SC/68C/GDR/01

Sub-committees/working group name: GDR

Databases and data holdings in the IWC: A review of the current status, issues and future plans

Iain J. Staniland, Mark Tandy, Imogen Webster, Moussa Ki Sow and Sarah Smith

Databases and data holdings in the IWC: A review of the current status, issues and future plans

Dr. Iain J. Staniland, Mark Tandy, Imogen Webster, Moussa Ki Sow and
Sarah Smith

*IWC secretariat, The Red House, 135 Station Road, Impington,
Cambridge, CB24 9NP, UK*

Introduction

Discussions at SC68B raised the issue of how data and databases were managed across the IWC. Several sub-committees discussed their desire to create new databases and requests were made for the Secretariat to take on the management of external databases. Whilst the SC research fund has been used to help fund the establishment of databases little or no resources have been set aside for ongoing maintenance and support.

Due to external factors the Global Databases and Repositories Group have been unable to meet during the last two scientific committee meetings, which coupled with staff changes within the Secretariat, has meant there has been disjointed progress in the diverse strands of IWC data management.

The IWC's database hosting architecture was reviewed in 2019 by an independent IT consultancy and it was concluded that the IWC had a large and overly complex IT infrastructure on mixed platforms that requires centralisation and simplification in order to mitigate ever-increasing development and data hosting costs and reduce the large server management burden faced by the Secretariat. The review's recommendations have now been actioned to allow the development of database projects in-house with the flexibility for external development in parallel with a standardised technical specification template that fits the architecture model if required. The IWC portal, which was heavily utilised by the Committee for SC meeting services, has been in use for over 8 years and the underlying architecture has now reached the end of its shelf life, so this has been redesigned and rebuilt in SharePoint to fit the needs of the IWC going forward. This presents an opportunity to begin centralising data held by the IWC and integrating workflows.

In the light of the extensive discussions in sub-committees relating to data reporting, collection and management, including proposals for new databases, IWC SC68B requested that the Ad Hoc Working Group on Databases and Related Issues develop an intersessional work plan to engage with the Secretariat on IWC databases and related issues. This paper aims to summarise the current status, highlight issues and stimulate discussions on establishing a more holistic approach to the capture, management and dissemination of data within the IWC.

Review of existing IWC databases

A summary of the IWC's existing databases and web applications can be seen in the following Table 1 and a brief overview of the major work areas are detailed below:

Southern Hemisphere Blue Whale Catalogue

Due to the sudden closure of the hosting account for this website in December, the database and associated applications were migrated to the Secretariat managed servers in their current form (Python application). Work has now begun to convert the application to PHP to allow for the Secretariat to maintain and administer the database going forward.

Ship Strikes database

The ship strikes database was previously managed by a contract funded through the SC research fund. However at SC68B (HIM) it was felt that this had now reached a stage where a dedicated full time co-ordinator was required. Using funds from a variety of sources including the SC and CC a new data manager has been employed, initially on a one year contract starting from the 19th April. The person will oversee the Ship Strikes database and relevant parts of the Strandings work programme. It is hoped that they will play a significant role in shaping how IWC facilitates data submission through National Progress Reports and from other non-governmental holdings.

Database of Recommendations

The Secretariat has, after extensive work, now made the DoR available online here. Historical records for Commission, SC and CC meetings are summarised in Table 2. This is still a work in progress but Resolutions and Recommendations from workshops will now be prioritised for back entry. Bulk upload via an Excel spreadsheet template has also be designed to assist data entry. It is hoped that, with use, the management of the DoR and the usefulness of the recommendations held within it will improve.

Conservation Committee

At its meeting in 2018, the CC endorsed a proposal from New Zealand to replace the current system of voluntary reporting with an integrated database. The Secretariat requested further input from the CC during its 2020 virtual meeting including on the purpose of reporting and database fields. Based on the results of these discussions the Secretariat is currently developing plans for a prototype database. An intersessional group was set up by the CC to assist in these efforts.

National Progress Reports

There has been a great deal of effort to improve the ease of data submission in recent years. Secretariat staff assisted countries in their submissions at IWC68B resulting in the submission of 17 National Progress Reports an increase from the 13 received in 2019. The Secretariat is launching a method for bulk upload for SC68C and the ability to clone data from previous years to aid data entry.

Global entanglement database

Most countries/regions already have a national or regional database and it is felt that there was currently no need for the IWC to develop its own global database and the decision to do so has been deferred.

Strandings database

The scientific committee asked the Secretariat to liaise with the strandings initiative in regard to the existing database development plans and agreed to consider this at SC68C. This work will be scoped by the newly appointed Ship Strikes and Strandings Data Manager (see above)

Catch Database

The catch database was updated this year including historical data from post war whaling in Japan. Previous discussions have been held in regard to making this publicly downloadable through a webform but to date data are only available by request from the secretariat.

Other Database Discussions

Infectious disease & pollution

The focus session on infectious diseases within the sub-committee on Environmental Concerns concluded that to fill data gaps on pathogen prevalence and disease in under-reported areas, there is a need to enhance sampling and diagnostic capabilities in those regions. They suggested a broad surveillance programme and database should be developed. The sub-committee also places value in a contaminant mapping tool which is linked via the IWC website but hosted at St-Andrews University a plan for the future of this tool is needed.

Marine Debris

Following an IWC workshop on Marine Debris the SC recommended intersessional efforts should look at data collection from post mortem examinations related to Marine Debris and how these data might be managed in the context of other IWC databases i.e., Strandings.

IWC POWER/SOWER

There are currently >160,000 images from POWER cruises between 2010-2020 and >60,000 from SOWER cruises between 1988 and 2010 held in a lightroom database. Pertinent information from natural marking / sighting / biopsy datasheets are coded into the image metadata, but not all information from the datasheets is included and not in every image has these data. However sightings data and biopsy data from these programmes are held in an outdated paradox database, paper records or on Excel spreadsheets. Money for any potential work could be made available from the cruise budget.

Future progress

The Ad Hoc Working Group on Databases and Related Issues has been requested to review the status of existing IWC databases and to liaise closely with those sub-committee members who are developing ideas for databases. These actions are critical in order to consider new database specifications in light of ongoing work in the Secretariat and the potential for interoperability with existing databases. In order to facilitate this work the following measures could be considered

- It would be useful to define what constitutes an IWC database and what rules and procedures are associated with this. For example how do we deal with external databases funded by the IWC? Do data availability expectations change when an external database is moved to become IWC database?
    - The IWC should develop a clear policy in regards to data access for its databases – particularly for internal and external databases that seek SC and CC funds.
- A workshop/meeting of this committee to scope all current database and data holdings within the IWC scoping areas of potential integration and consolidation including data submission. The Terms of reference for this would include identifying need, intended users, overlap with external data holdings, the benefit to the work of the IWC now and over the next 10 years.

- For example review the opportunity to fully integrate the CC and NPR reporting processes.
- Review the proforma and guidance regarding the development of new databases Annex i.

It is felt that the recent restructuring within the Secretariat coupled with the increasing needs of the users provides an opportunity to reassess the data management arrangements and practices within the IWC.

Table 1: Summary of IWC databases and data holdings

Database/Application	Status	Work Required	Priority *as set by Scientific Committee* *(If none set, defaults to Medium. Maintenance is always set to Highest)*
Portal	Live	Retire, Reaching End of Life (EOL) *- see below for work required* *Update the homepage with new background image*	
- Progress Reports	Live	**Improvement:** Create auto generated PDF summaries – to make the data more accessible to the public. - Not done **Maintenance:** Rebuild backend (EOL) as separate API based service at https://api.iwc.int/pr Break out public frontend from portal and host under https://data.iwc.int/pr with new unified OpenData theme Recommendations is working now https://recommendations.iwc.int Portal bulk import add from CSVhttps://portal.iwc.int/progressreports/importFromCsv)	**Improvement:** High – to finish ASAP, update 02/09/2020... **Maintenance:** Highest
- Ship Strikes	Live	**Improvement:** Bulk import tool (CSV) ' Not done Allow data to be submitted via API directly from country databases	**Improvement:** Medium (Default) **Maintenance:**

Database/Application	Status	Work Required	Priority as set by Scientific Committee *(If none set, defaults to Medium. Maintenance is always set to Highest)*
		Maintenance: Rebuild backend (EOL) as separate API based service at https://api.iwc.int/ss Create public frontend with mapping of confirmed cases and host under https://data.iwc.int/ss using new unified OpenData theme Break out private frontend from portal and host under https://dashboard.iwc.int/ss	<mark>Highest</mark>
- Research Requests	Live	**Maintenance:** Rebuild backend (EOL) as separate API based service at https://api.iwc.int/rr Break out public frontend from portal and host under https://data.iwc.int/rr with new unified OpenData theme Break out private frontend from portal and host under https://dashboard.iwc.int/rr	**Maintenance:** <mark>Highest</mark>
- Meeting Registration	Live	**Maintenance:** Rebuild backend (EOL) as separate API based service at https://api.iwc.int/meetings Break out admin frontend from portal and host admin frontend under https://dashboard.iwc.int/meetings Break out public frontend from portal and host on https://iwc.int on meeting page <mark>Using now eventleaf -</mark> https://www.eventleaf.com/iwc	**Maintenance:** <mark>Highest</mark>

Database/Application	Status	Work Required	Priority as set by Scientific Committee *(If none set, defaults to Medium. Maintenance is always set to Highest)*
- Meeting Documents	Live	**Maintenance:** SharePoint	**Maintenance:** Highest
- Meeting Schedule	Live	**Maintenance:** Eventleaf for registration, and SharePoint for agenda	**Maintenance:** Highest
- Meeting Room Entry System/Scanner	Live	**Maintenance:** Rebuild backend (EOL) as separate API based service at https://api.iwc.int/meetings Break out admin frontend from portal and host admin frontend under https://dashboard.iwc.int/meetings Using now eventleaf - https://www.eventleaf.com/iwc	**Maintenance:** Highest
IWC photographic cruise database and archive	Live	Updates only (in progress)	Medium (Default)
IWC biopsy sampling database	Under Development	Updates only (in progress)	Medium (Default)
SH Blue Whale Catalogue	Awaiting Deployment	Migrated to NeTxtra the current version. NeTxtra installation is working	
WNP gray whale catalogue	Under Consideration	Migration to new system	tbc
Document Web Archive	Live	Rebuild as API based service that better meets the needs of the IWC (see Archive discussion notes) **Other Considerations:** Eliminate PDFs in favour of plain text on website for easier reading on different devices and mobile compatibility. Also adds possibility of tracking	Medium (Default) **Other Considerations:** Requires Discussion

Database/Application	Status	Work Required	Priority as set by Scientific Committee (If none set, defaults to Medium. Maintenance is always set to Highest)
		changes for key documents. PDFs could still be generated 'on the fly' by the system. Examples: *Normal Document - https://mediawiki.iwc.int/index.php?title=Rules_of_Procedure* *Tracked Changes - https://mediawiki.iwc.int/index.php?title=Rules_of_Procedure&type=revisi on&diff=9&oldid=8*	
Bibliographic reference database (EndNote)	Live	Import commission papers and circulars?	Medium (Default)
Individual Catch Database	Live	**Improvements:** Rebuild as API based service at https://api.iwc.int/ic, admin frontend under https://dashboard.iwc.int/ic and public frontend under https://data.iwc.int/ic **Maintenance:** Document idiosyncrasies within data. Creation of a database of records for which there are no individual data and which conforms to all available summary data on the area, sex and month of these catches, to enable easy creation of catch series. Requires full documentation.	**Improvements:** Highest **Maintenance:** Highest
Catch Summary Database	Live	**Improvements:** Rebuild as API based service at https://api.iwc.int/cs, admin frontend under https://dashboard.iwc.int/cs and public frontend under https://data.iwc.int/cs	**Improvements:** Highest
Discovery Marking Data	Live		

Database/Application	Status	Work Required	Priority as set by Scientific Committee *(If none set, defaults to Medium. Maintenance is always set to Highest)*
Sightings Data (IWC-DESS)	Retired	Data to be integrated into the new integrated sightings, photo-ID, database	Medium (Default)
New integrated sightings, photo-ID, database	Under Development	Updates only (funding available)	Medium (Default)
Small Cetaceans Catches (Bycatch & Direct)	Retired		
Compendium of Whale Watching Regulations	Outdated		
Database of Recommendations	Live	Improved - https://recommendations.iwc.int	Done
Entanglement Response	Postponed	Develop database	
Cetacean Diseases of Concern Intranet	Stalled	Finalise website	Medium (Default)
WW Handbook	Live	**Maintenance:** Update to CraftCMS v3 and migrate database to PostgreSQL	**Maintenance:** Highest
Blue Whale Song Library	Awaiting Development	Develop database (£4000 available) – Assigned to tribalsystems.uk	High
Conservation Database	Under Consideration	Finalise Specification Develop database ….Update on 2/09/2020 à Review	Medium (Default)??
SC Report Drafting Tool	New Concept	Create WYSIWYG style online tool for rapporteurs to write up reports. Formatting will always be correct and could be auto-merged in to report requiring minimal input from Secretariat. Tool could have an 'Insert > Recommendation' type button which will insert a new recommendation into the drafting tool as well as adding it to the	Requires Discussion

Database/Application	Status	Work Required	Priority *as set by Scientific Committee* *(If none set, defaults to Medium. Maintenance is always set to Highest)*
		recommendations database eliminating duplicate entry requirements. This would also offer the option to lookup and/or supersede previous recommendations.	
Journal	Live Update to a new version	https://journal.iwc.int under test by publications teams	DONE
JCRM Submission Site	Live/Update august 2020	DOI published online	Medium (Default)
Informea project and IWC		IWC will join informea based On Drupal in development	

Table 2: Status of the Database of Recommendations

Meeting	Committee	Year
intersessional CC meeting	CC	2020
	SC	2020
	SC	2019
	SC	2018
IWC67	WKM&WI	2018
	ASW	2018
	Infractions	2018
	CC	2018
	F&A	2018
	Resolutions	2018
	SC	2017
IWC66	WKM&WI	2016
	CC	2016
	Resolutions	2016
	SC (partial)	2016
IWC65	Resolutions	2014
IWC64	Resolutions	2012
Marine Debris workshop	SC	2019
Indian Ocean BMI Workshop		2019
IMMAs and Ship Strikes	SC	2019
4th workshop on large whale entanglement issues	WKM	2018
SWG on E Acoustic Masking and Whale population dynamics	SC	2016
Non-Hunting Related Aspects of Cetacean Welfare	WKM	2016
Workshop to Develop Practical Guidance for the Handling of Cetacean Stranding Events	WKM	2016
Mitigation and Management of the threats Posed by Marine Debris to Cetaceans	CC	2014
Workshop on Small Cetaceans and Climate Change	SC	2010

Annex i

Database Request Proforma

This proforma is to be used for new database requests, and major alterations to existing databases.

Date	Request type	Related sub-committee
New database/alteration *(delete as appropriate)*		

1. Database title
Please provide the title of the database.

2. Brief overview of the database or alteration
Give a very brief overview on your proposal and its expected usage within the IWC community. Be succinct and clear as this may be used to summarise your request in a report.

3. Identified scope and usage within the IWC and its committees
Please explain what data the database will hold, and how the database will be used within the IWC and its committees; to what questions of importance to the IWC will this database contribute?

4. Proposed database schema or architecture
Please provide an overview of the proposed database schema or architecture. Where possible, please consider providing an entity relationship diagram.

5. Interaction with other databases
Will the database be required to interact with any other databases or applications?

6. How will the data be populated?
Please explain how the data will be populated. Please consider the following: will the data be entered by the public or a select group? Will the data be verified after entry? Will it be web or mobile accessible?

7. Other similar databases
Please list any other databases that capture similar data either within the IWC or externally.

8. Timetable for key planning activities

Activity to be undertaken	Key person(s)	Start (mm/yy)	Finish (mm/yy)

9. Proposed completion dates

Expected outputs	Completion data (mm/yy)

10. Associated people

Name	Affiliation	Role within database

11. Data ownership and sharing
Please state your expected data availability arrangements including data ownership and data sharing agreements.

12. Total costs
Please provide a breakdown of costs. These may require discussion within the IWC Secretariat.

Type	Description	Cost (GBP)
Planning costs (e.g. travel/subsistence)		
Development costs (e.g. salaries, contractors, software)		
Ongoing costs (e.g. servers, back-ups, maintenance)		
Equipment costs		
Expected ongoing data co-ordinator costs		

Expected ongoing and one-off data entry costs			
Other costs			
Total			

13. Funds and contributions

Please provide details of any existing funds or potential contributions

Type	Description	Cost (GBP)
Existing funds		
Potential contributions		
Total		

14. Have you read and agree to the IWC Database Guidelines? (please tick)

The guidelines can be found at the end of this pro forma. To ensure you have the latest version please contact the IWC Secretariat.

Yes	
No	

IWC Database Guidelines Version 2017.05

To standardise the creation of databases, repositories, catalogues and applications, the following guidelines should be adhered to.

- All new scientific database requests must be submitted to the scientific committee for discussion, approval, prioritisation and funding. This must be submitted using a database proforma **after** being discussed by the relevant sub-committee.
- All agreed development work will be overseen by the IWC Secretariat, and the Committee where specific input is needed.
- Where possible, all databases must use open source software.
- Programming languages, database engines and other technologies used must be discussed and agreed with IWC Secretariat to minimise development, infrastructure and maintenance costs.
- All completed source code and database schemas must be provided to and held by the IWC Secretariat.
- All databases will have a 5-year review cycle to ensure code and databases are kept up-to-date and secure.
- Where appropriate, new databases should be developed in a way that will allow expansion and interaction from other databases and applications. This is to be discussed with the IWC Secretariat.
- Where user accounts are required, they should be authenticated by a central IWC authentication server to minimise login credentials for users of multiple databases.