SCR Report 1

Preliminary Specifications for a Seabird Colony Database Retrieval System

J.W. Chardine & D.N. Nettleship

1987

THE CWS SEABIRD COLONY REGISTRY

Report No. 1

Preliminary Specifications for a Seabird Colony Database Retrieval System *

J.W. Chardine and D.N. Nettleship

Canadian Wildlife Service
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, N.S., Canada B2Y 4A2

November 1987

* An investigation associated with the program "Studies on northern seabirds", Seabird Research Unit, Canadian Wildlife Service, Environment Canada (Report No. 220).

ABSTRACT

A computerized system (Seabird Colony Registry, SCR) for the integration of data from surveys and censuses of colonially-breeding seabirds in Canada held by government (Canadian Wildlife Service (C&P-DOE), provincial/territorial wildlife divisions) and non-government (universities, environmental consultants, industry) organisations is described. The relational database structure and its functions are outlined, describing the input variables and how the system minimizes data redundancy and maximizes efficiency of access to information on a local, regional and national basis. Background and suggestions for implementation of the system are presented.

CONTENTS

THE CWS SEABIRD COLONY REGISTRY

Report No. 1

Preliminary Specifications for a Seabird Colony Database Retrieval System

J.W. Chardine and D.N. Nettleship

1. INTRODUCTION:

Existing data on breeding locations of seabirds in Eastern Canada are housed in files which are not easily accessible to investigators with management/scientific interests or for environmental emergencies such as oil spills. The first results of quantitative surveys of the breeding distributions of seabirds in northeastern North America were published as part of the Canadian Wildlife Service (CWS) "Atlas of Eastern Canadian Seabirds" in 1975. Those data were gathered largely through the efforts of a single CWS biologist with teams of summer students. However, the continued and increased exploitation of renewable and non-renewable natural resources over extensive areas of the continental shelf and arctic islands has resulted in a tremendous surge of interest in surveying, censusing and measuring breeding performance of seabirds during the last decade. Not only has CWS expanded its activities through the appointment of additional seabird biologists, but many other governmental and non-governmental agencies and organisations are actively involved in preparing inventories of seabird colonies in their geographical regions out of scientific interest or related to environmental assessment exercises. Thus, the immediate need is for the establishment of a centralized repository for the collection and dissemination of the enormous amounts of colony information produced by these diverse agencies and groups.

This document forms the preliminary specifications for a relational database system to store and retrieve seabird colony data derived from several files or databases in a standardized format. The Seabird Colony Registry (SCR) will archive all existing survey and census data for each breeding location and include early historical records. Most recent data are available in CWS files, but all agencies, organisations and investigators with an interest in breeding seabirds will be invited to contribute to the registry and

will have access to the data. The overall aim is to have a complete central repository for seabird colony data for northeastern North America that can be used by all interested parties, an information base that is comparable and compatible with U.S. Fish & Wildlife Service and other international registries.

The specifications given below take the form of a "user's manual" which describes the database structure, function, files and input/output parameters that make up the registry system. The basic design of the registry, contents of each database component, and composition/function of each individual input parameter, have been described in sufficient detail to allow potential users to comment on the implementation of the system. Although the internal logic of the programs has been defined, the precise layout of the files and input parameters are far from complete at this stage. Suggestions for clarification and/or revision are welcome.

2. DATABASE STRUCTURE AND FUNCTION:

A. Hardware and Software in Use

The Seabird Colony Registry is being developed on an IBM XT-286 microcomputer fitted with a 20 meg hard disk. dBase III+ is being used as the database management system allowing database design, input screen design and full retrieval and report generation. dBase III+ also includes a powerful programming language allowing many database functions to be automated. A program written in dBase III+ language is currently under developement that accepts data to be input to and retrieved from the database without the user having to be familiar with dBase III. The program also checks data before input and manages the relational database structure. Although adequate at present, we predict that the size of the hard disk currently in use will not be large enough to house the entire database once presently held data are input and ongoing work continually increases database size.

B. Design of Database

The SCR has been designed as a relational database with the emphasis on minimizing its size (by minimizing data redundancy) and maximizing efficiency of access to data.

Variables to be recorded in the database have been grouped according to origin and will be recorded in separate databases related to one another by "pointer" codes. Five

categories of variables have been identified: (1) census data, (2) location data, (3) species data, (4) observer data and (5) reference data.

The census database will be considered the MASTER file. Each record of the MASTER file will contain data for a single census of one species at a particular location and time. Codes for location, species, observer and reference will be recorded for each census, each of which will "point" to a specific record (containing the same code) in the particular database. Information about a location, species, etc., can then be recorded in the separate databases without enlarging the MASTER database. A series of small "lookup tables" will contain textual information about numeric codes used in each of the five databases. A schematic diagram representing the relational structure of the SCR is in Figure 1.

C. Inputs to Database (database contents)

The following gives details of each field (parameter/variable) contained in the five SCR databases mentioned above. The name of the field is given first in capital letters (field type and width in parentheses) followed by a brief description of the field.

MASTER database:

1. LOCATION (Character,8). Field codes for the location of the census. First 4 characters used for 1:250000 topographic map number (e.g., 065S) on which colony is located. Last 4 characters contain number assigned sequentially by database manager and uniquely identifying colony on map. Code is pointer to LOCATION database.

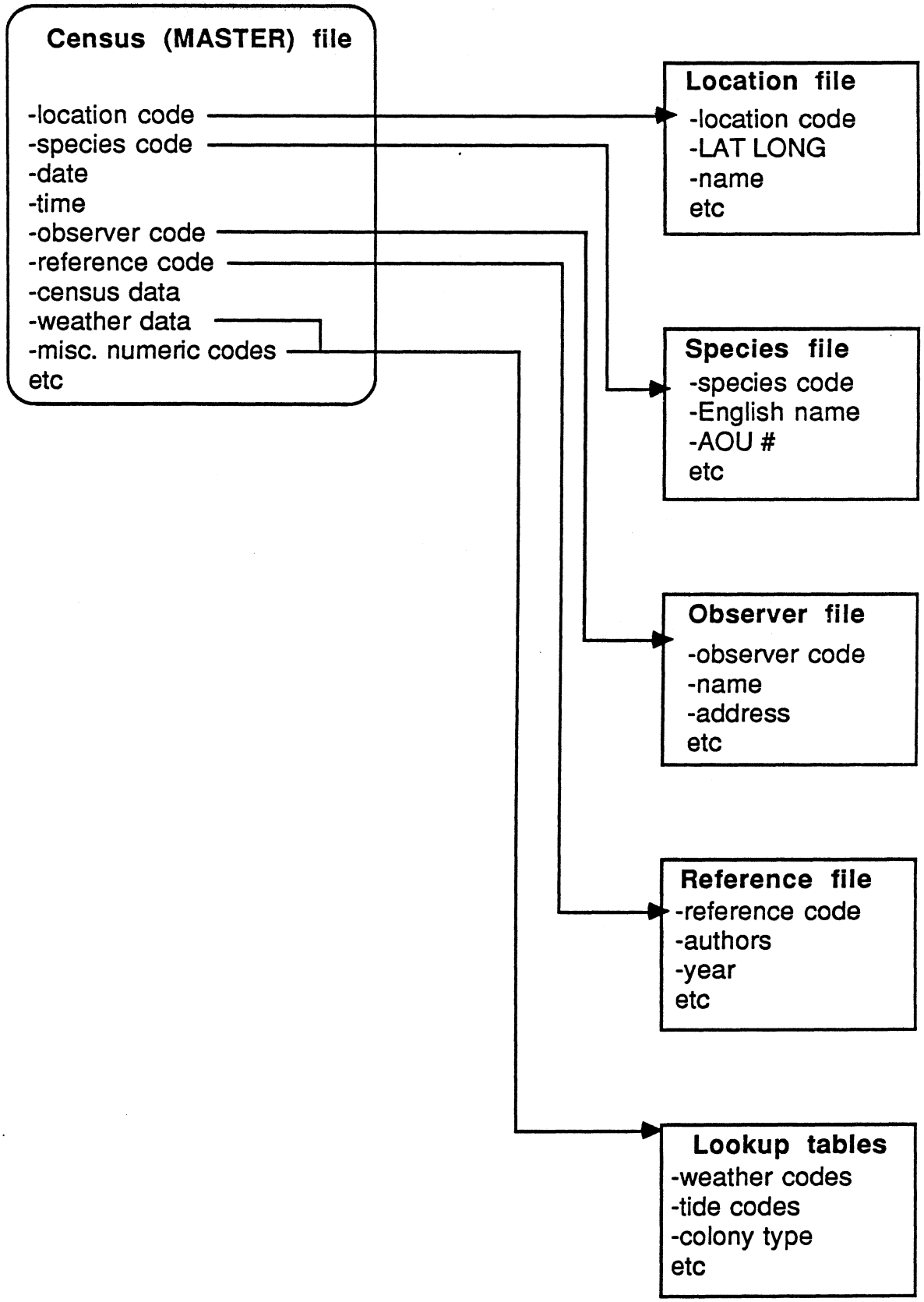
2. SPECIES (Character,3). Three letter abbreviation for species name (as used by CWS Seabird Research Unit; see Appendix 1). Code is pointer to SPECIES database.

3. COLONYSI (Numeric,4). Estimated area of colony in hectares.

4. NSP (Numeric,2). Number of other seabird species breeding in colony.

5. PHASE (Numeric,1). Code for phase of breeding cycle (i.e., pre-egg, egg-laying, incubation, etc.). Points to lookup table for description of phase codes.

FIG. 1: Relational organization for Seabird Colony Registry



6. DISTURB (Logical,1). Answer to question: "Was colony found disturbed upon arrival?" (N=not disturbed, Y=disturbed).

7. CENMETH (Numeric,2). Code for method of census. Points to lookup table for description of method codes.

8. DAY (Numeric,2). Day of census (1-31).

9. MONTH (Numeric,2). Month of census (1-12).

10. YEAR (Numeric,4). Year of census (e.g., 1987).

11. TIME (Numeric,4). Time of census: 24h clock (e.g., 1200)

12. ZONE (Character,3). Time zone for time (e.g., CDT, EST).

13. PI (Character,4). Initials of principal investigator (e.g., ABC_). If two PIs have same initials, end with number (e.g., ABC2). Points to OBSERVER file giving full details of each principal investigator.

14. OBS1 (Character,4). Initials of principal observer (e.g., ABC_). If two observers have same initials, end with number (e.g., ABC2). Points to OBSERVER file giving full details of each observer.

15. OBS2 (Character,4). Initials of second observer.

16. TEAMSIZE (Numeric,2). Total number of observers in team.

17. PHOTO (Logical,1). Answer to question: "Was colony photographed?" (Y/N).

18. SKETCH (Logical,1). Answer to question: "Was sketch map of colony made?" (Y/N).

19. REPRODAT (Logical,1). Answer to question: "Was reproductive biology data taken during census?" (Y/N). If "Yes", then: Type I census (Y/N); Type II (Y/N).

20. DATALOC (Numeric,2). Code for physical location of raw data.

21. NI (Numeric,8). Number of individual birds counted.

22. NIMIN (Numeric,8). Lower limit to NI.

23. NIMAX (Numeric,8). Upper limit to NI.

24. NIERROR (Numeric,3). Percent error of NI.

25. PRESENT (Logical,1). Answer to question: "Were individual birds present?" (Y/N).

26. NP (Numeric,8). Number of pairs counted.

27. NPMIN (Numeric,8). Lower limit to NP.

28. NPMAX (Numeric,8). Upper limit to NP.

29. NPERROR (Numeric,3). Percent error of NP.

30. PPRESENT (Logical,1). Answer to question: "Were pairs present?" (Y/N).

31. NPOP (Numeric,8). Population size (breeding pairs). [Note: to be checked or determined by database manager.]

32. NPOPMETH (Numeric,1). Code for method if determining population size. Points to lookup table for method codes.

33. GENREL (Numeric,1). Code for general reliability of census. Points to lookup table for reliability codes.

34. WEATHER (Numeric,2). Code for weather conditions during census. Points to lookup table of weather codes.

35. WIND (Numeric,2). Code for wind conditions during census. Points to lookup table of wind codes.

36. SEA (Numeric,1). Code for sea conditions during census. Points to lookup table of sea codes.

37. TIDE (Numeric,1). Code for tide conditions during census. Points to lookup table of tide codes.

38. REF (Character,50). Code for citation in which census data are published (e.g., SMITH et al. 1987a). Points to REFERENCE file giving full details of citation.

39. NOTES (Character,50). Miscellaneous notes.

LOCATION database:

1. LOCATION (Character,8). Contains location code pointed to from the MASTER database (see above).

2. MAPREF1 (Character,4). Contains map code, 1:500,000.

3. MAPREF2 (Character,4). Contains map code, 1:250,000.

4. LATDEG (Numeric,2). Latitude of location (degrees).

5. LATMIN (Numeric,2). Latitude of location (minutes).

6. LATSEC (Numeric,2). Latitude of location (seconds).

7. LONGDEG (Numeric,3). Longitude of location (degrees).

8. LONGMIN (Numeric,2). Longitude of location (minutes).

9. LONGSEC (Numeric,2). Longitude of location (seconds).

10. NAME (Character,50). Name of colony.

11. PROV (Character,4). Province in which colony is located.

12. WATER (Character,50). Name of body of marine water adjacent to colony

13. ZONE (Character,50). Geographical zone of colony.

14. COLTYPE (Numeric,2). Code for type of colony (e.g., cliff, scree, etc.). Points to lookup table containing colony type codes.

15. NOTES (Character,50). Miscellaneous notes.

SPECIES database:

1. SPECIES (Character,3). Three letter species code pointed to from MASTER database.

2. AOUNO (Character,6). AOU number for species.

3. ENAME (Character,30). AOU common name for species.

4. SNAME (Character,30). Scientific name for species.

5. FAMILY (Character,20). Family name of species.

OBSERVER database:

1. OBS/PI (Character,4). Code for observer and principal investigator pointed to from MASTER database.

2. NAMEINIT (Character,15). First name and initial(s) of observer/principal investigator.

3. SURNAME (Character,25). Surname of observer/principal investigator.

4. TITLE (Character,15). Title of observer (e.g., research scientist, biologist, professor, lecturer, etc.).

5. DEPART (Character,25). Department address.

6. INSTIT (Character,50). Institution address.

7. STREET (Character,50). Number and street.

8. CITYTOWN (Character,25). City or town.

9. PROVST (Character,25). Province or state.

10. CODE (Character,7). Postal/zip code.

11. COUNTRY (Character,25). Country.

12. NOTES (Character,50). Miscellaneous notes.

REFERENCE database:

1. REF (Character,50). Code for reference pointed to/from MASTER database.

2. AUTHORS (Character,50). Names and initials of authors in standard format.

3. YEAR (Numeric,4). Year of publication.

4. TITLE (Character,150). Title of citation.

5. JOURNAL (Character,40). Journal in which citation appears.

6. VOL (Numeric,4). Volume of journal.

7. PAGERANGE (Character,10). Page range of citation.

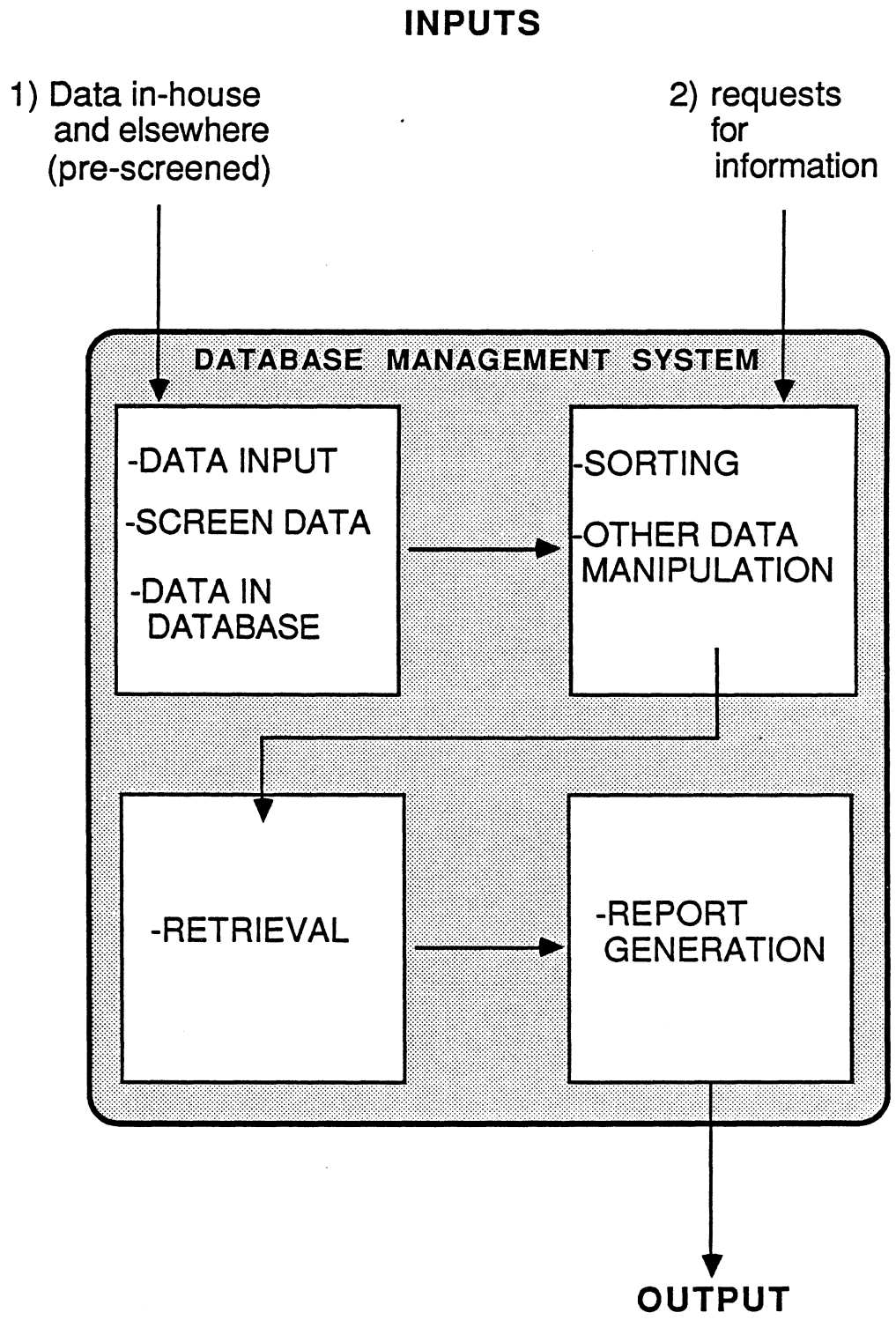
8. PUBLISHER (Character,25). If book reference, name of publisher.

9. PLACE (Character,25). If book reference, place of publication.

D. Database Function:

Despite the apparent fragmented structure of the SCR, the database will function as an integrated unit because of the relations established between each database (Figure 2). Pre-screened data (initially from in-house, but eventually supplied by ongoing studies performed by CWS and outside workers) will be in input to the database via a program. The program facilitates data input by providing custom designed screens that prompt the user for information in a well defined and systematic way. The program also checks for

FIG 2: SCR database function



errors before finally entering the data into the database automatically and also checks for internal consistency within the database (e.g., if a location is used in the census database that is not yet included in the location database, the user is prompted to enter the location data at the appropriate time).

dBase III+ will be used directly to manipulate data (e.g., sorts, calculations) and produce retrievals and reports from the database. These functions could be automated with a dBase III program and indeed we expect that many of the requests for information will be routine in nature and pertain only to a specific location (individual site or region), species or time frame. However, there will be sufficient variability in the type of information requested that it is probably not of benefit to automate this function.

The database has been designed so as not to limit the information that can be requested from it in the form of a retrieval or report. Any field or combination of fields contained in the databases can be used to produce a listing of field values for each retrieved record. For example, all Larus gull (species group) or Herring Gull (individual species) colonies censused between 1978-81 could be retrieved by accessing only the CENSUS database. These lists could be narrowed down to include only Newfoundland by selecting that province from the LOCATION database. There appears to be virtually no limit to the complexity of retrieval allowed in dBase III.

3. CONCLUSIONS AND IMPLEMENTATION

This document has presented the preliminary specifications for a seabird colony registry system to store and retrieve survey/census data in the form of a "user's" manual. To implement the system it is suggested that the following steps be undertaken:

(1) Review and revision of this document by CWS staff and other interested parties.

(2) After revisions of preliminary specifications, prepare "dummy" input modules which can then be used to test the system's design and structure characteristics. Once refinements have been made, a real dataset can be used (e.g., Lancaster Sound/Jones Sound region) to further assess data storage and return within the system. At this step, the sample database will be displayed to users within and outside government for final comment and approval.

(3) Select a number of CWS input files or databases for certain geographic regions (e.g., Labrador, Gulf St. Lawrence, PEI, NS, Newfoundland, etc.) that will be used for the first version of the system. Files or databases from nongovernmental sources can be added at a later date.

(4) Prepare the data sets selected in (3) for entry into system. Once input files are completed, enter and test storage/retrieval mechanisms of the SCR system.

(5) Continue program refinement and development.

Appendix 1: LISTING OF SPECIES ABBREVIATIONS

Common name	Species code
ATLANTIC PUFFIN	AP
ARCTIC TERN	AT
BLACK GUILLEMOT	BG
BLACK-HEADED GULL	BHG
BLACK-LEGGED KITTIWAKE	BLK
BONAPARTE'S GULL	BOG
CASPIAN TERN	CAT
COMMON MURRE	CM
COMMON TERN	CT
DOVEKIE	D
DOUBLE-CRESTED CORMORANT	DCC
GREAT AUK	GA
GREAT BLACK-BACKED GULL	GBG
GREAT CORMORANT	GC
GLAUCOUS GULL	GG
GREAT SKUA	GSK
HERRING GULL	HG
ICELAND GULL	ICG
IVORY GULL	IVG
LESSER BLACK-BACKED GULL	LBG
LAUGHING GULL	LG
LEACH'S STORM-PETREL	LSP
LONG-TAILED JAEGER	LTJ
MANX SHEARWATER	MS
NORTHERN FULMAR	NF
NORTHERN GANNET	NG
PARASITIC JAEGER	PAJ
POMARINE JAEGER	POJ
RAZORBILL	R
RING-BILLED GULL	RBG
ROSEATE TERN	RT
SABINE'S GULL	SG
THICK-BILLED MURRE	TBM
THAYER'S GULL	TG