacean database is desirable that brings existing data sets together in a common format. This database would allow analysis for conservation purposes and identify gaps in current distribution knowledge.

The first steps to creating a database are to clarify:

1. What is the purpose of the database?
2. What information does it need to store?
3. What are the main requirements of the database?
4. Who will hold and manage the database?

Clarity on these four questions is essential before any development of a database is undertaken, as the answers will greatly influence how the database is constructed and the cost of managing the database.

Another aspect to consider from the outset is that there are two sets of data; those that have been collected in the past and those that may be collected in the future. This is an important distinction because the merging of old data may be limited whereas standards and protocols can be put in place for future data collection. Having the latter in place will facilitate complete compatibility of all data collected in the future and greatly cut down on the management of any joint database.

This report assesses the technical feasibility of merging the main Irish cetacean databases and presents a database schema as an option for the design of a Joint Irish Cetacean Database to hold past cetacean data.

A key recommendation is made that an Irish standard for cetacean data collection is developed and that all future funding stipulates that data is collected to this standard.

2. INVENTORY OF IRISH CETACEAN DATASETS

There are a number of different Cetacean Sightings Databases pertaining to Irish waters (see Table 1). The largest ones are described below.

a. Irish Whale and Dolphin Group

The largest data sets are those held by the Irish Whale and Dolphin Group (IWDG). They have their data in four separate databases;

1. Land-based watches (casual sightings and effort watches),
2. Ships surveys/ISCOPE/PreCAST,
3. Ferry surveys and
4. Strandings data.

b. University College Cork

The European Seabirds at Sea (ESAS) is a collaborative European database that is managed by the Joint Nature Conservation Committee (JNCC) in the UK. University College Cork is the Irish node for this database. Although the main focus of the ESAS database is seabirds, cetacean data is also collected and stored in this database. The ESAS database for Irish waters covering the period 1979-2003 is summarised by Mackey et al. (2004).

c. Coastal and Marine Resources Centre

A number of projects led by the Coastal and Marine Resources have generated cetacean records, including the RAMSSI project (Roycroft et al., 2007), various trip reports from Ireland's Atlantic margin (Mackey et al. 2004), Broadhaven Bay and the SIAR survey.

d. SCANS I and II and CODA

Small Cetacean Abundance in the North Sea and adjacent waters (SCANS) was conducted in July 1994. SCANS generated the first large-scale abundance estimates for many cetacean species. Irish waters included in this survey were the Celtic Sea. SCANS II followed on from this in 2005 and extended the areas surveyed to include, amongst others, the Irish Sea and areas off the west and north coast of Ireland. The CODA project which follows on from SCANS began in January 2007. This project is currently undertaking surveys of offshore waters to include areas of the west of Ireland. SCANS I and II and CODA are co-ordinated by the Sea Mammal Research Unit at the University of St. Andrews, Scotland. The main funding came from EU LIFE-Nature funding but also from co-financiers including the National Parks and Wildlife Service. All data collected from these surveys will be made available to the public.

3. TYPES OF DATA

There are three types of data with different levels of analysis applicable to each. 1. Sightings data, 2. Effort-related sightings data, and 3. Covariate data. Joining datasets can be seen as merging these three types of data. The greatest degree of compatibility is among sightings data with less compatibility among effort and covariate data. Tables 2, 3, and 4 show the compatibility of the various attributes recorded by the four main databases (IWDG ship surveys, IWDG ISCOPE, ESAS, and SCANS). A description of these attributes can be found in Appendix 1.

i. Sightings data

Location, date, species identification and numbers of individuals are the four data items that are universally present in any sightings dataset. Collating these is the simplest means of joining datasets together. This data can be used to describe the range of a species. However, without information on the search effort, relative abundance cannot be determined.

All of the existing sightings databases have this minimum information (Table 1). Of the 25 variables recorded over all four datasets, six are unique to one or other of the four. Eleven are present across all four or can be obtained as they are implied by fields in the same or related tables (Table 2).

ii. Effort-related sightings data

Effort data allows the examination of cetacean distribution and population status in terms of relative abundance. Relative abundance of cetaceans may be expressed in terms of the number of animals sighted per unit distance travelled or the number of animals sighted per unit time surveyed. The basic requirements for obtaining relative abundance of cetaceans are effort data (time surveyed and/or trackline length, location, and date) and sightings data (date, location, species identity and number of animals sighted – assuming that the observer is vessel-based). Table 3 shows the compatibility of effort data for the four main data sets. Three data sets (ESAS, IWDG ship-based effort surveys, and SCANS) have information from which trackline length can be calculated. However, the IWDG ISCOPE is land-based and effort is measured in time watched. Animals-per-hour is the unit of relative abundance that best accommodates all effort-related data.

Some of the cetacean databases will not be suitable for analysing relative abundance of cetaceans in Irish waters including casual sightings collected by the Irish Whale and Dolphin Group, the Marine Institute, NPWS, EHS, and CEDaR (Table 1).

iii. Covariate data

Effort covariate data includes sea state or other weather conditions, animal behaviour, observer, and platform identity – all of which can affect sighting efficiency.

One of the main sources of potential bias arises from sea state conditions. Survey effort should be adjusted by an appropriate correction factor to compensate for this bias. Other covariate data that may influence observations are observer and visibility which are collected across all four data sets (Table 4). Swell height may also influence visual detection but this was only collected in three of the data sets examined. Other factors that may also affect visual detection include number of observers present, the speed of the observation platform, the eye height of the observer, and the observer's experience. These are more difficult to build in to a correction factor.

4. ATTRIBUTES OF THOSE DATASETS AND COMPATIBILITY

The merging of datasets into a common format is not entirely straightforward with the main problem being that data is not stored in a standard format.

There are two main issues that apply to the compatibility of most fields; units of measure, and duplication and semantics. Semantics is where different terms are used to describe the same item, and duplication is where, because of semantics, recording of the same item is duplicated. Different units of measure may be used in the different databases, for example hours, minutes, or seconds. These will need to be converted to one standard. Duplication and semantics will be one of the main difficulties in merging data for the Joint Irish Cetacean Database.

Figure 1: example of duplication and semantic difficulties

	SCANS	ESAS
Behaviour ID	LO	60
Behaviour	Logging/sleeping/resting	Resting or apparently asleep

The problem is apparent in the above illustration (Figure 1). When storing resting behaviour, the SCANS database stores the behaviour with an *ID* of *LO* and a *description of Logging/sleeping/resting*, while the ESAS database stores the behaviour with an *ID* of *60* and a *description* of *Resting or apparently asleep*. Each database will be mapped into the Joint Irish Cetacean Database with relevant fields being mapped to the standard.

The ESAS database is different from the other databases in a number of important areas. One of the key differences between ESAS and the other databases examined is that ESAS does not store point data, but rather records start, middle and end points and then uses time and vessel speed to interpolate the co-ordinate of the sighting. There are two methods used by ESAS to do this.

A. Use the last recorded position of the ship as the sighting lat/long of the ship. This method is accurate to approximately 3km resolution but this figure can fluctuate depending on the speed of the vessel.
B. Use the time measurement for the sighting to interpolate a position for the sighting. While this is more accurate than the above method, time for a sighting is rarely recorded.

Due to the fact that ESAS locations are estimations and not accurate point data, the Joint Irish Cetacean Database will be limited to the resolution of this data set.

Once the technicality of merging datasets has been addressed, the feasibility of merging data from different surveys should also be considered. For example, is it reasonable to directly compare effort data from a SCANS survey versus effort data from an IWDG ship-based survey?

5. STRUCTURE OF FUTURE JOINT IRISH CETACEAN DATABASE

The different options

There are two principal options for the storage of the Joint Irish Cetacean Database; one is the use of existing recording software, with specific customisation to cater for cetacean data, and second, a custom designed relational database. Both options would require the development of an online mapping interface for the display of the data.

Option 1.

A database that will have an interface with which ecologists can easily access and query the database themselves. This will either require the development of a user-interface or the development of existing software packages.

Existing software packages

There a number of different biological recording software packages available that provide standards-based tools for collection and collation of biological recordings. They ensure standardised collection of biological records and smooth integration of records into larger databases such as those held by Biological Records Centres. One has been specifically designed for marine data, Marine Recorder. However, Marine Recorder has been developed for rocky-shore and benthic data, it has a small capacity, and no field for attributes that are important in cetacean monitoring, i.e. sea state (Figure 1). A more powerful tool is Recorder 6 which has been built specifically to hold terrestrial data and is the software package recommended for use by the NPWS. Both packages would need development in order to make them suitable for cetacean data, but development of Recorder 6 rather than Marine Recorder would have clear advantages.

Figure 1. Advantages and disadvantages of two software packages for holding cetacean data.

Software packages	Advantages	Disadvantages
Marine Recorder	• Takes lat/long using WGS84 spheroid (generic worldwide system)	• Designed for rocky-shore type data • Based on Access database • Small capacity • Cannot store photos, etc. • No field for weather description, ship height, ship speed, glare, sea state
Recorder 6	• Existing, large piece of software, specifically developed by JNCC for biological records • Field for weather description (free text)	• Dorset software would have to write a special add-ins to make system suitable for cetacean data

The use of Recorder 6 for the Joint Irish Cetacean Database would require the design of custom input form/front-end (graphical user interface) to make it appropriate for storing the relevant data. The key requirements of this Recorder Add-in would be to:

a. Allow entry of relevant data for storage by Recorder;
b. Enable recall of entered data for review and further editing, and
c. Retention of selected entered data items to facilitate rapid entry of subsequent, similar creatures.

The development of the Recorder Add-in would need to be specified in detail, be agreed by the Recorder Steering Group, and developed by a suitably qualified software developer. Dorset Software was approached to provide an indicative cost of development of a Recorder Add-in for cetacean data. Using the Irish Whale and Dolphin data entry template (Figure 2) as a guide, the following indicative costing was proposed:

Design Phase:	St£5,500
Build Phase:	St£19,250 – St£24,250
Deployment Phase:	St£2,500 – St£3,000
TOTAL:	**St£27,250 (€34,408) – St£32,750 (€41,352)**

The delivery of the Add-in could be accomplished within 5 months.

Figure 2: Irish Whale and Dolphin Group's data entry template

DATA ITEMS

Contact details including:
- Name
- Address
- Telephone number
- Email address

Observer position and position(s) of sighting(s):

Environmental factors including:
- Sea state and swell
- Wind force and direction
- Visibility
- Glare

Watch details including:
- Date
- Start & finish times and duration
- Location with county, longitude & latitude, height above sea level
- Optics used- binoculars or telescope – and specifications

Details of sightings including:
- Sighting number
- Time
- Species
- Statement of certainty- definite, probable, possible
- Numbers – overall and adults/juveniles
- Behaviour
- Supporting comments

Option 2.

A relational database designed specifically for the Joint Irish Cetacean Database. This database would need to be managed by I.T. personnel and all data requests would be through them.

This would be a stand-alone database, managed by personnel with a high level of IT skills, with all data submission and extraction typically channelled through them. A relational database structure would ensure simplicity and expandability and reduce errors on data entry.

A potential schema for the Joint Irish Cetacean Database has been developed (Figure 2). This was developed having assessed the key attributes from the potential donor datasets and the compatibility of these attributes. The schema presented below has been designed to accommodate the different key attributes and would be suitable for the purposes of the Joint Irish Cetacean Database.

The cost of development of this stand-alone database is difficult to determine due to the unpredictability of many aspects. However, having obtained indicative costs for the Recorder option, it may be a more expensive option than developing Recorder.

The Preferred Option

Having explored both options in detail, and following discussions with key cetacean experts, the use of Recorder 6 for the Joint Irish Cetacean Database would have distinct advantages over a specially designed relational database. The primary advantage is that the Recorder option is likely to be more cost effective, but it would also assist the national biological data management systems being promoted by NPWS and the National Biodiversity Data Centre.

Design of online mapping system

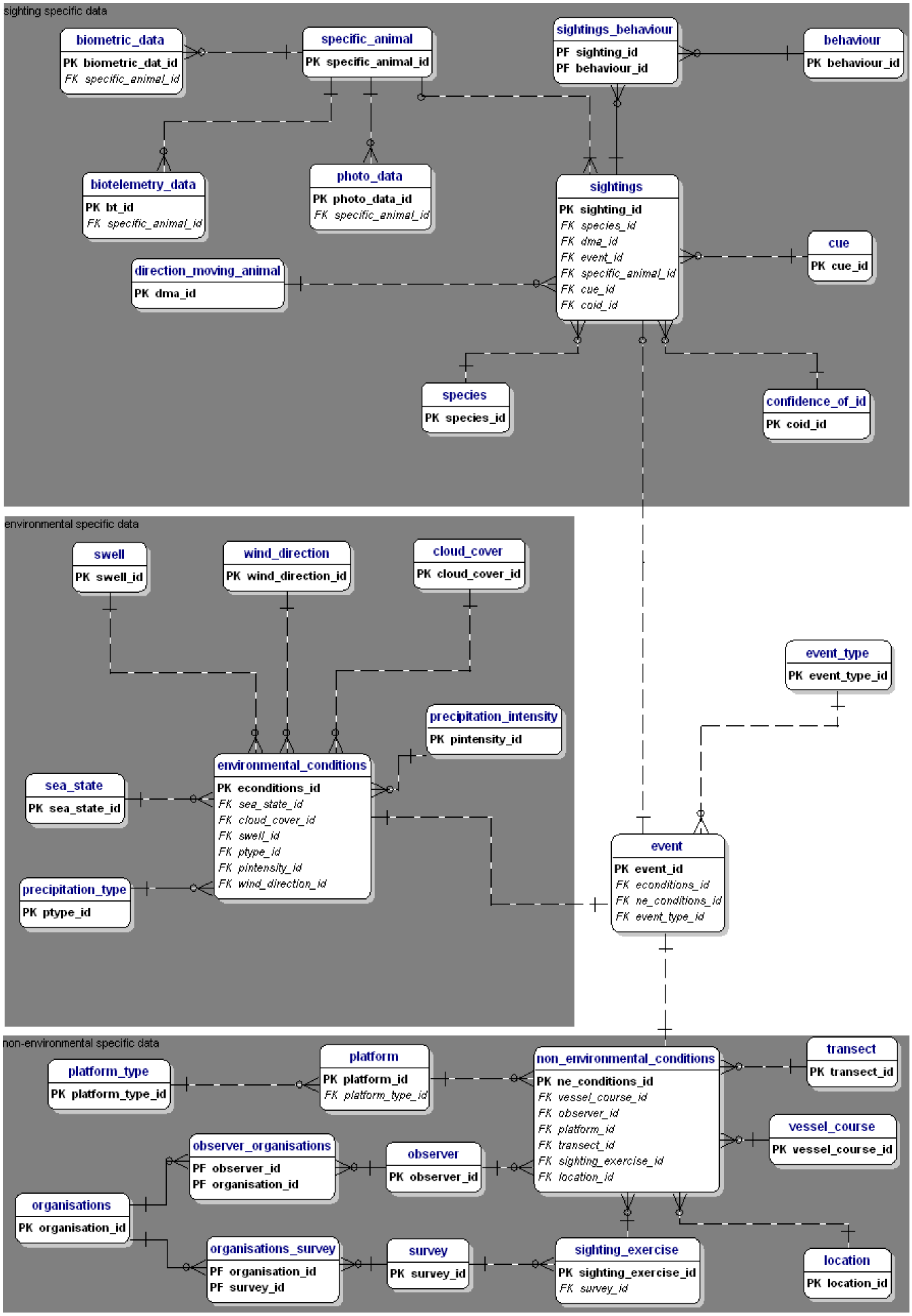
A clear requirement for the Joint Irish Cetacean Database is to have a public interface from which the public could view distribution information on Irish cetaceans. The National Biodiversity Data Centre has been funded by the Department of Environment, Heritage and Local Government, through the offices of the Heritage Council, to develop an on-line mapping system for the terrestrial environment. The core development work of this system is now complete and building on this development work to produce a marine equivalent would be a cost effective option. It would also have the advantage that such an approach would integrate the Joint Irish Cetacean Database into existing national biological recording management systems.

The on-line mapping system would allow:

- Display of species distribution on a grid basis,
- Mapping of species distribution as a GIS layer against other background GIS layers eg. bathymetry.
- Display of attributes of each data record, including date, recorder’s name, database source, record precision;
- Provision of metadata for each dataset;
- Generate dynamic dataset summary information, on spatial, temporal and species content of individual datasets.

The cost of this development work would depend on the detailed specifications, but a realistic effort would be between 40 to 60 days development work, at a cost of €28,000 - €42,000 excl. VAT.

Figure 2. A proposed database design/schema for a Joint Irish Cetacean Database.



Populating the Recorder Database

Were the preferred approach of an extended Recorder database to be adopted and delivered, the next stage would be the transfer of the existing datasets to this extended Recorder 6 database. This task will require transformations in order to transform the disparate datasets to a Recorder 6 compatible format. This would be performed by a Recorder 6 certified consultant. Initially this task would be performed on a one off transfer basis. However once the data has been transferred, add-ins (custom tools) can be developed by a Recorder consultant to allow the import of said datasets by the National Biodiversity Data Centre in future. Recorder 6 add-ins are cost effective as they remove the need to pay a consultant to import data every time a dataset update is received, which would be wasteful and impractical in the long-term.

The exact cost of these transformations is difficult to determine. However, as a rule of thumb if the data is coming from a database where the records are in a standard format then it would normally require 3 or 4 days work, irrespective of the number of records[1].

[1] Mike Weideli *pers comm.*.

6. OTHER DATA TYPES THAT NEED TO BE CONSIDERED FOR STORAGE

Acoustic Data

The storage of acoustic data is a possible requirement of a cetacean database. There are some challenges to storing this data. The immediate difficulty is the size of the data. Acoustic data can span from hundreds of gigabytes to terabytes in size. Storing this data would require the on-going purchase of hardware and its continued maintenance. Examples of the hardware required would be; servers, hard disks, software, and cooling equipment to dissipate the heat from the servers and hard disks.

Other considerable costs would be the set up and maintenance of the equipment. The sheer size of the data also raises another issue. It would be too large to share over the internet or even on the largest disposable disc media. In order to share it, the interested party would have to physically come to the centre or request that the centre send a hard disk containing the acoustic data to them.

The current PreCAST project is assessing issues around the storage of acoustic data. Their recommendations should be considered in designing the Joint Irish Cetacean Database, for implementation, if necessary, as a later phase of the project.

Biotelemetry Data

The Joint Irish Cetacean Database may also have to store biotelemetry data at some time. The data can include position, speed, time, water depth, water temperature, etc. This data can easily be linked to a specific animal in the Joint Irish Cetacean Database.

Photographic Data

A significant amount of photographic data exists from surveys of cetacean, including bottlenose dolphins, fin and humpback whales. The possibility of incorporating this photographic data to the Joint Irish Cetacean Database should be considered, even if it wasn't implemented in the initial development phase.

7. OTHER CONSIDERATIONS

In proposing a structure and design of a future Joint Irish Cetacean Database, additional issues were identified as important to consider at an early stage. These issues are identified below, and recommended course of action outlined.

- Central to having a joint database holding cetacean data from different sources is a data sharing agreement with the various organisations for joint analysis of cetacean data.

- Hosting of the Joint Irish Cetacean Database

The database should be hosted by an organisation centrally involved in cetacean recording, or biological data management. A further significant consideration is that the hosting organisation should be viewed as independent, but supportive of, the organisations providing the cetacean data.

- Frequency of updates

The frequency of updates to the database is about having the correct balance between ensuring the data is current while not placing too great a burden on the database manager. Accordingly, it is recommended that donor organisations send updates every six months to the Database. This will ensure that the Database is relatively modern but not requiring too much management from the donor organisation..

- Submission of data to the Joint Irish Cetacean Database

For publically funded surveys, it is recommended that there be a two-year timeframe between the end of the survey and submitting the data to the Joint Database. A shorter timeframe may jeopardise scientific publications of data. Agreeing a timeframe, however, is crucial for funding agencies so that data can feed effectively into conservation decisions.

- Availability of data and access rights

The assumption is made that access to raw data may need to be restricted, but as a minimum that basic sightings records should be publically available at the resolution of 5km but could be made available at a higher resolution to researchers provided specific access and usage rules are applied.

8. MAIN CONCLUSIONS OF THE SCOPING STUDY

The main aim of the Joint Irish Cetacean Database would be to bring existing and future datasets together in a common format so that an assessment could be made of the Conservation Status of cetacean species. It is likely that the Database would also be a key research tool for ecologists working in the area and may function as a means of disseminating information to the general public about these species

The merging of sightings data is a relatively straight forward process that will allow the known national range for a species to be mapped. The resolution of this will depend on the resolution of the original datasets. If all datasets are to be merged, the minimum meaningful common resolution would be 3km (as a result of the different formats of data collection).

The merging of effort data in order to calculate relative abundance of species is more technically difficult. However, the practicalities of merging data from different surveys using different methodologies should be considered.

It is recommended that Recorder is used to support the Joint Irish Cetacean Database, with a special add-in developed to cater for the specific requirements of cetaceans. An online mapping system should be developed to display the distributional data, and allow the public to view the data. The database should be managed either by the Responsible Authority or an organisation specifically involved in biological data management.

The online mapping system should provide the public user interface displaying sightings records at a resolution set at 5km. The mapping interface should have a series of layers, using GIS, which would show cetacean information against other environmental layers as well as data associated with other biological groups. The metadata should also be available for users to access to determine fitness of use of the data.

The final recommendation from this report is the development of standards for Irish cetacean data ensuring compatibility of future data. These standards should stipulate that records have a number of mandatory data items (i.e. location for sighting), that attributes such as behaviour and sea state are recorded to the same methodologies, and a standard for units of measure is followed.

References

Mackey, M., Perales i Giménez, D. and O'Cadhla, O. 2004. SEA678 data report for offshore cetacean populations. Coastal and Marine Resources Centre, Cork.

Roycroft, D., Cronin, M., Mackey, M., Ingram, S.N. and O'Cadhla, O. 2007. Risk Assessment for marine mammal and seabird populations in south-western Irish waters (R.A.M.S.S.I.). Coastal and Marine Resources Centre, Cork.

Mackey, M., Cronin, M.A. and O'Leary, D., 2004, Cetaceans and Seabirds of Ireland's Atlantic Margin. Trip Reports: April-July 2002. Annual Report. PAD, PSG, EEI/ Shell and GSI. Coastal and Marine Resources Centre,Cork,

Table 1. Survey name, data holder, database type, region, data type, number of records, contact, and other information pertaining to Irish cetacean datasets.

Survey	Data holder	Database type	Region	Data type	No. of records	Contact	Other
IWDG land-based watches (casual sightings)	IWDG	SQL	Irish coast	Sightings + effort	11,000+ sightings + 2,109 effort	Simon Berrow, Padraig Whooley	
IWDG ships surveys/ISCOPE/PreCAST	IWDG	Access and excel	Offshore & coastal Irish EEZ & EU waters	Sightings + effort	586 effort + 128 casual (6018 effort data points)	Dave Wall, Simon Berrow	
IWDG Ferry surveys	IWDG	Access and excel	Irish and UK coast	Sightings + effort	833 effort (2258 effort data points)	Dave Wall	
Strandings data	IWDG	Access	Irish coast	Strandings	1,833	Mick O'Connell, Simon Berrow	
PIP/Cetaceans and Seabirds at Sea (feeds into ESAS)	UCC/CMRC	ESAS type database (Paradox)	Offshore	Sightings + effort	772 effort	Mick Mackey (ESAS in JNCC: Tim Dunn)	
PIP acoustic surveys	UCC/CMRC		Offshore	Acoustic	671 acoustic detections	Natacha Aguilar de Soto	
Broadhaven Bay	CMRC/Shell Ireland	Access, excel	Inshore	Sightings + effort	223 effort + 59 acoustic detections	Oliver O'Cadhla (email: o.ocadhla@ucc.ie)	
RAMSSI	CMRC	Access, excel	Southwest Ireland Inshore	Sightings + effort	100+	Michelle Cronin	
SIAR survey	CMRC	Probably Access	Rockall Trough, Porcupine Bank & Shelf Slopes	Sightings + effort	116 effort + 10 casual sightings	Oliver O'Cadhla (email: o.ocadhla@ucc.ie)	
SCANS I	Sea Mammal Research Unit (SMRU)	Access based database	Celtic Sea & EU	Sightings + effort		Kelly Macleod, Emer Rogan	
SCANS II	Sea Mammal Research Unit (SMRU)	Access based database	Irish Shelf waters & Eu	Sightings + effort	Total (all areas) 481 effort & 543 aerial sightings	Kelly Macleod, Emer Rogan	

Survey	Data holder	Database type	Region	Data type	No. of records	Contact	Other
					+ 705 acoustic detections of harbour porpoise		
CODA	Sea Mammal Research Unit (SMRU)	Access based database	Deep water (200m+) west of Ireland, Porcupine Bank, Shelf slopes & EU	Sightings + effort	Total (all areas) 1500 ship sightings	Kelly Macleod, Emer Rogan	
Marine Institute – casual sightings	Marine Institute	All feed into IWDG		Sightings	----	Micheal O'Cinneide (email: Micheal.O'Cinneide@marine.ie)	
NPWS – casual sightings	NPWS	All feed into IWDG		Sightings	----	Dave Lyons (email: david_lyons@environ.ie) 021 4619902	Co-funded Coda
Seismic reports	Petroleum Affairs Division, DCENR (PAD)			Sightings		Peter Croker	All companies to forward reports to NPWS
Marine Mammal Observers data	JNCC/Petroleum Affairs Division (PAD)	Data sent to JNCC – some data from UK surveys in Irish waters held by IWDG		Sightings		Peter Croker	Now MMOs must send the data to PAD and NPWS
EHS and CEDar – various data	CEDaR	All feeds into IWDG			----	Lynne Rendle	
Historical info.	Sea Watch Foundation (SWF)	Access, excel, and paper		Sightings	Historical data - mid-70s onwards	Peter Evans	Presence/absence – also line transect data.
Ferry Surveys (Stena)	Sea Trust UK	Excel & Paper	Irish Sea	Sightings	Since 2005	Sea Trust UK – Cliff Benson (email: frederike.sjacob@virgin.net)	

Survey	Data holder	Database type	Region	Data type	No. of records	Contact	Other
France-Ireland surveys (Roscoff-Cork)	Oceanapolis, Brest/ORCA, UK		Celtic Sea	Sightings & effort	Since 2002	Sami Hassani (email: sami.hassani@oceanopolis.com) Dave Smith (d.w.smith@hotmail.co.uk)	
Arklow Bank windfarm monitoring programme	Cork Ecology/Airtricity	JNCC/ESAS format	Arklow Bank	Sightings + effort data		Clare Pollock, Cork Ecology	
Dutch Pelagic Fisheries	Wageningen-IMARES		South-west	By-catch and sightings		Bram Couperus (email bram.couperus@wur.nl)	
SOSUS Array	US Military / Cornell University		Offshore, West Coast	Acoustic		Chris Clark (email: CWC2@cornell.edu)	

Table 2. Compatibility table showing sightings structures of main databases pertaining to Irish waters (European Sea Birds At Sea (ESAS), Irish Whale and Dolphin Group (IWDG) ship-based surveys, IWDG land-based surveys and Small Cetaceans in the European Atlantic and North Sea (SCANS)) and their compatibility.

IWDG Ship Based Effort Surveys	IWDG ISCOPE Land Based / Casual	ESAS	SCANS('raw')	Compatibility Index
Date (DATETIME)	Sighting.sighted_when (DATETIME)	Origin.date_of_survey_effort (DATETIME)(Not Implied)	Date (DATETIME)	4
Record Number (not used at present but easily applied to Data)	Animal.record_id	Spec.species_key	Sightings Number	4
Species Code (Alphabetic code at present can be changed to numeric)	Species.code	Spec.euring	Species code	4
School Size(Best)	Sighting.bestest	Spec.number	Best Number	4
Time of Day	Sighting.sighted_when	pos.hours : pos.minutes	GPSTime	4
Latitude	Sighting.sighting_lat	implied	GPSIndex	4
Longitude	Sighting. sighting _long	implied	GPSIndex	4
Vessel Name & Observer Info	Observer.organisation/ platformtype.PlatformTypeName	Origin.numeric_code / Basename.base_code	Vessel	4
Meridian(implied)	implied	Meridian - implied	implied	4
Most Common Behaviour	Behaviour.behaviour_name	Spec.behaviour	Behaviour	4
Sighting.distance	Sighting.distance	Distance.distance	Radial distance	4
Number Juveniles / Calves	Sighting.juveniles	n/a	Calves	3
Speed of vessel	n/a	Implied	Implied	3
(Poskey) - n/a	Conditions.record_id	Pos.poskey	GPSIndex	3

Course of vessel	n/a	Implied	Heading	3
Number Adult	Sighting.adults	n/a	Implied	3
(Direction Travelling) - n/a	Sighting.direction_travelling	Spec.dir_ass	n/a	2
(High Number) - n/a	Sighting.maxest	n/a	HighNumber	2
(Low Number) - n/a	Sighting.minest	n/a	LowNumber	2
(Photo Record) - n/a	Animalphotosighting.record_id / Animalphotosighting.animal_id	n/a	n/a	1
(Max Length) - n/a	MaxLength(Sighting.mean_length)	n/a	n/a	1
(Min Length) - n/a	MinLength(Sighting.mean_length)	n/a	n/a	1
(Location ID) - n/a	Location.locationID	n/a	n/a	1
Second most common behaviour	n/a	n/a	n/a	1
(Form Number) - n/a	n/a	n/a	Form number	1

Note: Implied: The field is not stored explicitly but other fields exist which allow for the fields calculation. N/A: The field is not available.

Table 3. Compatibility table showing effort structures of main databases pertaining to Irish waters (European Sea Birds At Sea (ESAS), Irish Whale and Dolphin Group (IWDG) ship-based surveys, IWDG land-based surveys and Small Cetaceans in the European Atlantic and North Sea (SCANS)) and their compatibility.

IWDG Ship Based Effort Surveys	*IWDG ISCOPE Land Based / Casual*	*ESAS*	*SCANS('raw')*	*Compatibility Index*
Vessel	Observer.organisation/ platformtype.PlatformTypeName	Basename.base_name	Vessel	4
Date (DATETIME)	Sighting.sighted_when (DATETIME)	Origin.date_of_survey_effort (DATETIME)(Not Implied)	Date (DATETIME)	4
(Meridian) - implied	implied	Meridian - implied	implied	4
(Assoc Sightings) - Implied	implied	Assoc.sightings - Implied	Implied	4
(Time at Start) - Implied	Conditions.effort_start	Pos.hours : pos.minutes where pos.posmark = 'S'	Time at event	4
(Time at Middle) - Implied	n/a	Pos.hours : pos.minutes where pos. posmark = 'M'	Implied	3
(Time at End) - Implied	Conditions.effort_finish	Pos.hours : pos.minutes where pos. posmark = 'E'	Time at event	4
Record Number (not used at present but easily applied to Data)	n/a	Trip.trip_key	Effort.index	3
Implied	n/a	Pos.km_travelled	Implied	3
Latitude at start	n/a	Pos.lat,pos.posmark(S)	Latitude at event	3
(Latitude at middle) - implied	n/a	Pos.lat,pos.posmark(M)	Implied	3
Latitude at end	n/a	Pos.lat, ,pos.posmark(E)	Latitude at event	3
Longitude at start	n/a	Pos.long,pos.posmark(S)	Longitude at event	3
(Longitude at middle) - implied	n/a	Pos.long,pos.posmark(M)	Implied	3

Longitude at end	n/a	Pos.long,pos.posmark(E)	Longitude at event	3
(Poskey) - n/a	Condition..record_id	Pos.pos_key	GPSIndex	3
(Mins Watched) - n/a	Conditions.mins_watched	Calculation with pos.hours, pos.minutes, and pos.posmark	Implied	3
Speed of vessel	n/a	Implied	Implied	3
Course of vessel	n/a	Implied	Heading	3
(Count Type) - n/a	n/a	trip.count_type	n/a	1
(Transect Width) - n/a	n/a	Trip.transect_width	n/a	1
(Event) - n/a	n/a	n/a	Event	1
(Area Surveyed) - n/a	n/a	Pos.area_surveyed	n/a	1
(Form Number) - n/a	n/a	n/a	Form Number	1

Note: Implied: The field is not stored explicitly but other fields exist which allow for the fields calculation. N/A: The field is not available.

Table 4. Compatibility table showing covariate data structures of main databases pertaining to Irish waters (European Sea Birds At Sea (ESAS), Irish Whale and Dolphin Group (IWDG) ship-based surveys, IWDG land-based surveys and Small Cetaceans in the European Atlantic and North Sea (SCANS)) and their compatibility.

IWDG Ship Based Effort Surveys	*IWDG ISCOPE Land Based / Casual*	*ESAS*	*SCANS('raw')*	*Compatibility Index*
Sea state (Beaufort)	Conditions.sea_state	Pos.sea_state	BeaufortSea	4
Observer Code	Observer.record_id	Observer.observer_code	Observer	4
Visibility	Visibility.visibilty	Pos.visibility	Visibility	4
Implied from observer ID	Implied	Trip.number_of_observers	Implied	4
Swell height	Swell record_id	n/a	Swell height	3
Platform type	Platform type.record_id	Base type.base_type	n/a	3
Wind force	Beaufortscale.knots	n/a	Windspeed	3
Wind Direction	Conditions.wind_direction	n/a	Wind direction	3
n/a	Observer.organisation	Origin.numeric_code	n/a	2
Weather code	n/a	n/a	Weather code	2
Observation height(Implied from vessel ID)	n/a	n/a	Observation height(Implied from DeckHeight)	2
Cue	n/a	n/a	Cue	2
Angle/Bearing	n/a	n/a	Angle/Bearing	2
(Water Temp) - n/a	Conditions.water_temp	n/a	SeaSurfTemp	2
(Water Depth) - n/a	Conditions.water_depth	n/a	n/a	1
Cloud cover	n/a	n/a	n/a	1

Precipitation type	n/a	n/a	n/a	1
Precipitation intensity	n/a	n/a	n/a	1
(Swell Angle) - n/a	n/a	n/a	Swell angle	1
n/a	n/a	n/a	n/a	1
(Glare) - n/a	n/a	n/a	Glare	1
(Glare Direction) - n/a	n/a	n/a	Glare Direction	1
(Glare Width) - n/a	n/a	n/a	Glare Width	1

Note: Implied: The field is not stored explicitly but other fields exist which allow for the fields calculation. N/A: The field is not available.

Appendix 1

Sightings attributes

Date (4) - The date field indicates the date of the sighting and does not present any problems. It is compatible across all examined databases.

Record Number (4) - The record number is the unique identifier for the sighting. All tables have some form of record number. A possible issue could be if different organisations on the same vessel are using different record numbers for the same sighting which would lead to duplication of this sighting in the Joint Irish Cetacean Database.

Species Code (4) - All examined databases have a code identifying the species. However a problem arises if the databases are using their own unique identifier system for a species, i.e. if Seawatch identifies an Orca as species code 123, and ESAS identifies it as 124. When combining datasets it would be imperative to bring the species lists together as one and use only one identifier system.

Best Number / School Size (4) - This attribute is common to all examined databases and as it is simply a number it poses no difficulty to store.

Time of Day (4) - This is a field representing the time of day that the sighting occurred. All databases contain a field representing this data.

Latitude (4) - This is the latitude co-ordinate for the sighting record. All databases contain a field or fields either directly referencing or allowing the calculation of this value. With regards to databases such as ESAS who do not record the co-ordinate at the sighting point but rather at predetermined points along a transect, some transformation will have to be done to extrapolate the required data from the existing storage format. For example, the exact latitude will be extracted using software which will take into account the start point (lat, long), end point (lat, long), start time, end time, speed, course, sighting time etc. of the vessel.

Longitude (4) - This is the longitude co-ordinate for the sighting record. All databases contain a field or fields either directly referencing or allowing the calculation of this value. Some transformation needed for ESAS database as highlighted above.

Vessel / Platform (4) - All examined databases have an identifier for the platform or vessel from which the sighting was recorded. Once again it may require some investigation when designing the database to ensure that vessels do no suffer from the duplication issue described in the common issues section.

Meridian (4) - This field indicates the meridian or line of longitude along which a sighting record occurred. All databases contain a field representing this value or the necessary values required for the determination of this value as described in the longitude attribute above.

Most Common Behaviour (4) - This field represents the primary activity of the animal at the time the sighting was recorded. All databases examined had a field of this type, although a standard set of behaviour attributes would need to be compiled for a joint database system, with the donor's data being transformed to this new standard. This would prevent duplication. For instance consider the following example:

	SCANS	ESAS
Behaviour ID	LO	60
Behaviour	Logging/sleeping/resting	Resting or apparently asleep

In the above example we see the same behaviour being recorded using different ID's and using different descriptions. The Joint Irish Cetacean Database will need a uniform set of behaviours and indeed all other attributes.

Distance (4) - This field indicates the distance from the observer that the animal was observed. All of the databases observed had this field. A common measurement would be used in the Joint Irish Cetacean Database i.e. metres.

Number Juveniles (3) - Three of the databases examined have a field representing the number of juveniles recorded in a sighting. It is of type number and poses no difficulties.

Direction of Movement (2) - This field represents the direction the animal was travelling when the sighting was recorded. Two of the four databases examined contained this field.

Max Number (2) - This is the estimated number of animals seen in a sighting record. Two of the databases examined have a field that represents this figure. It is a number and so poses no difficulties.

Min Number (2) - This is the minimum estimated number of animals seen recorded in a sighting record. Two of the databases examined have a field that represents this figure. It is a number and so poses no difficulties.

Speed of Vessel (3) - This field indicates the speed at which the vessel, from which sightings were made, was travelling. This data was recorded explicitly by only one of the databases examined with the other two ESAS and SCANS requiring the calculation of the speed of the vessel using time and positional co-ordinates.

Position Key (poskey) (3) - The ESAS database has a table that maintains positional data while on a survey and all associated data with that position. ISCOPE and SCANS have similar tables. It will be difficult to cross reference the locations in these tables.

Course of Vessel (3) - This field indicates the course that the vessel was travelling when the sighting was recorded. Two of the examined databases recorded this data directly with the possibility of the calculation of the value existing in the third database ESAS. This however requires further research to ascertain if the values stored by ESAS can be extrapolated into a course bearing.

Number Adult (3) - Two of the databases examined have a field representing the number of adults recorded in a sighting. A third SCANS allows for its calculation. Once again this is a number and poses no problem with storage.

Description / Photo (1) - This field indicates the existence of a photo record of the sighting and also points to the photo record. Only one of the database examined, the IWDG land-based database, had this functionality. This functionality would also have size implications as there is very little comparison in storing a textual database to storing a file or picture database. The file database would be exponentially larger and standards regarding photo size and resolution would need to be adopted.

Max Length (1) – Again, this field represents the maximum estimated length of an animal recorded in a sighting record. ISCOPE's land based casual database was the only database that had a field explicitly representing this data.

Min Length (1) - This field represents the maximum estimated length of an animal recoded in a sighting record. ISCOPE's land based casual database was the only database that had a field explicitly representing this data.

Location ID (1) - This is a code referencing a specific area type. Only one database, ISCOPE, contains this field.

Second Most Common Behaviour (1) - Only one database examined contained this field. Once again a standard set of behaviours needs to be compiled and datasets with no value would simply be left at null.

Form Number (1) - The form number is only used in the SCANS database. Thus it will require a separate field for its storage which will be null for data which does not come from SCANS.

Effort Attributes

Vessel / Platform (4) - All examined databases have an identifier for the platform or vessel from which the sighting was recorded. Once again it may require some investigation when designing the database to ensure that vessels do not suffer from the duplication issue described in the common issues section.

Date (4) - The date field indicates the date of the sighting and does not present any problems. It is compatible across all examined databases.

Meridian (4) - This field indicates the meridian or line of longitude along which a sighting record occurred. All databases contain a field representing this value or the necessary values required for the determination of this value.

Associated Sightings (4) - This field represents any other animals sighted along with the sighting of the cetacean. This field was stored by one of the examined databases explicitly; however all of the examined databases stored the data to allow the calculation of this value. It is not essential information for the Joint Irish Cetacean Database and will need transformation and calculation.

Trip Key (3) - The trip key is the unique identifier for the effort trip. Three databases examined contained this key. Cross referencing would have to be done to ensure that duplication did not occur where different organisations stored the trip using different keys.

Time at Start (3) - This field indicates the time at the start of the survey. This is the beginning of effort watch data. Three of the examined databases had fields representing this data while one did not, instead favouring distance effort data. For those databases that do not store duration data which can be easily calculated using times, there is an alternative available. The duration can be calculated by calculating the distance travelled using the start end positional co-ordinates and then dividing this value by the average vessel speed. This will give a figure for the duration of a survey and thus provide time based effort data.

Time in Middle (2) - This field indicates the time at the middle of the survey. This field was present explicitly in one of the databases examined. It could be extrapolated for the other.

Time at end (3) - This field indicates the time at the end of the survey. This is the end of effort watch data. Three of the examined databases had fields representing this data while one did not, instead favouring distance effort data.

Km Travelled (3) - This field measures the km's travelled by a vessel during an effort survey. Three of the databases examined store this data explicitly or store the fields required to calculate it. As the field is a number and has a standard i.e. km it does not present any difficulty.

Latitude at Start (3) - This is the latitude co-ordinate at the starting point of the effort survey. All but one of the databases examined explicitly store this value.

Latitude at Middle (3) - This is the latitude co-ordinate at the middle point of the survey. All but one database examined stores this value explicitly or the values necessary to calculate it.

Latitude at End (3) - This is the longitude co-ordinate at the end point of the effort survey. All but one of the databases examined explicitly store this value.

Longitude at Start (3) - This is the longitude co-ordinate at the starting point of the effort survey. All but one of the databases examined explicitly store this value.

Longitude at Middle (3) - This is the longitude co-ordinate at the middle point of the survey. All but one database examined stores this value explicitly or the values necessary to calculate it.

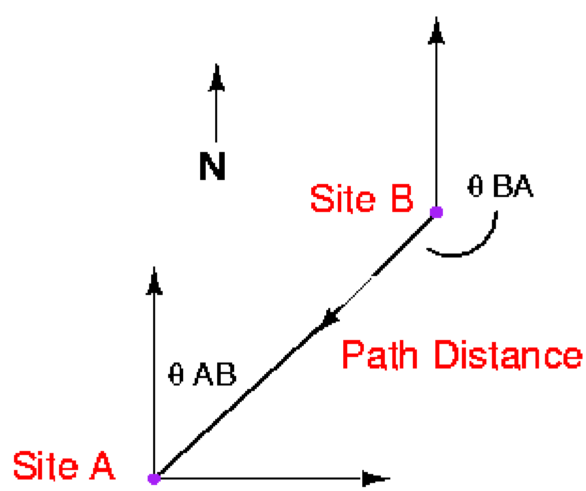
Longitude at End (3) - This is the longitude co-ordinate at the end point of the effort survey. All but one of the databases examined explicitly store this value.

Position Key (poskey) (3) - The ESAS database has a table that maintains positional data while on a survey and all associated data with that position. ISCOPE and SCANS have similar tables. It will be difficult to cross reference the locations in these tables.

Duration (3) - This field represents the time spent on the survey. Two of the examined databases stored this data or the required data to calculate it, i.e. start and end times. The minimum required data to extrapolate this field is either, start and end times or start and end location co-ordinates coupled with average vessel speed. This will allow for the calculation of the duration value. i.e. (distance/speed = time). Also a standard unit should be used e.g. minutes, hours etc.

Speed of Vessel (3) - This field indicates the speed at which the vessel, from which sightings were made, was travelling. This data was recorded directly by only one of the databases examined with two others, ESAS and SCANS, requiring the calculation of the speed of the vessel using time and positional co-ordinates.

Course of Vessel (3) - Two of the examined databases recorded this data directly with the possibility of the calculation of the value existing in the third database ESAS. The calculation can be done using trigonometry as shown below.



Count Type (1) - Once again this is a field unique to ESAS. It represents the type of count being carried out.

Transect Width (1) - Once again this is a field unique to ESAS. It represents the width of the transect in which a survey being carried out. ESAS may need this data to be stored by the Joint Irish Cetacean Database for reporting although further research is required to ascertain whether this is the case.

Area Surveyed (1) - This value is stored explicitly by only one of the examined databases, ESAS. For those databases that do not store the value explicitly, the following calculation can be performed. Calculate the distance of the line transect. Calculate the width of the area surveyed, i.e. maximum visible distance to either side of the ship multiplied by two. Multiply the distance travelled by the visible distance and this gives you area surveyed. The visible distance could also be decided upon before starting the survey.

Event (1) - This field refers to specified events along a transect, i.e. effort start, waypoint etc. This field is used by SCANS only although the field postmark used by ESAS is similar in concept but not stored in the same manner.

Form Number (1) - The form number is only used in the SCANS database.

Covariate Attributes

Sea State (4) - All databases store this value or their own representation of this value. The ideal way to store this value would be Beaufort scale measurements. A check will need to be made that every database records this in the same way.

Observer Code (4) - This field identifies the observer or observers who recorded a sighting. All databases store this data. The donor databases will need to be scanned to ensure the same observer is not entered twice into the Joint Irish Cetacean Database. This could occur if for instance John Doe is identified in the ESAS database as 123 and in the SCANS database as 124, both observers are the same but as far as a database is concerned they are not, thus this is a pitfall that will have to be overcome.

Visibility (4) - This field indicates the visibility at the time the sighting was recorded. All databases store this value explicitly.

Number of Observers (3) - This field represents the numbers observers involved in sighting or survey. Two of the examined databases explicitly store this data while one has the data required to calculate it. The minimum data required to calculate this would be a unique observer ID relating to each observer present at the time of the sighting or survey. These ID's could then be counted to give the number of observer's present value for the Joint Irish Cetacean Database.

Swell Height (3) - This field represents the height of the sea swell at the time of the sighting. Three of the examined databases store this value. Standardisation may be required e.g. cm, metres etc.

Platform Type (3) - This fields indicates the type if platform from which a sighting or survey took place. Three of the databases store this data. A standard set of platform types will have to be devised and the donor's data will need to be transformed to match this new set.

Wind Force (3) - This field represents the force of the wind at the time of the sighting. This field is recorded by three of the databases examined. Once again the Joint Irish Cetacean Database will have to adopt a standard unit of measurement for wind speed e.g. beaufort. Once the standard has been adopted then the donor databases will need to have their data transformed if necessary to fit the Joint Irish Cetacean Database database.

Wind Direction (3) - This field represents the direction of the wind at the time of the sighting. The field is stored explicitly by three of the databases examined. With regard to those that do not store it there is nothing to be done but to leave it set to null in the Joint Irish Cetacean Database.

Recording Group Code (2) - This field represents the group or organisation that the sighting was performed by. Two of the databases examined store this data. A standard set of organisations / recording groups will have to be devised and the donor's data will need to be transformed to match this new set.

Weather Code (2) - This is a field which represents a particular type of weather. Two of the databases stored this value or the numerous values allowing its calculation. A common set of weather codes will have to be adopted by the Joint Irish Cetacean Database. The donor databases then need to be cross matched with these new codes and any transformations necessary will need to be completed, to ensure uniformity in the Joint Irish Cetacean Database.

Cue (2) - This value represents the visual event/s that first caught the attention of the observer and subsequently lead to the sighting being recorded. Examples of cues would be blow, dorsal fin, body, breach, etc. Only two databases contained this data explicitly.

Angle (2) - This field represents the angle at which the observer sighted the animal. This can be important for discerning distance etc. The field is only stored explicitly by two databases.

Sea Surface Temperature (2) - This field records the surface temperature of the water at the point where the sighting was recorded. It is recorded by two of the examined databases. A standard unit of measurement would need to be adopted i.e. Celsius or Fahrenheit and donor data would then need to be transformed to the Joint Irish Cetacean Database standard units of measurement.

Water Depth (1) - This field details the depth of the water at the point where the sighting was recorded. This data is recorded by one of the examined databases.

Cloud Cover (1) - This field indicates the degree of cloud cover at the time of the sighting. One of the databases explicitly stores this field, while the others do not store it at all nor anything related to it.

Precipitation Type (1) - This field indicates the type of rainfall evident at the time of the sighting, e.g. rain, sleet, snow etc. Only one of the examined databases stores this field.

Precipitation Intensity (1) - Only one of the examined databases stores this field.

Swell Angle (1) - This value represents the angle of the sea swell at the time of the sighting.

Floating Matter (1) – This field represents the presence of fishing vessels during the recording period. This field was only recorded by one of the examined databases examined, ESAS.