BIRDIE

A BiodIveRsity Data pIpelinE
for wetlands and waterbirds

Report on the design and implementation of a data pipeline.

JRS Biodiversity Foundation

April 2024

BIRDIE:
A BiodIveRsity Data pIpelinE for wetlands and waterbirds

Report on the design and implementation of a data pipeline
April 2024

Prepared by: Nancy Job, Francisco Cervantes, Res Altwegg, Francis Strobbe, Michael Brooks, Sediqa Khatieb, Shae-Lynn Hendriks, Carol Poole, Emily Botts, Douglas Harebottle, Andrew Skowno, Vernon Visser, Fhatani Rhanwashe, Brenda Daly.

Prepared for:

Collaborators:

Partners:

ACRONYMS

AEWA	Agreement on the Conservation of African-Eurasian Migratory Waterbirds
CWAC	Coordinated Waterbird Counts
DFFE	Department of Forestry, Fisheries and Environment
GIS	Geographical Information System
IUCN	International Union for Conservation of Nature
NBA	National Biodiversity Assessment
RLI	Red List Index
SABAP	Southern African Bird Atlas Project
SANBI	South African National Biodiversity Institute
SDGs	Sustainable Development Goals
SoER	State of Environment Reporting
WCV	Waterbird Conservation Value

TABLE OF CONTENTS

1 INTRODUCTION

Effective monitoring and reporting mechanisms are essential to fulfil national and international commitments to secure the conservation of wetlands and waterbirds and to promote sustainable development. A key limitation reported in the latest National Biodiversity Assessment (NBA 2018) was the paucity of high-quality biodiversity field monitoring data specifically aimed at building an understanding of national trends (SANBI, 2019), including for waterbirds and wetlands. Waterbirds and wetlands play crucial roles in ecosystem health and biodiversity conservation. Waterbirds are among the groups that have declined globally and are of international conservation concern (Wetlands International, 2023), while the wetland habitats on which waterbirds depend in South Africa are among the most threatened and least protected of all ecosystem types (SANBI, 2019). These ecosystems are among the most impacted by human activity (SANBI, 2019; Albert et al., 2021) and yet they lag behind in terms of data availability, compared to terrestrial ecosystems (Cervantes et al., 2023). For example, sufficient data on water quality, or documentation of the wetland plant species present across even a small number of the roughly 700 sites represented on the BIRDIE platform was not available for analysis.

Citizen science has made a significant contribution to monitoring of waterbird observations from wetland or river systems. However, the contributions from this source are declining (Lee, pers comm). Citizen science data does have certain biases and gaps, while the raw data still need to be analysed and interrogated to bridge the disconnect between the production of data and its application into effective conservation policy in South Africa. The translation from data to action within conservation programmes and management decisions is a process that can be complex and time consuming.

The aim of the BiodIveRsity Data pIpelinE, or BIRDIE, project is to develop a data-to-decision pipeline for wetlands and waterbirds that uses statistical analyses to compute policy-relevant indicators. The data pipeline initiates from different databases and repositories where data are hosted, data then move through cleaning and validation steps, and into statistical analyses to produce indicators that are finally made available via a website (**Error! Reference source not found.**). These indicators are useful to support South Africa's national and international reporting requirements, management of wetland sites, and broader interest in wetlands and waterbirds.

When the high cost and many obstacles to establishing new biodiversity monitoring programmes are considered, it is clear that existing, robust monitoring programmes need to be maintained and fully utilised as a matter of priority. A goal of the BIRDIE project is to effectively leverage existing waterbird monitoring into South Africa's reporting instruments, and contribute to filling the current gap in wetland monitoring in the country. Investment in strategic and co-operative biodiversity monitoring programmes is essential to strengthen our ability to detect and report on trends, plan accordingly and manage effectively.

1.1 Purpose, audience and structure of the technical report

This data pipeline design document was developed and updated regularly throughout the BIRDIE project to date (September 2020 to March 2024). At each stage of the project development, various team members captured pertinent details of project progress and decisions. The version of the document presented here serves the following purposes:

1. It is a summary of the development phase of the BIRDIE project and records the foundations of the project, conceptualisation of the pipeline, the outcomes of project workshops, project methodology, and decisions made during working sessions.
2. It serves as an information manual and repository for technical details, such as statistical routines and IT architecture and specifications up until the end of March 2024.

The data pipeline design document is a resource for users of the BIRDIE data pipeline who wish to understand the technical details and data processing during the development phase. It is also useful for those who may be developing a similar data pipeline in the future. However, it is important to note that there are other documentation sources for BIRDIE that elaborate on certain aspects of the pipeline and go into more detail. For example, a journal article (Cervantes et al. 2023), as well as a "Frequently Asked Questions" and other guidance sections on the BIRDIE website itself. Those readers interested in the R code and modelling details should visit https://africabirddata.github.io/BIRDIE/ where they can find all technical details and, very importantly, the most recent code and other specifications adopted by the pipeline since March 2024.

This document is structured into sections that correspond to the data pipeline development steps for this project (Figure 1). Beginning with a record of the initial user needs, and then progressing from the data sources, through data processing and analysis, the development of the data visualisations on the dashboard and then to collecting user feedback on the working platform and platform rollout among users. Finally, the needs for long-term maintenance and future updates are considered.

Section 1: Introduction **Section 2:** User needs assessment	**Section 3:** Project datasets	**Section 4:** Data processing and analysis	**Section 5:** Website dashboard	**Section 6:** User testing and training	**Section 7:** Maintenance and upgrades

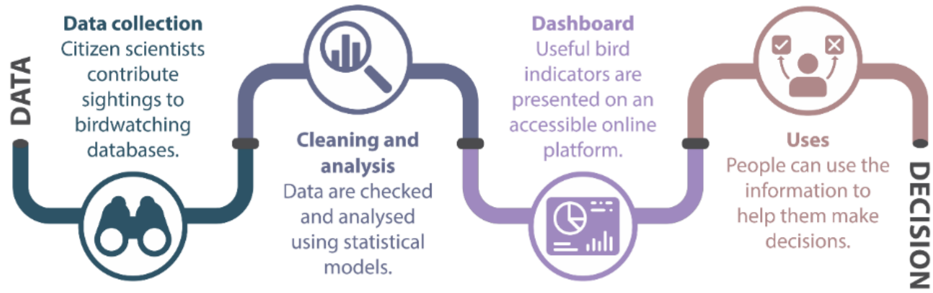


Figure 1: Structure of the report according to the data pipeline, with hyperlinks.

2 USER NEEDS ASSESSMENT

Engagement with stakeholders was maintained for the duration of the development of the BIRDIE project to ensure the data-to-decision pipeline was relevant and would provide information that meets users' needs. Concerted stakeholder engagement through a series of workshops at the beginning of the project brought focus to a way forward, and confirmed potential reporting indicators. Then, once the technical construction of the data pipeline had progressed to a workable version, stakeholders were again engaged to test and provide feedback on the prototype web pages and receive initial training on the modelled data (**7 USER TESTING AND TRAINING**), and we ended this phase of the project with a major stakeholder project representative of stakeholders from site managers through to international reporting.

The BIRDIE project established a comprehensive stakeholder list that included:

- Department of Forestry, Fisheries and Environment (DFFE) and BirdLife South Africa (**both are key partners)**
- Officials from government involved in national and international reporting, including from the DFFE, SANBI and the Department of Water and Sanitation (DWS)
- Officials from South African National Parks (SANParks)
- Provincial managers in each of the nine provinces of South Africa
- Citizen science champions contributing to the main databases
- Bird club members
- Counterparts from other African countries involved in waterbird counts

2.1 Types of users and their needs

Several workshops were held with relevant stakeholders to better understand the user needs and guide the development of relevant indicators. In brief, the majority of users are interested in waterbird population size, distribution and trend data, as well as links to potential ancillary data that can help to explain the trends. Most are looking for easily accessible processed information in the form of indicators, graphics and maps.

Based on the workshop interactions, user needs may be better understood in terms of several general groupings and their requirements, and this informed the BIRDIE project. The groups are outlined below.

2.1.1 National officials involved in reporting

The user needs assessment workshop in February 2021 drew in intelligence from important stakeholders and implementing authorities involved in meeting national and international reporting mandates.

Officials involved in species-related reporting require data on waterbird population size and abundance trends. They are looking for distilled information, graphics and maps, that are accompanied by messages and interpretation notes. To reach this goal, raw input datasets first need to be collated, cleaned, validated and analysed before feeding into national and international indicators.

There are two key waterbird-related international reporting instruments that directly require information on wetlands and waterbirds – the Convention on Wetlands of International Importance (Ramsar) and the African-Eurasian Migratory Waterbird Agreement (AEWA) (see **3.1 International International** reporting).

An additional use at national level is for those who are involved in undertaking and updating species Red List assessments (see **3.1 International International** reporting). These assessments contribute to the National Biodiversity Assessment and consequently towards various national and international reporting requirements. In addition, similar information may also have relevance to reporting into other multilateral environmental agreements, such as the Convention on Biological Diversity (CBD), the Bonn Convention on Conservation of Migratory Species of Wild Animals (CMS), and the Sustainable Development Goals (SDGs).

In terms of ecosystems (wetlands), officials ideally require wetland as well as species information to be relatable and available in one place, with the option to view them together, spatially.

2.1.2 Provincial officials involved in wetland management

Provincial and municipal conservation authority officials were well represented at stakeholder workshops. These officials are responsible for site-based management, regulation and planning. Provincial officials need to make decisions about the management of wetland sites, as well as on regulated activities such as hunting quotas. Throughout the project, they offered crucial guidance on what information is needed to inform practical conservation management questions and to address current challenges.

Provincial managers are looking for information at site and sub-national level related to waterbird population size and distribution. They are particularly interested in tracking trends over time and linking these to potential drivers. This will allow them to understand both the short-term impact of events such as sewage spills, but also long-term trends related to human impact and climate change. For this reason, they proposed that ancillary information related to water, climate and land cover change would also assist them in their work. They require easily accessible and well-structured indicators in graphical and tabular format, as well as maps. It became clear that site-based data from citizen science databases is less accessible as disparate local information than when it is organised into summarised national statistics, so, to meet the needs of provincial officials, it would require additional data cleaning and analysis to provide useful indicators.

2.1.3 Researchers

Researchers seek wetland and waterbird data to answer a wide range of potential research questions. Researchers require access to raw data so that they can undertake analyses themselves. They are looking for easy access to citizen science data that has been cleaned and validated. Researchers need data on population size and trends in tabular outputs that can be imported into statistical or analysis platforms. They also want data on species distribution in appropriate spatial formats to analyse in GIS. An additional support to researchers would be interfaces to facilitate access to citizen science data, as well as appropriate functions and models in R statistical code for analysing such data to produce occupancy and abundance models, confidence statistics, change statistics and more.

2.1.4 Birdwatchers and tour guides

Enthusiastic birders are the contributors to citizen science datasets that make the information that powers BIRDIE available. They are interested in seeing the results of their contributions to citizen science used in meaningful ways. Birdwatchers and tour guides are interested in information on recent observations at specific sites, sightings of rare species and species richness. They use this information to plan birdwatching trips. They prefer easy online access.

2.1.5 Barberspan Nature Reserve

A second workshop, specifically with managers and stakeholders associated with Barberspan Nature Reserve, was held in July 2021. The aim of the workshop was to focus on one site to delve more deeply into site-specific user needs for the data pipeline. Barberspan Nature Reserve is a large body of water located in North West Province. It falls within a nationally Important Birding Area, and it is also a Ramsar Wetland of International Importance for migratory birds and waterfowl. Over 365 species have been recorded at the site, including a number of rare migrants.

During the workshop, the following potential uses of the data pipeline were identified, to support both local management and assist users to comply with reporting processes at various scales:

- Monitoring waterbird population trends at different spatial scales (national, provincial and site-specific)
- Developing the potential of a waterbird indicator for assessing wetland health, for example the development and potential application of waterbird feeding guilds and their link to water quality and wetland physical habitat.

The Barberspan Nature Reserve site workshop revealed how complicated on-site questions could be. For example, there can be a positive relationship between pollutants and bird abundance in some sites, which may complicate the manager’s needs to strengthen motivations for improved management of water quality. In this, as well as subsequent meetings, discussions were increasingly focussed on which reporting questions the available data and the BIRDIE project could feasibly answer.

Species present at the Barberspan site were subsequently used as a species subset to develop the initial statistical routines for the BIRDIE data pipeline (**5.1**

Step 2: Running statistical routines to produce indicators). Estimates of population sizes and trends for all species recorded at Barberspan were produced since the beginning of the CWAC programme in 1993 until 2020. In addition, the occupancy of these species at a national level was estimated between 2008 and 2011. In total, 106 species were analysed, one of which is Endangered (Cape Cormorant), one is Vulnerable (Maccoa Duck), and six are Near Threatened at an international level, according to IUCN criteria. Trends of all waterbirds at Barberspan Nature Reserve were assessed using state-space models to examine how various indicators work for this case.

3 REPORTING AND INDICATORS

The two user need workshops helped underscore that indicators are needed for a variety of uses. Figure 2 illustrates the broad range of user needs gathered from the two workshops which span different spatial and temporal resolutions and are linked to different reporting needs.

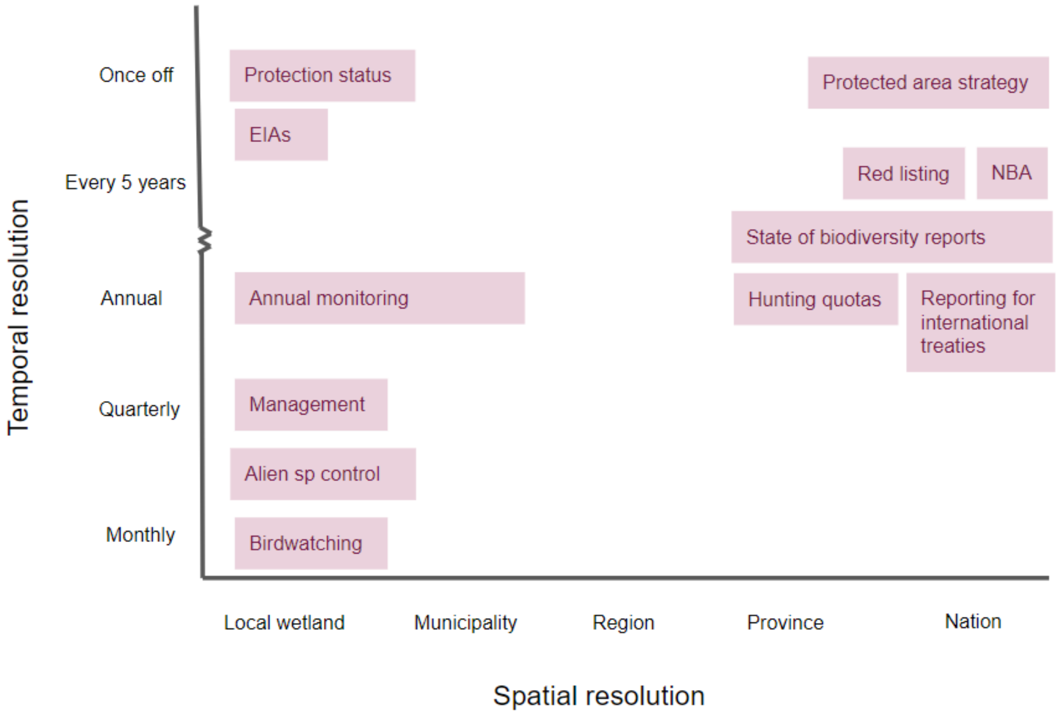


Figure 2: The broad range of user needs at different spatial and temporal resolutions, linked to their reporting needs, based on two user need workshops at the project outset.

3.1 International reporting

South Africa has a number of waterbird-related reporting requirements. South Africa is a party to several multi-lateral environmental agreements related to wetlands and waterbirds. The country has reporting obligations under these agreements that require relevant and current evidence. Waterbird population sizes and trends are important reporting parameters for these international agreements.

3.1.1 Convention on Wetlands of International Importance (RAMSAR)

The Convention on Wetlands (Ramsar, Iran, 1971) is an intergovernmental treaty with the vision: "Wetlands are conserved, wisely used, restored and their benefits are recognized and valued by all" and mission: "the conservation and wise use of all wetlands through local, regional and national actions and international cooperation, as a contribution towards achieving sustainable development throughout the world". South Africa became the fifth contracting party to the Convention in 1975, with the declaration of the country's first Ramsar site at Barberspan. Since becoming a signatory to the Convention, South Africa has designated 30 wetlands (as of 2024) to the Ramsar List of Wetlands of International Importance.

Article 2 of the Convention lists nine criteria by which wetlands of international importance may be designated. A wetland should be considered internationally important if:

1. It contains a representative, rare, or unique example of a natural or near-natural wetland type found within the appropriate biogeographic region.
2. It supports Vulnerable, Endangered, or Critically Endangered species or threatened ecological communities.
3. It supports populations of plant and/or animal species important for maintaining the biological diversity of a particular biogeographic region.
4. It supports plant and/or animal species at a critical stage in their life cycles, or provides refuge during adverse conditions.
5. It regularly supports 20 000 or more waterbirds.
6. It regularly supports 1% of the individuals in a population of one species or subspecies of waterbird.
7. It supports a significant proportion of indigenous fish subspecies, species or families, life-history stages, species interactions and/or populations that are representative of wetland benefits and/or values and thereby contributes to global biological diversity.
8. It is an important source of food for fishes, spawning ground, nursery and/or migration path on which fish stocks, either within the wetland or elsewhere, depend.
9. It regularly supports 1% of the individuals in a population of one species or subspecies of wetland-dependent non-avian animal species.

Many of South Africa's Ramsar wetland sites are in need of updated baseline information and do not have a monitoring programme in place. This lack of information was highlighted by the Department of Forestry, Fisheries and the Environment (DFFE) and the Water Research Commission in 2017 as a priority research area (Malherbe et al. 2017). The DFFE has indicated future plans to support the development of management effectiveness measures across

South African Ramsar sites and it is critical that these sites are monitored and that management plans are actively implemented.

Officials involved in the species-related reporting require data on waterbird population size and abundance trends. Indicators need to correspond to the reporting metrics of the conventions, some of which require information on sites, specific waterbird species, or national level summary statistics. In particular, information was needed towards the following metrics:

- Species present at Ramsar sites, including Vulnerable, Endangered or Critically Endangered species or threatened ecological communities (Criterion 2).
- Species at critical stages of their life cycles, or refuge during adverse conditions (Criterion 4).
- Total population size of waterbirds (Criterion 5) or as a percentage of global population for a single species (Criterion 6).

3.1.2 Agreement on the Conservation of African-Eurasian Migratory Waterbirds (AEWA)

The African-Eurasian Migratory Waterbird Agreement (AEWA) is an international conservation treaty between European and African governments that protects migratory waterbird species and their habitats along the important Africa–Eurasia flyway. AEWA covers 255 species of migratory waterbirds that are ecologically dependent on wetlands for at least part of their annual cycle, which cross international boundaries during their migration and require good quality habitat for breeding, as well as a network of suitable sites to support their annual journeys. The Agreement classifies each waterbird species according to criteria concerning a population's size, range, habitat, and trends, as well as species' migratory status, and their status on the IUCN Red List of Threatened Species. These classifications have implications for Parties' obligations in respect of each population. In particular, those obligations concerning the provision of legal protection and the development and implementation of species action plans.

AEWA has extensive reporting needs requiring multiple institutions and reporting instruments. However, a key area of overlap with Ramsar, and for which the BIRDIE project can contribute were identified to be:

- Change in migratory bird populations over time.
- Identification of preferred sites for species where a percentage of the migratory population occurs at a particular site.

3.1.3 Red List Index (RLI) for Waterbirds of South Africa

Red Listing concerns the identification of threatened spec
ies using recognised categories and criteria of the IUCN Red List of Threatened Species. In South Africa, threatened species information is used widely in conservation planning and environmental screening processes which directly impact the outcomes of reactive land use decision making and influence strategic land-use planning. The Red List Index measures change in aggregate extinction risk across groups of species. It is based on genuine changes

in the number of species in each category of extinction risk. BirdLife South Africa leads this process for South African bird species, with support from SANBI. Computing the Red List Index requires repeat assessments, and currently BirdLife South Africa is undertaking the third national Red List for birds.

The red listing process calls for information on species occupancy and trends in abundance.

3.2 National and Provincial reporting

The Red List Index is a crucial indicator in other multilateral environmental agreements, including the UN Sustainable Development Goals (15.5.1), the Global Biodiversity Framework of the UN Convention on Biological Diversity (Headline Indicator for Goal A) and the UN Convention to Combat Desertification. It is also considered a headline indicator in South Africa's National Biodiversity Assessment.

3.3 Preliminary Indicators for BIRDIE

Responding to the user needs, and following a workshop in February 2022, a set of preliminary indicators were identified, for both national and international reporting, as well as for local and ad-hoc users (Table 1). These indicators were interrogated for feasibility during the BIRDIE project, namely, if they would be feasible to implement out of the available citizen science waterbird data (**4 PROJECT DATASETS**). This was confirmed through setting up the initial model for a subset of data relevant to the Barberspan pilot site, and a subset of indicators were adopted to take forward. Computing these indicators at the smallest scale possible, which is for each species at the site level, and then aggregating these into broader scales, ensures that all indicators are provided at the necessary scales. It was proposed that information could be aggregated to site types, species groups, or nationwide.

Table 1: Preliminary indicators identified to meet user needs and that are feasible to achieve from the available citizen science datasets.

Indicator	Domain	Extent	Period	User
Occupancy	Taxonomic	National	Annual	Site management
Extent of occurrence	Taxonomic	National	Annual	Reporting, Red Listing
Area of occupancy	Taxonomic	National	Annual	Reporting, Red Listing
Change in area of occupancy	Taxonomic	National	Annual, 15 year, 30-40 year	Reporting, Red Listing
Abundance	Taxonomic, geographic	National, regional, site	Annual, monthly	Reporting, Red Listing, site management
Abundance in winter	Taxonomic, geographic	National, regional, site	Annual, monthly	Reporting, site management
Abundance in summer	Taxonomic, geographic	National, regional, site	Annual, monthly	Reporting, site management
Population rate of change	Taxonomic	National, regional, site	Annual, 15 year, 30-40 year	Reporting, Red Listing, site management

Population rate of change in winter	Taxonomic	National, regional, site	Annual, 15 year, 30-40 year	Reporting, site management
Population rate of change in winter	Taxonomic	National, regional, site	Annual, 15 year, 30-40 year	Reporting, site management
Population proportion	Geographic	National, site	Monthly	Reporting
Waterbird Conservation Value	Geographic	Site	Monthy	Reporting
Living Planet Index				
Number of occurrence sites	Taxonomic	National	Annual	Reporting
Hydrological Drought Index	Geographic	Regional, site	Annual, monthly	Site management
WetHealth Score	Geographic	Site	Static	Site management
Surface Water Extent	Geographic	Site	Annual, monthly	Site management
Surface Water Recurrence	Geographic	Site	Annual, monthly	Site management

4 PROJECT DATASETS

South Africa needs to monitor its biodiversity to inform decisions that affect the environment, and to fulfil reporting obligations linked to international conventions. Due to South Africa's size and exceptional biological diversity, monitoring is a challenge. Fortunately, there are several long-running citizen science projects that collect waterbird data at wetlands (Barnard et al 2017). At its core, the project leverages two bird-related datasets: the Co-ordinated Waterbird Counts (CWAC) and the African Bird Atlas Project (ABAP, Brooks et al 2022).

The pipeline provides four main types of indicators that are estimated from different datasets. There is some information flowing between indicator types and datasets but we can roughly associate:

- Species **abundance indicators** with the Co-ordinated Water Counts project (CWAC)
- Species **distribution indicators** with the African Bird Atlas Project (ABAP)
- Species **diversity indicators** combine information from ABAP and CWAC.
- Ancillary **environmental indicators** with the Google Earth Engine repository of environmental layers, as well as the National Wetland Map 6 with associated wetland condition information.

4.1 Citizen science datasets

Large, long-term data sets are immensely valuable and establish a convincing evidence base for action now and into the future. The power of citizen science data has not yet been adequately harnessed. In the words of Barnard and de Villiers (2012), "time is short to build the tools we need to scan the horizon for trends in biodiversity. We have most of the right ingredients, but need to combine them in the right way, supported by sound science and carefully applied statistical analysis. South Africa is in a good position with excellent large-scale, long-term biodiversity databases". These long-term and large-scale spatial biodiversity databases may be built into a national early warning system as a central support tool for South Africa's policy, planning, management and reporting needs.

Large-scale citizen science projects, such as atlases of species distribution, are an important source of data for macroecological research, for understanding the effects of climate change and other drivers on biodiversity, and for more applied conservation tasks, such as early-warning systems for biodiversity loss (Barnard et al., 2017). Enlisting citizen scientists in monitoring offers a dual advantage. First, the increased number of observers allows for more data to be collected from more places than surveys carried out by professional scientists could. Second, by engaging ordinary citizens, there is an opportunity to turn these people into ambassadors for biodiversity as they gain first-hand information on the biodiversity around them.

However, citizen science data are challenging to analyse for two main reasons (Yoccoz et al., 2001, Altwegg & Nichols, 2019, Johnston et al., 2022):

- Sampling usually does not happen in a probabilistic way (Hugo & Altwegg, 2017). Instead, citizen scientists prefer to visit areas that are easily accessible (e.g. close to roads and human settlements) or interesting for some reason (e.g. protected areas).
- Citizen scientists do not detect all species that are present in the area they survey. Imperfect detection is also an issue with almost all other types of surveys, but in citizen science projects, the observation process is typically more variable as the observers vary in their level of skills and the effort they are prepared to put into a survey (Altwegg & Nichols, 2019).

The biological process one is interested in is confounded with the observation process. Analysis must therefore account for the observation process so that the biological signal can be extracted from the data. Properly analysed, however, citizen science data are an invaluable source of information about large-scale biodiversity trends (Devictor et al., 2010, Isaac et al., 2014).

Many environmental decisions are based on lists of species that have been encountered in a particular location. For example, environmental impact studies, decisions about the level of protection that should be afforded to a particular area, etc., often rely on such lists. The problem with this approach is that it ignores the fact that we have incomplete information on biodiversity at any given location. It is unknown whether a short species list reflects low biodiversity or low sampling effort at a particular site. Robust environmental decisions therefore need to be based on monitoring protocols that account for the observation process.

4.2 Co-ordinated Waterbird Counts (CWAC)

The Co-ordinated Waterbird Counts (CWAC; https://cwac.birdmap.africa/) dataset consists of counts of all wetland-related bird species at sites (wetlands, dams, estuaries) across South

Africa. CWAC data are specific to a site, i.e. a species record shows that the species occurs at the wetland in question. Observers count all individuals they encounter at the site twice per year. Waterbird species have diverse habitat requirements and life histories. Some use the same wetlands year-round, whereas others are migratory. To capture this diversity, CWAC counts are carried out twice per year: once in mid-summer and once in mid-winter.

The CWAC has accumulated a long time-series for many wetlands. Bird abundance has been recorded with waterbird counts taken twice a year at 731 sites across South Africa since 1992. However, not all wetlands have been monitored since the start of the project and some regions are better represented than others. There are also gaps in these time series, which means there may not be a count for every year and every season at a particular site. However, when a count was conducted, it was complete and all species were counted.

While observers have aimed to count all individual birds that were present, it is possible that some individuals were missed or double counted. These data are noisy because it is very difficult to count waterbirds precisely, but with appropriate statistical analyses (see **4.3 Statistical routines**), they can reveal long-term temporal trends and seasonal fluctuations in all waterbird species across many different types of wetlands.

At the time of writing, the statistical routines of BIRDIE assume that the counts are made under the CWAC protocol, with one count in mid-summer and one count in mid-winter. In the future, these models could be expanded to be able to incorporate provincial or individual information outside of the CWAC data infrastructure, such as more frequent counts.

4.3 African Bird Atlas Project (ABAP)

The African Bird Atlas Project consists of detection / non-detection data, rather than counts. It can be used to infer where a species occurs but does not directly contain information on abundance. The BIRDIE analysis is currently restricted to South Africa, and therefore the Southern African Bird Atlas Project (SABAP2) component of ABAP. However, in the future it would be possible to expand functionality to cover other countries contributing data to ABAP, such as Kenya or Nigeria. The main adaptation that would be necessary is to exchange covariates that are specific to South Africa with the local equivalents. Many of the covariates we used in the models are based on global data sets, however, and should require little adaptation.

The southern African version of ABAP, SABAP2, started in 2007, is still ongoing and is implemented at the scale of pentads (5’ x 5’ minute grid). SABAP2 had nearly 12 million records by 2018, and collects more than 1 million new records per year. SABAP2 has a predecessor, SABAP1, which was carried out between 1987 and 1992. SABAP1 used quarter-degree grid cells as spatial units and the protocol differed from SABAP2 in a few other ways. BIRDIE currently does not make use of SABAP1 but could do so in the future.

ABAP data are collected by citizen scientists. Volunteers collect checklists of all birds observed within specific grid cells over a grid of pentads covering different African countries. A full protocol list consists of a complete list of all species that were encountered and identified over at least two hours of intense birding, and covering up to five days. Observers are asked to cover a grid cell as thoroughly as possible by attempting to visit all habitats within the grid cell. They are also asked not to list any species they have not been able to identify with certainty. It is not recorded which areas within a grid cell were visited by an observer and species are not necessarily recorded at the wetland in question, but an ABAP record shows that the species occurs at least in the vicinity of wetlands within the grid cell.

Checklists are vetted and unexpected sightings trigger a request for more specific information from the observer. This limits false positive records in the dataset, in other words reports of species that were not actually encountered. On the other hand, there are a lot of false negatives in these data that happen when a species is not observed even though it was present. Our statistical modelling approach is designed to account for such false absences.

The spatial and temporal extent of ABAP allows us to examine how bird distributions are changing over the years. The rigorous protocol of this project makes it possible to account for imperfect detection, and obtain estimates of the true distributions and how they change over time (see **5.2.2 Species distributions**). For example, these data have been used to test the effectiveness of protected areas (Duckworth et al., 2018), and to monitor bird migration (Bussiere et al., 2015). But perhaps more importantly, ongoing data collection can be used for monitoring bird distributions "in real time".

4.4 Ancillary data

The BIRDIE project makes species (waterbird) information accessible together with information about the ecosystems that support them (wetlands).

Information wetland type, area and condition derived from the National Wetland Map (SANBI 2023) was drawn into the Sites and Ramsar reporting tables (see **4.5.2**).

In addition to the National Wetland Map data, the waterbird-related data is combined with other informant datasets around themes of climate, surface water and land cover. The purpose of the ancillary datasets is two-fold. They:

- Support the statistical routines (**5.1**
- **Step 2: Running statistical routines to produce** indicators) to develop links between waterbird data and land cover, climate change, and surface water extent.
- Are made directly available on the web platform (**6 DASHBOARD WEBSITE**) to support users who might use them as background layers for deeper interpretation of the indicators and site information.

There are potentially a very wide range of additional environmental data that could be used to complement and explain trends in BIRDIE indicators. However, it is not possible to host an unlimited number of additional datasets, so the selection of additional information had to be prioritised to that most useful and accessible for modelling waterbird distribution and abundance at a national scale (Table 2). The options were explored during a series of working sessions with the BIRDIE team and other experts. It was agreed that the BIRDIE project would focus mainly on core waterbird-related indicators that are not available through other well-established channels, but help connect the BIRDIE dashboards to other platforms that are already tracking applicable environmental variables, such as surface water extent, land cover change, drought indices and others.

Most additional layers are accessed through Google Earth Engine (GEE) data catalog (https://developers.google.com/earth-engine/datasets). They then needed to be processed to bring them to a common resolution. For example, to model species distributions BIRDIE uses the ABAP pentad as the smallest spatial unit, but some of the environmental layers come at a resolution of 1km x 1km or even 30m x 30m. Therefore, these layers with higher resolution need to be summarised taking the mean value per pentad.

Table 2: Ancillary datasets used in the modelling of waterbird distribution.

Ancillary dataset	Description
Average maximum temperature	Monthly temperature and precipitation from the Terraclimate dataset. TerraClimate is a dataset with high-spatial resolution (~4km) providing monthly climate information from 1958 – ongoing. https://www.climatologylab.org/terraclimate.html
Average minimum temperature	
Precipitation	
Normalized Difference Vegetation Index (NDVI)	MODIS Vegetation Index products are based on satellite imagery. NDVI quantifies the amount of live green vegetation to measure plant health and density. https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13A2
Surface Water Extent	The occurrence of surface water dataset (Global Surface Water Explorer) provided by the European's Commission Joint Research Centre provides the frequency with which water was present on the surface from March 1984 to December 2020. https://global-surface-water.appspot.com/
Surface Water Recurrence	
Elevation	The MERIT 3 arc second Digital Elevation Model https://developers.google.com/earth-engine/datasets/catalog/MERIT_DEM_v1_0_3
Distance to coast	Computed in R using the package *rnaturalearth* (Massicotte P, and South 2023 and https://www.naturalearthdata.com/)
Human population density - WorldPop Global Project Population Data	Estimated Residential Population per 100x100m Grid Square https://developers.google.com/earth-engine/datasets/catalog/WorldPop_GP_100m_pop

4.5 Master lists

The BIRDIE project developed two master lists – one for waterbird species and one for wetland sites. These master lists were used to set the scope of the BIRDIE project based on available data and to organise information into suitable categories or groupings.

The master lists provide static, contextual information about birds and wetlands that is drawn into the tables on the BIRDIE Sites and Species webpages. The master lists are defined by the BIRDIE experts and draw information from recognised sources, such as the IUCN conservation status of species. They are used throughout the website and map viewer as the reference tables for all the selections that are possible. The master lists are also made available for direct download from the Resources page on the BIRDIE website:

https://biodiversityadvisor.sanbi.org/contentmanagement/index?guid=e22e53ba-172d-4f7e-b52d-9e511ce5d63a

4.5.1 Bird master list

The bird master list included 143 indigenous South African bird species that have an association with water or wetlands. These are species that are encountered at wetland during CWAC counts. For each bird species, static information was gathered about their taxonomy and life history (Table 3).

Table 3: Attributes collected in the bird master list of the BIRDIE project.

Attribute	Description
Common_group	The common grouping for the species, e.g. Duck or Stork. This assists users in filtering lists according to common groups.
Common_species	The specific part of the species common name, e.g. African Black. This assists users in selecting a specific species after filtering by common group.
Common_name	The complete common name for the species, combining the common species and common group in order.
Genus	The recognised genus. The first term in the scientific name.
Species	The recognised species. The second term in the scientific name.
Family	The taxonomic family that the species belongs to.
Order	The taxonomic order that the species belongs to.
Species_group_guild	Grouping of species according to common groups of species, e.g. Ducks and geese. This allows information to be aggregated by types of birds.
Foraging_habitat_guild	Grouping of species according to their foraging habitat, e.g. Open water. This allows information to be aggregated by the type of habitats that species prefer.
Feeding_guild	Grouping of species according to their feeding habits, e.g. Herbivore. This allows information to be aggregated by how species feed.
Migrant type	Grouping of species according to their migrant status, e.g. Nomadic. This allows migrating species to be selected for analysis.
Status	The conservation status of the species based on the latest available IUCN Red List of Threatened Species assessment.
WetlandIntCode	The code assigned to the species by Wetlands International. The code is used to link information to the 1% thresholds provided by Wetlands International.
#Ref	A unique reference number.
Assess for AEWA	Whether the species is assessed for the AEWA agreement. (Yes or No).

Two new guild groupings were developed for the BIRDIE project and used in the bird master list. Guilds are groupings of species based on shared or similar traits such as diet, migratory status and/or taxonomy/morphology. Guilds made it possible to group bird species according to their preferred habitat or feeding habits to better analyse data at aggregate levels.

The new guilds had the following categories:

- Foraging habitat guild
    - Marsh
    - Open water
    - Open water swimmer
    - Pelagic
    - Reedbed specialist
    - Riparian
    - Shallow-deep water
    - Shoreline and mudflats
    - Shoreline-shallow water
    - Vegetated riparian edge
    - Wetland edges
- Feeding guild
    - Herbivore
    - Invertebrate feeder
    - Piscivore
    - Predator

4.5.2 Wetland sites master list

The wetland site master list contains 727 wetland sites where CWAC counts have been conducted one or more times. In some cases, this list contain multiple sites at the same wetland (such as different locations or vantage points from which counts are done), however, they have been assigned an additional common ID number so that data for these sites can also be summed for an overall analysis of the site. Further improvements in grouping this data were recommended in the final project workshop and have been identified as a future project task. For each wetland site, static information is gathered about its size, location, protection status and type (Table 4). Some static info derived from the National Wetland Map (SANBI, 2023) was manually carried over to the sites master list for inclusion into the web tables for the Sites and Ramsar reporting tables.

Table 4: Attributes collected in the wetland sites master list of the BIRDIE project.

Attribute	Description
Province	The province that the wetland is located in.
CWAC site#	The unique CWAC site number.
Ramsar site	Whether the site is a recognised Ramsar wetland (Yes or No).
ICA code	The unique code for International Bird and Biodiversity Areas, if relevant.
Management	The responsible management of the site, including both private and government.
Protection status	Whether the site is protected, partly protected, or not protected.
Protection type	If the site is protected, what type of protected area is it, e.g. Nature Reserve.
Overall type	The overall type, either coastal, dam, wetland, estuary, river, saltpan or wastewater treatment.

Wetland ecosystem type	The ecosystem type of the wetland according to ecosystem classification.
Ha	Size of the wetland in hectares.
Condition	Condition of the wetland according to the standard condition categories of the Department of Water and Sanitation (A – F).
Fresh or saline	Whether the wetland is freshwater or saline.
Hydroperiod	The period that the wetland is inundated.
Pressures	A list of pressures on the wetland.

5 DATA PROCESSING AND ANALYSIS

The core objective of the BIRDIE project is to process and analyse citizen science waterbird datasets (**4.1 Citizen science datasets**) and provide a set of useful indicators that are accessible and understandable by managers, decision makers and other users (**2.1 Types of users and their needs**). To achieve this objective, we have defined several sequential steps that involve data access, statistical modelling, and display of indicators, together with the necessary information technology associated with each of them.

BIRDIE's data, code and outputs are stored and run on three separate servers, that generally correspond to steps in data processing that are described in sections below (Figure 3). The servers include:

1. **African Bird Data server** is part of a different project, but links closely to BIRDIE and hosts its main data sources, the CWAC and ABAP data.
2. **BIRDIE Server B** connects with the Africa Bird Data server to obtain CWAC and ABAP data. It also connects to Google Earth Engine to annotate these data with environmental information. It then runs the main analytical modules of the pipeline, where occupancy and state-space models are fitted. The outputs of these models are stored in an FTP directory from where they are accessible to server A.
3. **BIRDIE Server A** hosts BIRDIE's database and data mart, and runs certain analytical modules to calculate high-level indicators. It also hosts the online dashboard.

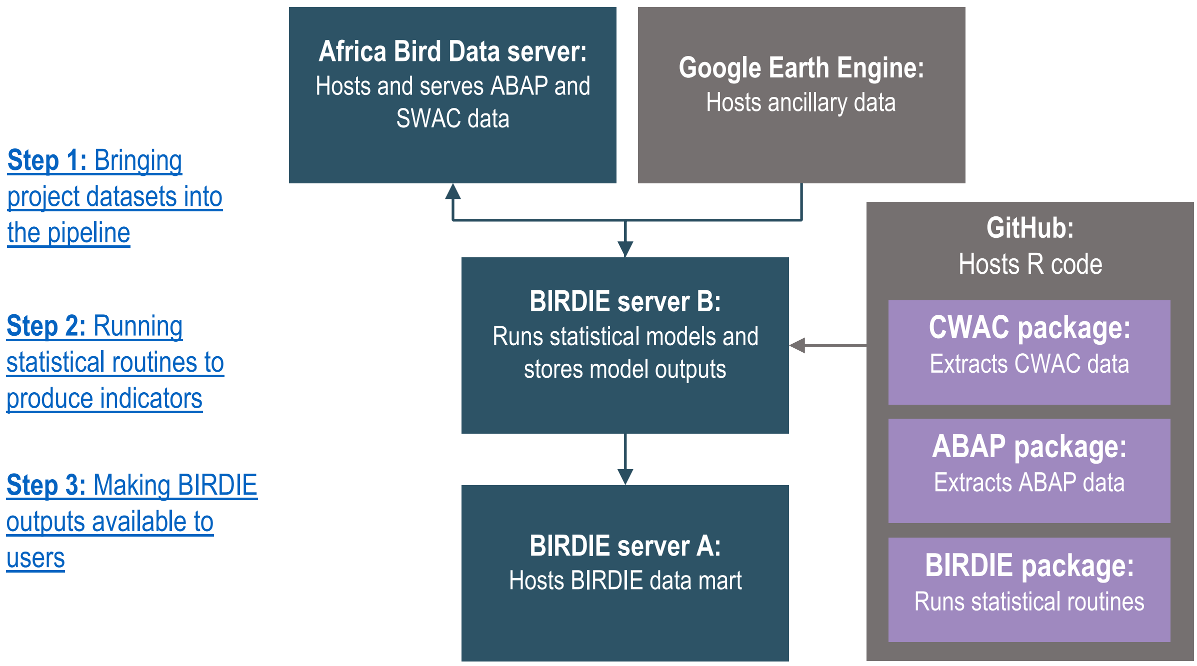


Figure 3: Overview of BIRDIE's software systems.

5.1 Step 1: Bringing project datasets into the data pipeline

The first step of data processing is to access the citizen science databases from the FitzPatrick Institute of African Ornithology (FIAO) (Box 1). The institute hosts a dedicated data server, secured behind the University of Cape Town firewalls. The server is only accessible through the University of Cape Town internal network and validated user machines. All external data systems and third-party access is restricted to the Application Processing Interface (API) calls. All data is backed up daily to three locations (one on an alternative machine in the server room, one repository on campus, and one cloud based).

Box 1: Technical specifications of the FIAO data host server.

- Intel Xeon Dual 8 core; 32 GB RAM; 4 TB HDD
- Ubuntu Linux 16.04 LTS Server Edition
- MySQL 5.7 and MySQL 8.0
- PHP 5.6
- Python 3.8.10

The Africa Bird Data server hosts an API that is dedicated to connecting the main servers of the CWAC and ABAP programmes with the BIRDIE systems (Box 2). There has also been an improvement in the APIs that serve these datasets, which makes access faster and more flexible.

Box 2: Technical specifications of the Africa Bird Data server.

- Intel Xeon Dual 8 core; 64 GB RAM; 8 TB HDD
- Ubuntu Linux 18.04 LTS Server Edition
- NGINX + PHP 7.3 platform and Flight (PHP) routing framework

- PHP 7.2
- Python 3.8.10
- PHP Flight routing framework

Examples of API's running on the CWAC database are:

- The 'CWAC Data extraction - Species data' API, that extracts all the species that are in the CWAC database: https://pipeline.birdmap.africa/cwac/species/
- The 'System assigned reference for the submitted card' API, which has all the info of the different submitted cards: https://pipeline.birdmap.africa/cwac/dictionary
- The 'CWAC table fields description' API, which explains all the headers of the table fields: https://pipeline.birdmap.africa/cwac/metadata

At the same time, several packages for the "R" programming language were developed to be able to pull data from the ABAP and CWAC databases straight into R (R Core Team, 2024). The packages are also named ABAP (https://github.com/AfricaBirdData/ABAP) and CWAC (https://github.com/AfricaBirdData/CWAC) and are hosted on GitHub repositories under the umbrella of the AfricaBirdData GitHub organisation. These R packages improve the overall accessibility of ABAP and CWAC data, as they allow users to download and manipulate data from the ABAP or CWAC from R.

In addition, another R package, ABDtools (https://github.com/AfricaBirdData/ABDtools), was designed to connect to the Google Earth Engine Catalog of environmental layers and pull those data into R. ABDtools also provides functionality to annotate ABAP and CWAC data with this environmental information, making analyses easier to automate and reproduce. The ancillary environmental datasets from Google Earth Engine used by BIRDIE datasets are also presented on the BIRDIE website (see **5.3 Step 3: Making BIRDIE outputs available to users**).

The GitHub repositories under the AfricaBirdData organisation (https://github.com/AfricaBirdData/) provide details on how to use all R packages developed for BIRDIE. Also note that although the CWAC and ABAP packages can be used independently, in BIRDIE they are called from the BIRDIE R package during our automated workflows (see articles at https://africabirddata.github.io/BIRDIE/ on how BIRDIE's data acquisition and analytical routines work).

5.2 Step 2: Running statistical routines to produce indicators

The BIRDIE pipeline uses statistical models to estimate various indicators associated with waterbird occurrence and abundance. The BIRDIE repository in the AfricaBirdData GitHub hosts an R package with the functions required for data acquisition and statistical routines needed for the pipeline.

These tasks run on BIRDIE server B (Box 3), which is hosted by the FitzPatrick Institute of African Ornithology at the University of Cape Town, and include:

1. Connecting to the API of the Africa Bird Data server, downloading CWAC and ABAP data and annotating them with environmental information from the Google Earth Engine catalogue, using the ABAP and CWAC R packages.
2. Fitting occupancy and state-space models to the data using the BIRDIE R package.
3. Diagnose and summarise model fit and extract estimates of occurrence and abundance.
4. Store modelling outputs and indicators in a directory with a File Transfer Protocol (FTP) that enables communication with other systems, notably with BIRDIE server A. This directory is also accessible from the web via HTTPs. Outputs for each species are stored in directories named after the species SAFRING (South African Bird Ringing Unit) code.
5. Version control for the AfricaBirdData packages via Git/GitHub.

Box 3: Technical specifications of BIRDIE server B.

- Intel Xeon Dual 8 core; 64 GB RAM; 8 TB HDD
- Ubuntu Linux 20.04 LTS Server Edition
- Apache 2.24
- PHP 7.3
- R version 4.2.0 (2022-04-22)
- Python 3.8.10

All models are fitted in a Bayesian framework, and estimates are always accompanied by the level of uncertainty around them. Further technical details on how models are fitted can be found in Cervantes et al. (2023) and the most updated information at https://africabirddata.github.io/BIRDIE/.

5.2.1 Abundance and population trends

Abundance of a species refers to the number of individuals of that species at a site. Abundance, or population size, and trends are estimated using state-space models fitted to CWAC data that are collected twice per year (mid-winter and mid-summer). A few sites are counted more often and could potentially be analysed at finer resolution, but currently all analyses are run on a two-season basis. These statistical methods for time series can separate ecological and observation processes, reducing the noise in the data and estimating abundance even when no counts were conducted. In essence these methods are used to deal with gaps in the data and with imperfect counts.

State-space models are used to describe and understand dynamic systems that evolve over time, and may not be perfectly observed (i.e., they might be distorted by some imperfect observation process). For example, estimating the number of birds present at a certain site, may be complicated by field observers who miss some individuals. By counting repeatedly over time, and assuming that the process of interest evolves slowly compared to observation error, it is possible to disentangle these two processes.

Our models consider that an abundance time series can be decomposed in multiple elements: a “long-term” mean abundance, perturbations in the population that produce deviations from that mean, and observation error. Abundance deviations from the long-term mean must be relatively smooth, so in general, frequent, large changes in a population are not expected. On the other hand, observation errors behave erratically. These assumptions help us separate errors in the counts from changes in the populations.

For BIRDIE, the objective is estimating the underlying bird abundance using counts from the CWAC dataset (**4.1 Citizen science datasets**

Large, long-term data sets are immensely valuable and establish a convincing evidence base for action now and into the future. The power of citizen science data has not yet been adequately harnessed. In the words of Barnard and de Villiers (2012), "time is short to build the tools we need to scan the horizon for trends in biodiversity. We have most of the right ingredients, but need to combine them in the right way, supported by sound science and carefully applied statistical analysis. South Africa is in a good position with excellent large-scale, long-term biodiversity databases". These long-term and large-scale spatial biodiversity databases may be built into a national early warning system as a central support tool for South Africa's policy, planning, management and reporting needs.

Large-scale citizen science projects, such as atlases of species distribution, are an important source of data for macroecological research, for understanding the effects of climate change and other drivers on biodiversity, and for more applied conservation tasks, such as early-warning systems for biodiversity loss (Barnard et al., 2017). Enlisting citizen scientists in monitoring offers a dual advantage. First, the increased number of observers allows for more data to be collected from more places than surveys carried out by professional scientists could. Second, by engaging ordinary citizens, there is an opportunity to turn these people into ambassadors for biodiversity as they gain first-hand information on the biodiversity around them.

However, citizen science data are challenging to analyse for two main reasons (Yoccoz et al., 2001, Altwegg & Nichols, 2019, Johnston et al., 2022):

- Sampling usually does not happen in a probabilistic way (Hugo & Altwegg, 2017). Instead, citizen scientists prefer to visit areas that are easily accessible (e.g. close to roads and human settlements) or interesting for some reason (e.g. protected areas).
- Citizen scientists do not detect all species that are present in the area they survey. Imperfect detection is also an issue with almost all other types of surveys, but in citizen science projects, the observation process is typically more variable as the observers vary in their level of skills and the effort they are prepared to put into a survey (Altwegg & Nichols, 2019).

The biological process one is interested in is confounded with the observation process. Analysis must therefore account for the observation process so that the biological signal can be extracted from the data. Properly analysed, however, citizen science data are an invaluable source of information about large-scale biodiversity trends (Devictor et al., 2010, Isaac et al., 2014).

Many environmental decisions are based on lists of species that have been encountered in a particular location. For example, environmental impact studies, decisions about the level of protection that should be afforded to a particular area, etc., often rely on such lists. The problem with this approach is that it ignores the fact that we have incomplete information on biodiversity at any given location. It is unknown whether a short species list reflects low biodiversity or low sampling effort at a particular site. Robust environmental decisions therefore need to be based on monitoring protocols that account for the observation process.

Co-ordinated Waterbird Counts (CWAC)). In addition to estimating abundance and filtering out observation errors, state-space models allow us to estimate the rate of change of a population, with its associated uncertainty The rate of change of a population is calculated as the ratio between the number of individuals at the end of a time period and the number of individuals at the beginning.

Unfortunately, not all CWAC sites have good enough data to estimate abundance for all species. Therefore, we have put in place some filters to remove those sites that do not have enough counts for modelling population trajectories.

The state-space models are fit in a Bayesian framework using the R package *jagsUI* that uses the software *JAGS* in the background. For details on how current filters are implemented and on the most recent state-space model specification, please visit https://africabirddata.github.io/BIRDIE/ and check the articles related to the abundance modules of the pipeline.

5.2.2 Species distributions

The distribution of a species is the area over which it occurs. Distribution maps are based on ABAP full-protocol checklists, which provide information on whether a species was detected or not during visits to pentads. Species occurrence is estimated using occupancy models, which are statistical models that account for false negatives.

Occupancy models are fitted to detection/non-detection data from the APAB to delineate the distribution of waterbird species and its dynamics over time. Within the ABAP framework, observers visit pentads and make a list of the bird species detected during the visit. It is assumed that observers do not misidentify species or otherwise list species that were not actually detected (the rigorous vetting process of ABAP data justifies this assumption), but non-detections may be caused by either the species not being present in the pentad or by the observers not detecting it, although it was present.

Therefore, occupancy models aim to decouple two processes that detection/non-detection data capture: the ecological process whereby a species is present at a site (a pentad), and an observation process that inevitably leads to imperfect detection and results in species being overlooked. They describe two types of occupancy probabilities:

1. **Estimated occurrence** (also known as unconditional occupancy probabilities) is the probability that a species occurs in a grid cell based on the environment (covariates) and known patterns of detections across all grid cells. The models use environmental

data, such as extension of surface water, temperature, altitude or distance from the coast (**4.4 Ancillary data**), associated to those sites where a species was observed to analyse the probability of that species being present in sites where the species was not detected. An estimated occurrence close to 1 means that the species is likely to occur at the site whereas an estimated occurrence close to 0 means that the species is unlikely to occur there.

2. **Predicted occurrence** (also known as conditional occupancy probabilities) is based on estimated occurrence but also takes into account the actual sampling effort and detections at a particular site. If a species was recorded at a site, predicted occurrence is 1. If it has never been recorded, the number of checklists collected at that site, the probability of detecting the species when it is present and the estimated occupancy probability are taken into account. Where a site has a lot of checklists and none of them contain a species that is otherwise easy to detect, this species probably does not occur at this site. On the other hand, if a species is hard to detect and a site only has a few checklists, there is a good chance that it still occurs there and has simply been overlooked.

Each checklist is treated as an independent survey, but occupancy is assessed on a yearly basis. This means that if a species is detected in any one survey it is considered present that year. Therefore, missing a species because it left the site is considered part of the observation process and not the occupancy process.

Probability of occurrence is only estimated based on ABAP data for those species that have been detected in at least 22 pentads each year. These probability estimates range between zero and one, and they are represented by a colour gradient in the maps. For those species that were detected in less than 22 pentads, there is not enough data to provide reliable estimates of occurrence probability. Therefore, we present the raw data that is either zero if the species was not detected, one if the species was detected or `no data` if nobody visited that pentad.

The R package *spOccupancy* is used to fit occupancy models, because it has comprehensive built-in functionality that gives many options in terms of the type of models that can be fitted and also provides model diagnostics algorithms. For details on how current filters are implemented and on the most recent model specification, please visit https://africabirddata.github.io/BIRDIE/ and check the articles related to the distribution modules of the pipeline.

5.2.3 Data and model diagnostics

The pipeline needs to run for a multitude of species, with different ecological requirements and geographical distribution. Therefore, finding a model that suits all species is challenging. Not only may a model not be a good fit for a particular species, but the algorithms used for fitting the model may fail to converge due to characteristics of the data. For example, this type of issue occurs when fitting non-linear effects to occurrence data that is better suited for linear relationships.

Thus, for occupancy models, the first measure taken to accommodate multiple species is defining a set of models to fit each species rather than a single one. The algorithm will try to fit the most complex model first and if it finds convergence issues, it will try the next simpler one and if it runs into problems again, then it will try the last, even simpler option. Unfortunately, none of the models will be appropriate for some species. This typically happens for species with very few records. If a species was detected in less than five pentads in the three-year period defined for occupancy modelling, this species is not analysed, and a note is produced and stored in the output directory allocated to this species.

We work on a Bayesian framework and rely on Monte Carlo Markov Chain (MCMC) algorithms to estimate parameters. This type of algorithm requires chains to converge to steady-state distribution. Gelman-Rubin convergence diagnostics were computed for each MCMC chain produced to sample parameters for each species. These diagnostic values were tabulated and explored visually. Distinctive characteristics of the species presenting convergence problems were explored and new features incorporated to the models, when possible, to accommodate these problematic data.

Posterior predictive checks were used to evaluate goodness-of-fit of occupancy models. This technique entails simulating data using the fitted model computing some quantity of interest and comparing it to the same quantity computed using the raw data. A chi-squared statistic represents how different each observation or simulation is from the expected value modelled for each data point. Diagnostics for state-space models are conducted mostly manually, but posterior predictive checks can be implemented.

These diagnostic measures help to accommodate needs for multiple species, but also facilitate the automation of the pipeline analysis. To check details of current diagnostic techniques used please visit https://africabirddata.github.io/BIRDIE/.

5.2.4 Species richness

Species richness is the number of species that occur at a site. For most grid cells it is not certain whether a species is present or absent, so an estimation of the number of species present is based on the probability of occurrence estimates (**4.3.2 Species distributions**). This is done by summing the occurrence probabilities across species. Estimated occurrence (from ABAP data) is used for this calculation, because predicted occurrence would bias species richness towards grid cells sampled more frequently (i.e., with higher detection probabilities).

More sophisticated diversity metrics take into account evenness and might be useful. For example, Hill numbers would be flexible and elegant. They unify a variety of diversity measures, including species richness, Shannon's index and Simpson's index. However, at the moment, BIRDIE only presents the simple species richness measure described above.

5.2.5 Waterbird Conservation Value

Waterbird Conservation Value (WCV) is an experimental new indicator proposed by the BIRDIE project to support Ramsar and site management reporting. The BIRDIE project has piloted this quantitative method to assess wetland avifaunal importance, originally developed by Harebottle (2016) and taken forward by him within this project.

Traditionally, species richness, species diversity, total count, biomass, energy consumption and the '1% threshold' have been used to assess the importance of wetlands for waterbirds (**3.1.1 Convention on Wetlands of International Importance (RAMSAR)**). Designation of wetlands of international importance (Ramsar sites) based on waterbirds has focused largely on those species meeting the 1% population threshold levels. Only a very few species meet this Criterion against international numbers. This has led to prioritisation of only a subset of species as being important, with little consideration of the contributions of the remaining species' populations.

The index sums the ratio of each species' abundance to its published (by Wetlands International, the acknowledged custodian of bird population data for Ramsar reporting) 1% threshold across all species to give an overall measure of the 'value' of the waterbirds at a wetland. It is a measure of how important a wetland site is for the population of waterbirds, and it offers an alternative to the Ramsar 1% criteria (Criteria 6). The WCV index takes into consideration the proportion of the population of the full variety of waterbird species that are present at the site. Thus, a site can be important because it hosts a great proportion of the population of a single species, but it can also be important if it hosts a relatively large proportion of the population of multiple species. The contributions of all species are summed to calculate the WCV index for the site. Large values indicate that large proportions of the total populations of waterbird species are present at the wetland. Indices can be evaluated at site and species levels. The WCV is a more nuanced approach, sensitive to actual species' abundance rather than counts of '1% threshold' species, and considers all species in the assessment. Data inputs come from standard waterbird surveys as part of CWAC with reference to the latest 1% thresholds published by Wetlands International (http://wpe.wetlands.org/) in order to calculate the ratios for the WCV index.

5.3 Step 3: Making BIRDIE outputs available to users

The final step of data processing is to store the model outputs in a place that can be accessed by users. The BIRDIE data mart is on the BIRDIE server A. It was developed and hosted through the duration of the project by RBINS and is currently hosted at SANBI (Box 4). This server is used for running the OpenAPI, Wildfly and the MySQL database. The **OpenAPI** is a *specification* (OpenAPI 3.0.3 (OAS3)) which is the latest version of the OpenAPI Specification, an open-source format for describing and documenting APIs. OpenAPI has a user interface that guides and allows a users to look for specific data in the MySQL database. This can be the masterlist of species or RAMSAR sites, but this can also be the estimates and counts of a waterbird at a specific wetland at a specific year. The **Wildfly Application Server** is there to build and run all the applications (back end / front end) that were developed, namely, the website, the map viewer, and the OpenAPI user interface. This is a freely available, powerful, modular and lightweight application server. **Apache Maven** is a WildFly Maven Plugin and is used to deploy the web applications to a running WildFly Application Server. The deployment can be executed by using the Maven Plugin, which makes building and deploying new instances and updates easy. The **MySQL Database** is the environment in which the data mart (see **5.3.1 Data mart**

A data mart provides easy access to frequently requested data via APIs. This includes the underlying data described above, in addition to the data drawn through statistical analyses to the web dashboard (Figure 3)(**5 DATA PROCESSING AND ANALYSIS**). The data mart stores raw data (and metadata) optimised for further use and for meeting the requests for data analysis. The data are loaded into this storage area from a data staging area where the data from different data sources have been validated, cleaned, reconciled, transformed and standardised (Figure 3).

The process of building the data mart involved the following 5 steps:

1. **Design:** The first step involved agreeing on the need for the data mart and collecting information about the requirements. Other tasks involved identifying data sources, such as site and species information, pentads, statistical model output data, ancillary data. The multidimensional data is organised following a snowflake schema.
2. **Construct:** During this step, both the physical and the logical structures for the data mart were created, including tables, indexes, fields, and access controls (Table 5).
3. **Populate:** The data mart was populated by transferring data into it. The source information was extracted, cleaned, transformed, and loaded into the data mart. At first this was done manually by transferring the modelling outputs and indicators from the FTP server to the data mart for all sites (currently 741) and all species (currently 147) and all years that had been analysed (currently 1993 - 2022). Once the procedure was in place, this data transfer process could be repeated very quickly (a matter of hours) for all sites, all species and all available years. To ensure that information stored in the structure is clean, it is always overwritten during the population of the data mart.
4. **Data access:** The data, once loaded into the data mart, was put into active use. Activities included querying, generating maps, plots and reports, and publishing. For BIRDIE, two pathways lead on from the data mart:
    - RESTful APIs were designed allowing users to directly access the raw data. An OpenAPI Specification (OAS) using a standard format was created to describe these RESTful APIs. An OAS is machine-readable, but is also easy for humans to read and understand. The specification describes the elements of an API and is commonly used to generate API documentation.
    - The pipeline takes data from the data mart and subjects them to statistical analyses that produce metrics needed for conservation decisions. These metrics are presented on the BIRDIE website.
5. **Manage:** The last step involved the following ongoing tasks:
    - Controlling user access.
    - Refining and optimising the target system to improve its performance.
    - Adding new data into the data mart and managing it (e.g. model outputs for 2023, a new RAMSAR site, a new CWAC site).
    - Configuring recovery settings and ensuring that the system is available even after a possible failure.

- Set the frequency with which new data transfer will be done in an semi-automated way. Currently, this is set to yearly updates of the model output data for each new year.

Table 5: Tables in the BIRDIE data mart.

Sites, species and pentad tables: t_guild, t_json_1_pc_levels, t_json_ramsar_report, t_json_sites, t_json_species_area, t_json_waterbirds, t_json_wcv_sites_species, t_migrant, t_ramsar_report, pentad
Ancillary data tables: t_gee, t_geo_parametres_gee, t_site_convention, t_site_group, t_site_list, t_site_list_full_cwac, t_site_master_list, t_site_type, t_sites_birds_status, t_species, t_status, t_waterbirds, t_wcv_sites_species, t_migrant, t_ramsar_report, t_ratio, t_year, t_season, t_richness
Statistical models output data: t_geo_parametres, t_rat_est, t_rat_est_ci_lower, t_rat_est_ci_upper, t_ratio, t_season_count, t_season_est, t_season_est_ci_lower, t_season_est_ci_upper, t_geo_parametres_ancillary
Linking tables: i_site_convention_list, i_site_group_list, i_site_type_list, i_species_conserv, i_species_guild, i_species_migrant, i_species_status

The bulk of the data is an automated pipeline via the two citizen science projects. In the future, basic additions or corrections to this citizen science framework can also revert to direct automation.

) is hosted. This database is not only chosen because it is the main database type that is being used by SANBI, but it also has a flexible structure, high performance, security and storage management, JSON support, replication and high-availability, manageability and ease of use, OLTP (Online Transaction Processing) that is defined by numerous low-complexity transactions (INSERT, UPDATE, DELETE), and has geo-spatial support.

Box 4: The technical specifications of BIRDIE server A (was at RBINS -> now moved to SANBI)

- Ubuntu Linux 20.xx LTS Server Edition
- OpenAPI Version 3.0.3
- WildFly 26 final (Java EE 8 certified application server)
- JDK 11 (LTS) (Java Development Kit 11, Long Term Support, Extended Support Until January 2032)
- Apache Maven 3.5.4
- MySQL Community Server 8.0.27, with 5 GB RAM

5.3.1 Data mart

A data mart provides easy access to frequently requested data via APIs. This includes the underlying data described above, in addition to the data drawn through statistical analyses to the web dashboard (Figure 3)(**5 DATA PROCESSING AND ANALYSIS**). The data mart stores raw data (and metadata) optimised for further use and for meeting the requests for data analysis. The data are loaded into this storage area from a data staging area where the data from different data sources have been validated, cleaned, reconciled, transformed and standardised (Figure 3).

The process of building the data mart involved the following 5 steps:

6. **Design:** The first step involved agreeing on the need for the data mart and collecting information about the requirements. Other tasks involved identifying data sources, such as site and species information, pentads, statistical model output data, ancillary data. The multidimensional data is organised following a snowflake schema.
7. **Construct:** During this step, both the physical and the logical structures for the data mart were created, including tables, indexes, fields, and access controls (Table 5).
8. **Populate:** The data mart was populated by transferring data into it. The source information was extracted, cleaned, transformed, and loaded into the data mart. At first this was done manually by transferring the modelling outputs and indicators from the FTP server to the data mart for all sites (currently 741) and all species (currently 147) and all years that had been analysed (currently 1993 - 2022). Once the procedure was in place, this data transfer process could be repeated very quickly (a matter of hours) for all sites, all species and all available years. To ensure that information stored in the structure is clean, it is always overwritten during the population of the data mart.
9. **Data access:** The data, once loaded into the data mart, was put into active use. Activities included querying, generating maps, plots and reports, and publishing. For BIRDIE, two pathways lead on from the data mart:
   - RESTful APIs were designed allowing users to directly access the raw data. An OpenAPI Specification (OAS) using a standard format was created to describe these RESTful APIs. An OAS is machine-readable, but is also easy for humans to read and understand. The specification describes the elements of an API and is commonly used to generate API documentation.
   - The pipeline takes data from the data mart and subjects them to statistical analyses that produce metrics needed for conservation decisions. These metrics are presented on the BIRDIE website.
10. **Manage:** The last step involved the following ongoing tasks:
    - Controlling user access.
    - Refining and optimising the target system to improve its performance.
    - Adding new data into the data mart and managing it (e.g. model outputs for 2023, a new RAMSAR site, a new CWAC site).
    - Configuring recovery settings and ensuring that the system is available even after a possible failure.
    - Set the frequency with which new data transfer will be done in an semi-automated way. Currently, this is set to yearly updates of the model output data for each new year.

Table 5: Tables in the BIRDIE data mart.

Sites, species and pentad tables: t_guild, t_json_1_pc_levels, t_json_ramsar_report, t_json_sites, t_json_species_area, t_json_waterbirds, t_json_wcv_sites_species, t_migrant, t_ramsar_report, pentad
Ancillary data tables: t_gee, t_geo_parametres_gee, t_site_convention, t_site_group, t_site_list, t_site_list_full_cwac, t_site_master_list, t_site_type, t_sites_birds_status, t_species, t_status, t_waterbirds, t_wcv_sites_species, t_migrant, t_ramsar_report, t_ratio, t_year, t_season, t_richness

Statistical models output data: t_geo_parametres, t_rat_est, t_rat_est_ci_lower, t_rat_est_ci_upper, t_ratio, t_season_count, t_season_est, t_season_est_ci_lower, t_season_est_ci_upper, t_geo_parametres_ancillary
Linking tables: i_site_convention_list, i_site_group_list, i_site_type_list, i_species_conserv, i_species_guild, i_species_migrant, i_species_status

The bulk of the data is an automated pipeline via the two citizen science projects. In the future, basic additions or corrections to this citizen science framework can also revert to direct automation.

5.3.2 Backend process

Creating a dashboard, especially an interactive one, involves a complex backend process to ensure seamless data retrieval, processing, and presentation. Figure 4 shows a breakdown of the backend process of how data is retrieved to create such a dashboard and further discussed in detail.

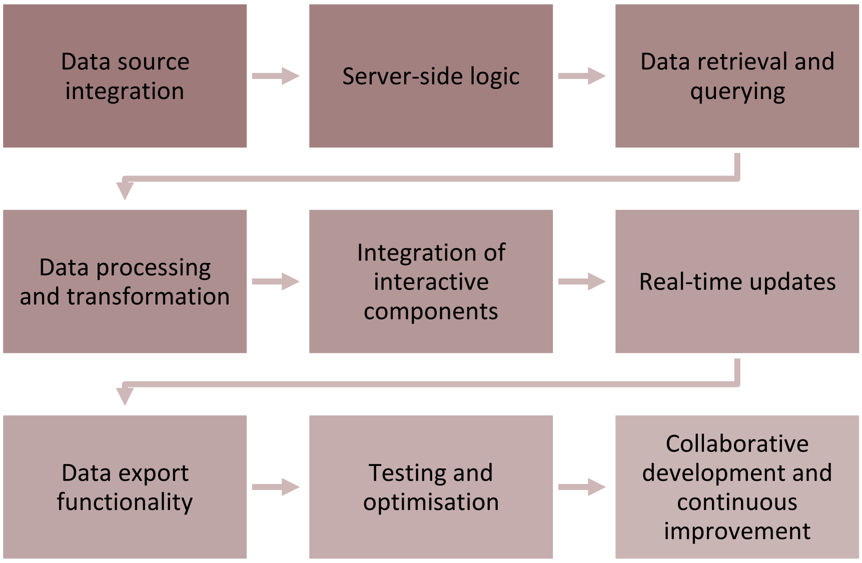


Figure 4: Backend process of the BIRDIE dashboard.

To start the backend process, the first step was to integrate data sources into the dashboard's architecture. These data sources encompassed databases (CWAC, ABAP), RESTful APIs (Google Earth Engine), external data providers (master lists), and data files (Excel file with ancillary data). These sources contained information related to species, sites, indicators and associated data points. Establishing connections to these sources was crucial for data retrieval.

Once the data was retrieved, the backend processed it to make it suitable for presentation in interactive maps and plots. This processing involved data cleaning, quality control, and transformations (pivot tables). For example, the modelled data has to be checked on the data type, the correct format, and numerical data needs to be in the correct formatting for plotting.

The next step was to employ the backend scripting languages to develop the server-side logic. This logic acts as an intermediary between the dashboard's frontend and the data sources. It handles data requests and responses, ensuring that data is retrieved efficiently and securely. For example, when a user selects specific species or indicators in the frontend website, the backend processes these requests and retrieves the corresponding data.

The backend logic performed data retrieval by sending queries to the connected data sources. These queries were tailored to fetch the desired data based on user interactions. API requests were used for getting all data from the datamart. For instance, if a user selected a particular species, the backend logic uses the available APIs to retrieve all relevant data related to that species.

Backend developers integrated interactive libraries like OpenLayer and Plotly into the dashboard. These libraries provided tools for creating dynamic maps, plots and charts. The backend communicates with these libraries to populate them with processed data. For instance, when a user selects a specific site or indicator, the backend updates the OpenLayers map and Plotly charts with the data coming from APIs to display the relevant information.

User interaction is crucial for an interactive dashboard. Backend logic handles user interactions by continuously listening for frontend requests and updating the displayed data accordingly. If a user zooms in on a map or selects a different combination of species and indicators, the backend ensures that the requested data was fetched and presented.

The backend also managed the data download functionality. When a user requested data downloads in CSV format, the backend generates the requested file and served it to the user for download. Thus, this is dynamically generating files based on user preferences.

In summary, the backend process of creating an interactive dashboard involved integrating data sources, developing server-side logic, retrieving, and processing data, integrating interactive components, handling updates, providing data export functionality, conducting rigorous testing, and fostering collaborative development. These backend processes were essential for delivering a web application that empowered users to explore and download data conveniently while navigating species, sites, indicators, and their combinations within an intuitive user interface.

5.3.3 Integration with SANBI systems

The development web platform (hosted throughout the duration of the project at RBINS) was migrated to the final production website (hosted by SANBI) by uploading the pages to the SANBI server during mid-September 2023 and updating the production database.

The final upload of the Map Viewer and the final database export and import (including the 2022 year) successfully took place during the project extension in early 2024.

This was a crucial pilot to investigate the capacity for SANBI to successfully host a website created through open source software (different software to the core software of the main platform) on the SANBI Biodiversity Advisor platform. Hosting by SANBI also ensures that the BIRDIE website will receive the maintenance associated with maintaining a functioning website.

See the BIRDIE Technical Design Document, based on SANBI's *technical design template* and the BIRDIE Installation and Management Document, based on SANBI's *installation and management template*.

6 DASHBOARD WEBSITE

The dashboard is a web application (https://birdie.sanbi.org.za/) that provides interactive maps and plots available for users to put together, explore and download, at their convenience. Creating an interactive dashboard involves a detailed process of data retrieval to enable users to access interactive maps and plots while providing options for exploration and data download. Additionally, the website allows users to navigate through various components, including species, sites, indicators, and combinations thereof. The development of this dashboard, represented by the prototype (https://dataviz.naturalsciences.be/birdie), underwent a collaborative team effort to ensure it fulfilled the desired functionalities. All code development for the dashboard website was conducted using HTML5, CSS, and widely used JavaScript libraries such as OpenLayers and Plotly.

6.1 Website structure

The website's structure consists of multiple primary sections and subpages, enabling users to easily navigate and engage with diverse information, including tables and maps (Figure 5).

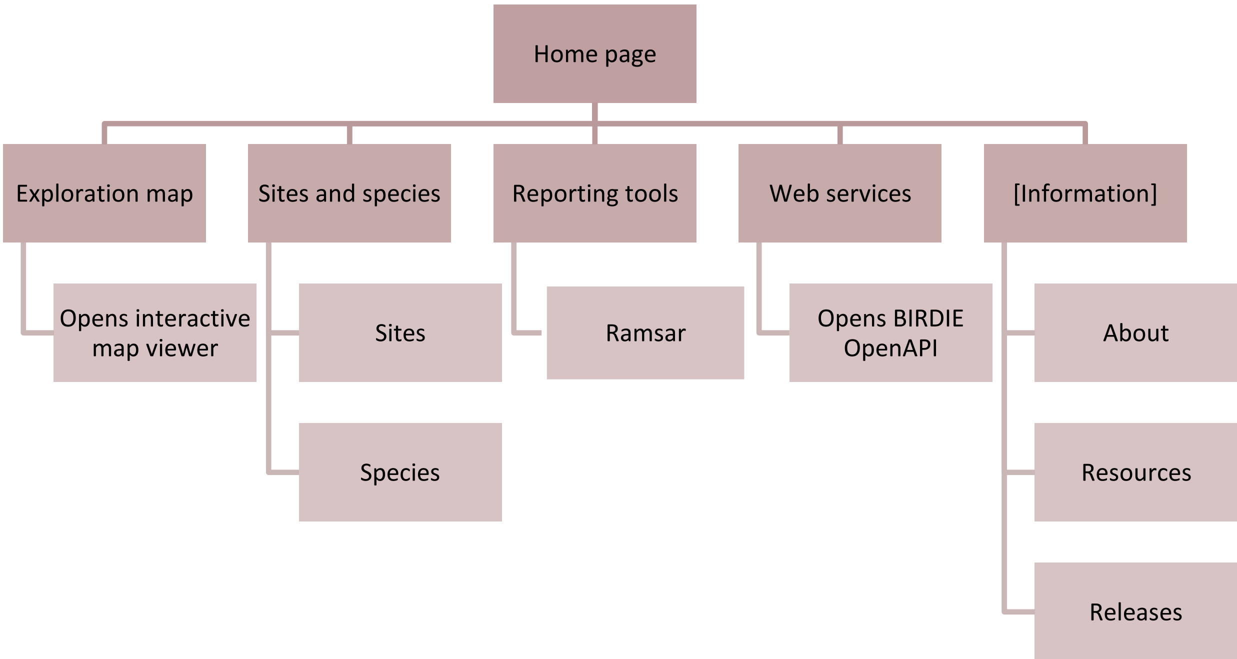


Figure 5: The pages of the BIRDIE website.

The structure of the website can be broken down for each component as follows:

a. Home page:

   i. The home page serves as the entry point to the website
   ii. It contains buttons or links that direct users to various sections or pages of the website.
   iii. It provides a brief introduction about the BIRDIE project.
b. Exploration Map:
   i. Clicking on the “Exploration Map” button takes users to an interactive map viewer.
   ii. This map viewer allows users to explore indicators, species and sites according to user-specific selections.
c. Site and Species pages:
   i. This page includes links to subpages that provide summarised information:
      1. Sites: This subpage contains information about various monitored wetland sites, including tables and descriptions of wetlands and the species found at the site.
      2. Species: This subpage provides information about different waterbird species, accompanied by images, and tables containing information about the species.
d. Reporting Tools:
   i. The “Reporting Tools” section includes subpage links to reporting tools.
      1. Ramsar reporting
e. Web Service:
   i. The “Web Service” section includes a link to the BIRDIE OpenAPI.
   ii. The link provides access to an API that allows developers to interact with and retrieve data programmatically.
f. The "Information" section includes links to various informational pages:
   i. About: This page provides information about the website, its purpose, and its creators and sponsors.
   ii. Resources: The "Resources" page contains additional materials, documents, or links related to the website's content on waterbird species.
   iii. Releases: This page provides information on the latest and previous release versions. It provides an indication of what has been updated for each version.

The website’s structure is designed to provide users with easy access to different types of information and tools. The home page serves as a central hub, offering navigation buttons to explore interactive maps, detailed summaries of sites and species, reporting tools related to international agreements, access to a web service, and important informational sections about the website itself. This structured layout aims to guide users efficiently to the content and resources they are interested in exploring on the website.

6.2 Visualisation of indicators

The indicators are primarily presented in three formats: interactive maps, plots and tables. These allow the user to explore the information in a format that is most relevant to their needs.

6.2.1 Spatial indicators

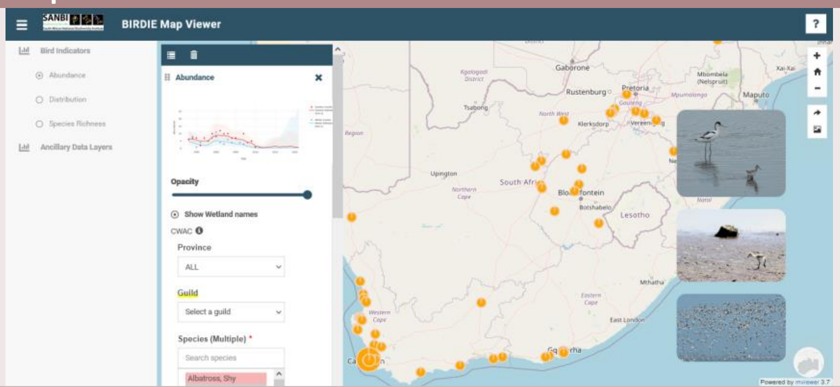
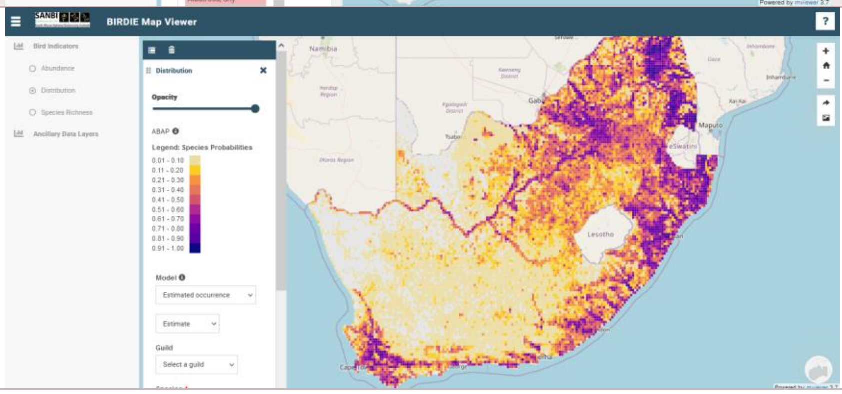
A cartographic tool, known as a map viewer, was used to visualise the spatial indicators (Table 6). OpenLayers, an open-source JavaScript library, similar to Google Maps or Bing Maps, was utilised for displaying map data in web browsers.

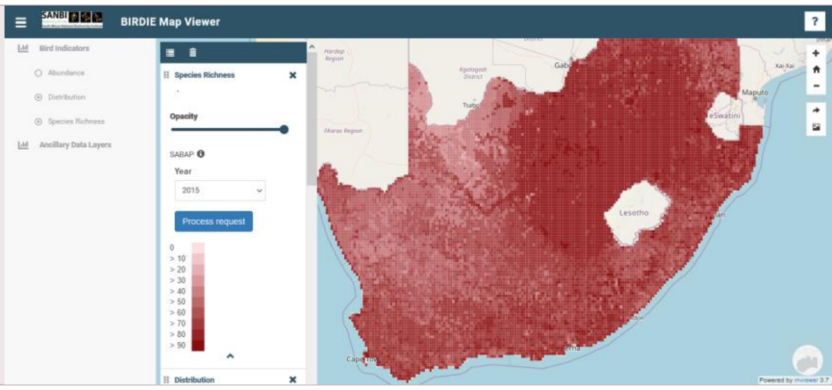
The map viewer is based on mviewer (https://mviewer.netlify.app/en/), a free and open-source cartographic application. This application featured an intuitive and user-friendly interface while adhering to established standards and APIs. Furthermore, it harnessed the capabilities of Bootstrap, recognized as the most widely adopted CSS Framework for crafting responsive and mobile-first websites.

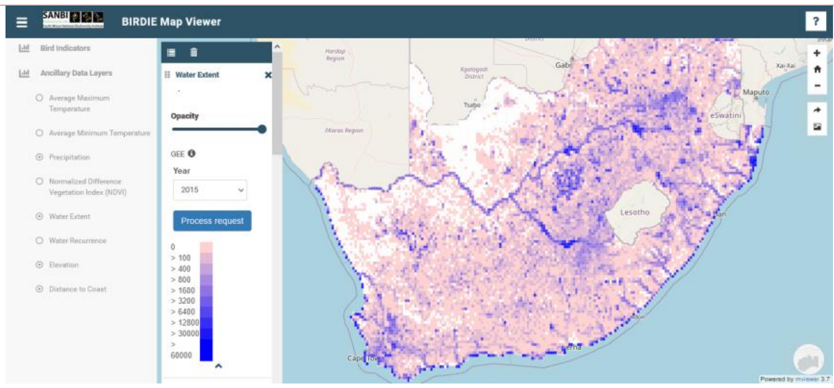
Upon request from a user, the dashboard pulls information from the BIRDIE Server A in the GeoJSON format via an HTTP request from our OpenAPI, the resulting GeoJSON contains the geometries of the pentads for the maps and the information required to draw the plots.

This approach involved utilising OpenLayers to construct a mashup, combining a tiled basemap representing South Africa with a layer that was specifically requested. A "requested layer" represented an entire web service tasked with extracting data from the data mart through BIRDIE Server A, treating it as a layer, and subsequently presenting it within the web application. This method seamlessly converged map visualisation and data presentation within the dashboard.

Table 6: Visualisation of spatial indicators in the map viewer.

	Indicator	Map viewer
Bird indicators	Abundance	
	Distribution	

	Species richness	

Ancillary data layers	Average maximum temperature	
	Average minimum temperature	
	Precipitation	
	Normalised Difference Vegetation Index (NDVI)	
	Water extent	
	Water recurrence	
	Elevation	
	Distance to coast	

6.2.2 Time series indicators

Abundance plots show how the population of waterbird species have changed over time in the different CWAC sites. Plotly is used for displaying the different time series plots and indicators. Plotly is an open-source JavaScript graphing library for displaying plots, including statistical graphs.

A logarithmic scale is used for plots as abundance at CWAC sites often stays relatively low, but has large peaks in certain years. This data structure requires a large axis range but most data with relatively low values are displayed at the bottom of the plot. The log transformation is used to compress the large range of the data into a narrower axis range, making changes over time to be more evident and interpretable. However, in a logarithmic scale, changes in large values appear to be smaller relative to changes in small values.

In all these plots, users have access to basic functions that enhance interaction. Next to the plotare buttons for actions such as downloading the plots as a PNG image, zooming in and out, panning, box selection, lasso selection, autoscaling, and resetting axes.

6.2.3 Interactive tables

There are multiple tables with information on BIRDIE. Most of the tables are static. However, the bird information table on the Site pages and the abundance tables on the Species pages can be filtered and ordered. Users can interact with these tables by using either the search box to search for a particular species, or using the drop-down menus below each heading to filter according to the information they are interested in.

7 USER TESTING AND TRAINING

7.1 Internal testing

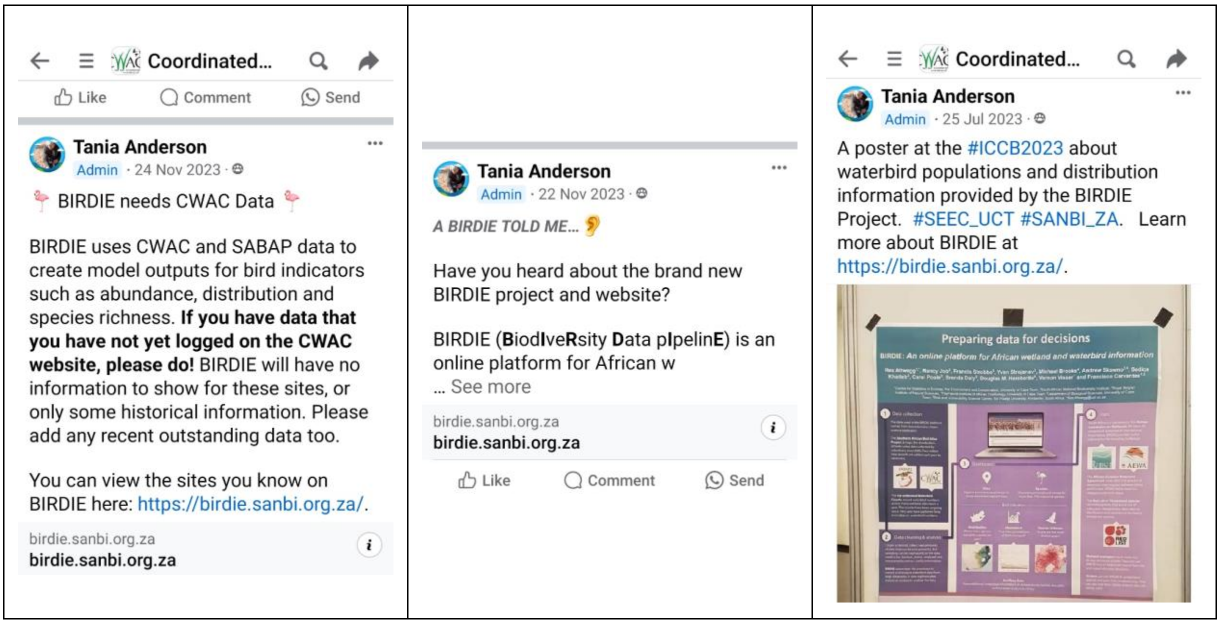
Iterative internal testing through feedback from the project team took place once the prototype dashboard was in place. The testing of the dashboard was a collaborative effort with the BIRDIE team. Two in person team workshops were held, which focussed on the website development. Team members worked together to design, implement, and maintain the backend processes. Continuous feedback and improvement cycles were essential to refine the dashboard's functionality and ensure it aligned with user requirements. Workshops were held to first describe these specifications and after development were organised to check on any open issues and to do the stress testing.

Throughout the website development process, extensive internal testing by the web design team was conducted to ensure that data retrieval, processing and presentation functioned smoothly and efficiently. This testing involved selecting all species and all available years to see if the map viewer would perform correctly and would in a reasonable time show the results of this request. The map viewer was also tested against how quickly the distribution layers showed when selecting a species and a specific year. On the website, the reporting pages were tested to see if they would still perform correctly, and show the correct information, without a time-out from the server, after many times of selecting a site or a species. The OpenAPI UI was also tested on several occasions, with multiple dummy tests of querying the APIs to see if the OpenAPI and the webservices behind it performed efficiently. Performance optimisation by using the JSON data type field for the occupancy model data (ABAP) was vital to reduce server load times and enhance the responsiveness of the dashboard, ensuring a seamless user experience.

7.2 User testing undertaken in the last phase of the project

Following the full development of the dashboard, there was a short period of iterative improvement of the website functionality based on user testing and training. Consultation with users was critical to ensure the data pipeline outputs were fit for purpose, and to allow the project to benefit from the advice and critique of the managers and champions knowledgeable about their sites and the species likely to occur there.

During the project extension a research assistant, Marianne Stevens, was brought on board to support SANBI engagement with stakeholders, to introduce and guide users through the website and solicit user feedback. Dr Tania Anderson of BirdLife South Africa who supports the FIAO on SABAP as well as CWAC citizen science engagement assisted with stakeholder contacts, advice and oversight. Through the increased awareness of this post at the FIAO and through BIRDIE user engagements, a number of new CWAC sites were added, data errors identified and corrected, and location information and maps improved. Some Facebook posts by Dr Anderson to CWAC champions are pictured below:



The BIRDIE project engagement with Coordinated Waterbird Count citizen scientists included an outreach email to approximately 150 CWAC champions to introduce the project and share the website url followed by a meeting. The majority of engagements were online and included a powerpoint presentation, a guided walk through the website, and a discussion component to encourage users to ask questions and share their responses to the website. The main outreach was to bird clubs around South Africa, as the group that hosts the majority of CWAC champions. Provincial Conservation and SANParks scientists were also engaged. In total, meetings were held with champions in the Western Cape Province (5 meetings), Gauteng Province (3 meetings), Mpumalanga Province (2 meetings), Free State Province (2 meetings), Eastern Cape Province (2 meetings), Northern Cape Province (2 meetings), NorthWest Province (2 meetings), KZN Province (1 meeting). The meetings took place in November and December 2023 as the website came on board. Email engagement continued through February, with a few additional meetings in March 2024.

The queries raised by stakeholders during the meetings were taken to the project team of Francis Strobbe (if they were website and database questions), Pachi Cervantes (if they related to the models and stats), Doug Harebottle (if they related to waterbird terminology and habitat preferences and the Waterbird Conservation Value, and Tania Anderson and Michael Brooks (if they related to the CWAC protocols or CWAC in general).

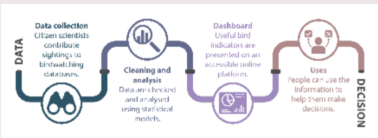
Importantly, the queries were captured into the expansion of the Frequently Asked Questions document and this document is now updated and available on the website. They were also used to generate tasks on the github internal project development site.

On February 6 and 7 2024, a major project stakeholder workshop was held to reflect on the project achievements, discuss the overarching data needs and monitoring gaps and, following this, identify priorities for funding and implementation in a comprehensive way. This is captured in a workshop proceedings document. The overall aim of the workshop was to review the achievements of the BIRDIE project and develop a roadmap for the way forward. The workshop confirmed the keen engagement at all levels of the data pipeline. It was agreed that effective biodiversity monitoring and reporting are essential for wetland and waterbird conservation efforts in South Africa and that BIRDIE is making a contribution to addressing challenges to enhance data accuracy, accessibility, and usability, ultimately supporting informed decision-making and sustainable management practices.

7.3 Development of communications materials

A wide range of communications materials were developed to make users aware of the BIRDIE website and to help them to understand how to navigate and use the dashboard (Table 7). These included developing a unique "look and feel" for the BIRDIE website, building from the project representative photo of Lesser Flamingos taken by Doug Harebottle.

Table 7: Communications materials developed by the BIRDIE project.

Item	Thumbnail	Description
Icon		An icon to be used when visually communicating the BIRDIE project (not a project logo). Design by SANBI graphics using the project “look and feel” and the flamingo representation.
QR code		A QR code was developed to take people directly to the BIRDIE website on scan. The QR code can be used to track user interaction with communications products.
Infographic		A simple representation of the BIRDIE pipeline to communicate the overall process of taking data, cleaning and analysing it, and presenting it on a dashboard for users. Used on the website in communications materials.
Brochure		A 4-page, A5 brochure explaining the BIRDIE project for a wide audience including users. The brochure was printed and has been distributed at the BirdLife bird fair and other user engagements.
Poster		The centre spread of the BIRDIE brochure also works as a standalone poster that visually describes the pipeline process. It has been used for poster displays communicating the BIRDIE project.
Bookmark		A bookmark with simple information about the BIRDIE project, including the web address and QR code. The bookmark was printed and distributed at the BirdLife Bird Fair and other user engagements.

FAQ page		A detailed FAQ page for the website with all the information a user would need structured into easily understandable sections and questions.
User guides		A set of four user guides targeted at specific user groups to show them how BIRDIE is relevant to them and how they can go about using the website. User guides were developed for: (1) CWAC, (2) Ramsar, (3) site managers, and (4) wetlands. All of the user guides were printed for distribution at user engagements.
Bird poster templates		A template in MS Publisher to allow others to make their own bird information posters using the information from the BIRDIE page. Information draws directly from the BIRDIE species master list and Exploration Map. The posters can be printed or used as social media cards. A number of different species were developed and several were printed.
Wetland poster templates		A template in MS Publisher to allow others to make their own wetland site information posters using the information from the BIRDIE page. Information draws directly from the BIRDIE site master list. The posters can be printed or used as social media cards. A number of different wetlands were developed and several were printed.

8 MAINTENANCE - ANNUAL UPDATES AND BUG FIXES

In this section we describe the maintenance necessary to ensure that the BIRDIE systems run smoothly through regular annual updates and bug fixes. It is important to understand that annual updates should be predictable, happen at the beginning of each year, and could take a couple of months. On the other hand, bug fixes must be able to happen at any time and are inherently unpredictable in terms of complexity and time required for the fixes. Therefore, resources allocated for the maintenance of the BIRDIE pipeline should contemplate both regular updates and unexpected fixes. New functionality should be planned with the necessary resources in mind (**9 VERSION UPGRADES**).

We describe the most important maintenance tasks organised according to the different processing steps of the pipeline they belong to.

8.1 Data acquisition

Regular annual updates of BIRDIE's outputs start with accessing updated waterbird data collected by the ABAP and CWAC programmes, as well as the various environmental data pulled from the Google Earth Engine catalog. This requires that the following tasks (and responsible institutions) take place:

- ABAP and CWAC programmes are running, and data are accessible through the API (FIAO). Any changes that affect the functionality of the R packages must be reported to the person maintaining those packages.
- ABAP, CWAC and ABDtools R packages (currently on GitHub, see https://github.com/AfricaBirdData) can access the ABAP, CWAC and GEE APIs, respectively (SEEC).
- Addition of new functionality to ABAP and CWAC APIs and bug fixes (FIAO).
- Addition of new functionality to ABAP, CWAC and ABDtools R packages and bug fixes (SEEC). New functionality must be properly documented and tested, if applicable.

8.2 Data analysis

Once the data enters the system, R modules are run to estimate waterbird distribution and abundance. These modules run on server B and require the following tasks (and responsible institutions):

- Server B needs the capacity to run the analytical routines (FIAO), for which we must ensure we at least have technical specifications similar to the current ones (see **5**
- **Step 1: Bringing project datasets into the data** pipeline). The software R should be installed with the necessary packages to run the pipeline - see the AfricaBirdData packages and their dependencies (https://github.com/AfricaBirdData). All packages and security updates should be kept up to date.
- We have to run the R modules using the BIRDIE R package (SEEC), which is currently on GitHub at https://github.com/AfricaBirdData/BIRDIE. Instructions on how to run all the modules can be found at the website https://africabirddata.github.io/BIRDIE/. Note that server B currently runs an Ubuntu operating system and therefore all analyses are run from the command prompt rather than from a graphical interface, such as RStudio.
- There are some functions in the BIRDIE package that need to be looked at on every annual update (SEEC). For example, we must make sure that the environmental layers we use from GEE have information for the current year or else, we must use information from the last year available. Please, follow the instructions on https://africabirddata.github.io/BIRDIE/, carefully.
- Streamlined models can be unstable, and the outputs must be carefully and critically examined after each annual run to make sure they are working as intended (SEEC).
- Addition of new functionality to the BIRDIE R package and bug fixes (SEEC). New functionality must be properly documented and tested, if applicable.

8.3 Database

Regular general maintenance is essential for a well-functioning MySQL database. This means we need to:

- Ensure data safety by scheduling regular backups.
- Optimise SQL queries for efficiency and monitor performance using tools that track CPU, memory, disk I/O, and query execution.
- Keep MySQL server software up-to-date to benefit from security patches.
- Periodically check and repair tables for potential corruption, and manage indexes to enhance query performance.
- Review and maintain disk space, also by considering purging old data or old backups to free up storage and truncating log files.
- Make sure secure user privileges are set.
- Periodically review and update MySQL configuration settings based on your application's evolving needs and resource changes.

These practices collectively contribute to a robust and efficient MySQL database. Most of these tasks will be covered by the Cloud Service contract (AZURE), however some aspects are BIRDIE datamart specific, rather than MySQL related. Within SANBI, the Biodiversity Information Management Team are the key team to take care of this.

Besides the regular maintenance of the database system, the regular annual updates of the datamart itself is an important element in the data analysis pipeline. After data analysis takes place, the model output data needs to go in the datamart. This is done through stored procedures that are documented in the BIRDIE Technical Design Document. (See the BIRDIE Technical Design Document, based on SANBI's *technical design template* and the BIRDIE Installation and Management Document, based on SANBI's *installation and management template*.)

Finally, bug fixes can happen at any time and consist of corrections that need to be made inside the BIRDIE datamart on records in e.g. the species master list table, the site/RAMSAR master list table, or the Waterbird Conservation Value table.

8.4 Updating the live website

Maintaining a website built with HTML5, CSS, and JavaScript, incorporating OpenLayers, Plotly, and an OpenAPI UI involves front-end and back-end tasks. For content updates (text), front-end adjustments are typical, while data-driven sections (tables, maps, plots) often require back-end involvement.

Front-End updates include:

- Text Content Updates:
    - Modify HTML and CSS for text content changes.
    - Ensure responsive design for optimal display on various devices.
- API Integration (OpenAPI UI):
    - Update and enhance user interface experience by updating the OpenAPI descriptions file (birdie_application.json).
    - Ensure OpenAPI UI components reflect any changes in API specifications.
- Map Viewer:
    - JavaScript updates for dynamic map interactions.

  - Enhance user experience with more options to choose from in the navigation panel.

Back-End updates include:

- Data-Driven Sections:
  - Update server-side code (Java, JavaScript) for dynamic content.
  - Integrate with APIs to fetch and present updated data for tables and plots.
- API Integration (OpenAPI):
  - Update server-side logic to align with any changes in the OpenAPI specifications.
  - Ensure data consistency and accuracy when interacting with APIs.
- Plotly Charts:
  - Retrieve and process updated data on the server.
  - Update Plotly charts dynamically based on all available data.

8.5 Ongoing requirements for keeping the system running

A first ongoing requirement has to do with licensing and space on servers. This is an ongoing requirement of concern to System Administrators and License Managers. The current system, at **SANBI**, runs on Azure, necessitating proper licensing arrangements to ensure compliance and uninterrupted service. Sufficient space on servers must be maintained, with monitoring systems in place to continuously evaluate storage usage, anticipate future needs, and prevent any capacity-related disruptions.

A number of ongoing risks and challenges have been identified. Firstly, ongoing overall risks and challenges of concern to System Administrators and the IT Security Team at **SANBI** include undertaking regular updates, which are essential to ensure the system remains compatible with the latest software and application versions, optimising performance, and user experience. Monitoring of server performance should be also conducted regularly to identify potential bottlenecks, resource constraints, or anomalies, enabling prompt resolution and efficient resource allocation. Security updates are imperative to safeguard against evolving threats and vulnerabilities, necessitating continuous vigilance and proactive measures to ensure data integrity and user privacy.

Secondly, ongoing overall risks and challenges of concern to Project Managers at **SEEC** include the necessary challenge of securing sustainable funding to sustain BIRDIE maintenance. This highlights the importance of strategic fundraising efforts and partnership collaborations to support ongoing operations and development initiatives. In addition, long-term maintenance necessitates a proactive approach to resource mobilisation, innovation, and stakeholder engagement to ensure the system's viability and relevance in addressing emerging conservation and research needs.

Specific long-term maintenance needs for BIRDIE firstly include ongoing responsibilities for the **FIAO**, which is responsible for maintaining critical components of BIRDIE such as SABAP and CWAC data generation, API functionalities, and R packages, ensuring seamless integration and reliable performance. Regular updates and monitoring of Server B's performance and security parameters are essential to mitigate risks, optimise resource utilisation, and uphold service reliability and data integrity.

A second ongoing long-term maintenance need for the **SANBI** System Administrators and **FIAO** is to ensure ancillary datasets from Google Earth Engine are automatically updated, leveraging scheduled processes and monitoring mechanisms to detect and rectify any disruptions promptly, thereby ensuring the availability and accuracy of geospatial data resources for analysis and visualisation.

A third ongoing long-term maintenance need for **SANBI Data Analysts** and **FIAO-SEEC-SANBI Personnel** relates to R Package and Python Models on Server B. Maintenance activities encompass ongoing support and optimisation efforts by SANBI Data Analysts and FIAO-SEEC-SANBI personnel, focusing on enhancing model accuracy, scalability, and usability. Regular training sessions for new staff members and continuous monitoring of system performance and user feedback are critical to fostering a collaborative and adaptive maintenance culture that prioritises user satisfaction and operational efficiency.

A fourth ongoing long-term maintenance need for **SANBI** Data Managers and System Administrators relates to Datamart Import Procedures**.** Import procedures should ideally be automated or semi-automated, leveraging scripting and scheduling tools to streamline data ingestion processes and ensure consistency, reliability, and traceability in data management workflows. Implementation of a robust versioning system is necessary to facilitate effective change management, auditability, and reproducibility of data transformations and model outputs, enhancing transparency and accountability in decision-making processes.

A fourth ongoing long-term maintenance need for **SANBI** System Administrators and Database Administrators is general maintenance for Server A. Regular updates of software and applications on Server A are essential to address software vulnerabilities, performance bottlenecks, and compatibility issues, safeguarding system stability and resilience in dynamic computing environments. WildFly updates and database management activities should be performed diligently to minimise service disruptions, data corruption risks, and compatibility conflicts, prioritising system reliability and data availability for stakeholders. Scheduled backups of data mart and export folders are critical to mitigating data loss risks associated with hardware failures, software errors, or malicious activities, enabling timely data recovery and continuity of operations in the event of unforeseen incidents.

A final ongoing long-term maintenance need for **SANBI** Quality Assurance Team, Technical Review Committee is Regular Testing and Technical Review**.** Scheduled testing of the pipeline involves comprehensive validation and verification activities to ensure compliance with functional requirements, data quality standards, and performance benchmarks, fostering confidence in system reliability and user trust. Establishment of a BIRDIE tech reference committee facilitates cross-functional collaboration, knowledge sharing, and peer review processes, enabling continuous improvement, innovation, and best practice dissemination across the user community.

9 VERSION UPGRADES

9.1 Updates and versioning system

Regular programming conventions have been adopted to record versioning for the BIRDIE platform. The "major.minor.patch" system has been applied, so that version 1 is v1.0.0.

Any **fixes and patches** will increase the last digit (e.g., v1.0.1). **New features** and **database updates** would increase the minor version (second digit, e.g., v1.1.0). **Major changes that significantly affect existing functionality** would increase the major version (first digit, e.g., v2.0.0). For more information refer to https://semver.org/.

The version code will be displayed next to the BIRDIE name to the left top of the webpage, so that the development team and users can keep track of when changes happen in BIRDIE and what those changes are. All version changes will be logged and stored by SEEC and/or SANBI.

In principle, the plan is to update the database and website with new information once a year. This will be initiated by SEEC and SANBI. This would translate into minor version updates. Further ideas about additional functionality into BIRDIE could be part of major updates in the future.

9.2 Potential future updates

A number of issues arising were listed as they arose on to the internal team github project. Most were dealt with during the project, but some were flagged as tasks to be addressed through future funding (see Phase 2 roadmap for full list). Some of the most important potential future tasks, from the perspective of the current team and based on user feedback, are highlighted below.

Important records to add: sightings of rarities; invasive species. This does not require any statistical processing but would simply flag particular records and make them more visible. The usefulness of this depends on how quickly new records travel through the pipeline and into the data mart. An added useful functionality could be to allow stakeholders to sign up for alerts on these.

For our next step we would like to develop bespoke model diagnostics that will help automatically flag problems. This is an important step towards automating the pipeline. What we do with species that we can't model will depend on the specific indicator in question. The three options are 1) using simpler statistical methods that ignore the observation process, 2) use the raw data, 3) leave out the ‘problem’ species / observations. For some species and situations, 1) might be feasible, where not, we choose 2) or 3), whatever makes more sense for the specific situation (see Section 4.3 and note within the Phase 2 roadmap on way forward for expanding to neighbouring countries).

For consideration when rolling out to other countries, the lessons learned from dealing with data sparseness in the South African data will be important when we extend the pipeline to include Kenyan waterbird data. While their data are collected in a similar format as the SABAP and CWAC data, their data sets are generally sparser (see Section 4.3 and note within the Phase 2 roadmap on way forward for expanding to neighbouring countries).

Temporal resolution will need some consideration. Ideally, we would like to estimate occupancy on a monthly basis but this is limited by data availability. Annual occupancy probabilities are definitely obtainable and the current pilot analyses are being undertaken yearly. As such, the project is only a glimpse of what is possible for an ecosystems component of a data pipeline. It is hoped that the methods developed through this project provide sufficient information to pursue and construct the ecosystem component (see Section 4.3 and note within the Phase 2 roadmap on way forward for expanding to neighbouring countries).

10 FINAL REFLECTIONS

10.1 Project impact

BIRDIE is a data-to-decision pipeline with few equivalents in the biodiversity sector worldwide. Few countries have started to implement operational workflows for periodically producing high-level indices, including trends in population size and distributions of species at the national and subnational scale (e.g. the UK, Boyd et al 2023, Biological Reviews 98: 1492-1508). BIRDIE has put South Africa at the forefront in this regard. With the pipeline now functional, we have a base on which to build more sophisticated tools and reports going forward. Overall, stakeholders have encountered the project with excitement and have endorsed it as a good foundation with promise for the future. As a pilot for a complex data pipeline, the BIRDIE team considers the project a resounding success.

The project also considered as successes:

- the bringing together of a multidisciplinary team to combine several very distinct technical project elements, all critical to the success of the pipeline;
- the embedding of the project within (and together with) government and within the team responsible for reporting and development of national biodiversity indicators (a major advantage even though procurement within a government entity is not agile and requires persistence and planning);
- that this was a successful proof of concept that an externally developed, open source, partner website can be hosted by SANBI; and
- that the full range of stakeholders responsible in some way for the status of waterbird data in South Africa were consulted and engaged in the project.

The final stakeholder workshop in February 2024 was a critically important workshop, bringing together all the partners and representative stakeholders to reflect on the project achievements, discuss the overarching data needs and monitoring gaps and, following this, identify priorities for funding and implementation in a comprehensive way. This is captured in a workshop proceedings document. It would be extremely useful to revisit the outcomes of the final workshop as an initial task when picking up on Phase 2 of BIRDIE. The workshop confirmed the keen engagement of stakeholders at all levels of the data pipeline. It was agreed that effective biodiversity monitoring and reporting are essential for wetland and waterbird conservation efforts in South Africa, and that BIRDIE is making a contribution to addressing challenges to enhance data accuracy, accessibility, and usability, ultimately supporting informed decision-making and sustainable management practices. However, overall there is gap and an urgent national need for the development of an over-arching coordination strategy and field monitoring framework.

10.2 Lessons learnt and way forward

The workshop also confirmed the need for more work and funding across all BIRDIE project components. A BIRDIE Phase 2 project would add enormous benefits and pick up on the many components identified by the project team and during the stakeholder engagements (see also **9.2 Potential Future Updates**). It was agreed to initiate the development of one or multiple project funding proposals based on the national priorities identified during Workshop 3, and that future project lead organisations would be identified and lead the development of each proposal.

Notable lessons were learned, however, that will be valuable for future projects of a similar nature. They are summarised as follows:

Dedicated project coordinator for BIRDIE project and to a dedicated project coordinator to manage and grow the CWAC citizen science project: To effectively manage a project to improve and expand BIRDIE, would need a dedicated project coordinator. The long-term management and growth of the CWAC and SABAP datasets on which BIRDIE depends is also of vital importance, and it was envisaged that, at a minimum, this would require a dedicated CWAC project coordinator, as well as a number of regional support coordinators.

User feedback and engagement is a priority task: It shapes a project and ensures it is relevant through regular feedback on errors and gaps.

*Commencement phase takes approximately six months***:** The project commencement phase – i.e. the phase after a donor approves a project – takes longer than one anticipates during the project development phase. The initial contractual arrangements take substantial time. This commencement phase time period is an important consideration for any future projects.

User testing and roll-out was delayed substantially from the original plans, however, a pilot case study remains an important goal: It took a lot longer than originally anticipated (i.e. articulated in the project proposal phase) to develop the complex models, data pipeline and dashboard to be ready for showing to potential users. Originally we thought we would engage users earlier on, and built this into the project plan through a pilot case study, but ultimately took the decision that it was more important to have the working models and develop the MapViewer as far as possible. This departure from the initial workplan of having a pilot site prototype available for early user feedback did, however, have important consequences, as the team were not able to see the work of the data developer visualised until late in the project and so were unable to mock up site, species, and reporting pages until the fifth reporting period, setting back the dependent components of user testing, feedback loop, and uptake phase of the BIRDIE Project, which took place within a six-month extension period.

Challenges running the complex models for numerous species: During the course of the project, we realised that the models could only be run on a super-computer. The energy crisis in South Africa took its toll on BIRDIE's servers. When the back-up system also fails, running analyses stop, and they need to be resumed manually, with subsequent delays. For this project, the UCT High Performance Cluster was used, and later, an account was opened on a High Performance Computing Cluster from the National Integrated Cyber-infrastructure System (NICIS). Thanks to this extra flexibility we managed to run all planned analyses and are in a good position to promptly fix any problems with model outputs we may encounter.

Regular internal team engagement: In such a complex project, the team is dependent on each other for the necessary set of complementary skills. With the interdependence of the different parts, delays in one part of the project has knock-on effects on the entire pipeline, which makes effective project management challenging. Setting time for face-to-face discussions, regular focus topic discussions, and regular team meetings is extremely valuable. The Covid-19 pandemic placed an unforeseen burden on the team members, resulting in travel restrictions lasting well into 2022, and with fewer in-person contact sessions than originally planned. Francis (based in Belgium, at SSBE/RBINS) was only able to first visit Cape Town in November/December 2022, and we saw how necessary face-to-face workshops are in making progress with integrating the different parts of the pipeline. It would have been better to have such interactions sooner.

Ongoing updates and maintenance, and ongoing user support: We are concerned about the capacity for ongoing user support as well as updates and maintenance on the system. User uptake takes substantial time and likely results in ongoing ideas for updates of the system. We would like to maintain the capacity to expand and modify BIRDIE over the next few years, through a number of reporting cycles for some of our main stakeholders, e.g. officials reporting for RAMSAR. AEWA and State of Biodiversity. However, this depends on our ability to keep adapting BIRDIE for a while past the end of the current funding period. While we have delivered on the objectives in the proposal, we ideally would want to support users and adapt BIRDIE while it is being used in real-world applications, beyond testing.

Hand-over to project members that have not been involved in the development phase: As BIRDIE moves into the production phase, many of the systems put in place need to be actioned by new team members, who are in charge of updates and maintenance. This transition tests the systems themselves, as well as the documentation prepared for them. This is a great time to polish any outstanding functional issues and clarify the documentation.

11 REFERENCES

Albert, J.S., Destouni, G., Duke-Sylvester, S.M., Magurran, A.E., Oberdorff, T., Reis, R.E., Winemiller, K.O., Ripple, W.J. 2021. Scientists' warning to humanity on the freshwater biodiversity crisis. Ambio 50(1), 85-94. doi: 10.1007/s13280-020-01318-8.

Altwegg, R., Nichols, J.D., 2019. Occupancy models for citizen-science data. Methods in Ecology and Evolution 10, 8–21. https://doi.org/10.1111/2041-210X.13090

Barnard, P., Altwegg, R., Ebrahim, I., Underhill, L.G., 2017. Early warning systems for biodiversity in southern Africa – How much can citizen science mitigate imperfect data? Biological Conservation 208, 183–188. https://doi.org/10.1016/j.biocon.2016.09.011

Brooks, M., Rose, S., Altwegg, R., Lee, A.T., Nel, H., Retief, E., Reynolds, C., Ryan, P.G., Shema, S., Tende, T., Underhill, L.G., Thomson, R.L., 2022. The African Bird Atlas Project: a description of the project and BirdMap data-collection protocol. Ostrich 93, 223–232.

Bussière, E.M.S., Underhill, L.G., Altwegg, R., 2015. Patterns of bird migration phenology in South Africa suggest northern hemisphere climate as the most consistent driver of change. Global Change Biology 21, 2179–2190. https://doi.org/10.1111/gcb.12857

Cervantes, F., Altwegg, R., Strobbe, F., Skowno, A., Visser, V., Brooks, M., Stojanov, Y., Harebottle, D.M. and Job, N., 2023. Birdie: A data pipeline to inform wetland and waterbird conservation at multiple scales. *Frontiers in Ecology and Evolution*, *11*, p.1131120.

Devictor, V., Whittaker, R.J., Beltrame, C., 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. Diversity and Distributions 16, 354–362.

Duckworth, G.D., Altwegg, R., 2018. Effectiveness of protected areas for bird conservation depends on guild. Diversity and Distributions 24, 1083–1091. https://doi.org/10.1111/ddi.12756.

Harebottle, D. M., & Underhill, L. G. (2016). Assessing the value of wetlands to waterbirds: exploring a population-based index at flyway and regional levels. *Ostrich*, *87*(1), 7–21. https://doi.org/10.2989/00306525.2015.1104396

Hugo, S., Altwegg, R., 2017. The second Southern African Bird Atlas Project: Causes and consequences of geographical sampling bias. Ecology and Evolution 7, 6839–6849. https://doi.org/10.1002/ece3.3228

International Union for Conservation of Nature & BirdLife International. 2021. *SDG indicator metadata.* January 2021. Available: https://unstats.un.org/sdgs/metadata/files/Metadata-15-05-01.pdf [2021, August 6].

Isaac, N.J.B., van Strien, A.J., August, T. a., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods in Ecology and Evolution 5, 1052–1060. https://doi.org/10.1111/2041-210X.12254

Johnston, A., Matechou, E., Dennis, E.B., 2022. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. Methods in Ecology and Evolution 2022, 1–14. https://doi.org/10.1111/2041-210X.13834

KBA Standards and Appeals Committee. 2020. *Guidelines for using A Global Standard for the Identification of Key Biodiversity Areas. Version 1.1*. Gland, Switzerland: IUCN. Available: https://portals.iucn.org/library/sites/library/files/documents/2020-033-En.pdf [2021, August 27].

Malherbe W., Ferreira M., Van Vuren JHJ., Wepener V., Smit NJ. 2017. The aquatic biodiversity and tourism value of selected South African Ramsar wetlands. Unit for Environmental Science and Management, North-west University. Potchefstroom. WRC Report No. TT 732/17. ISBN 978-1-4312-0920-0.

Marnewick D., Stevens C., Antrobus-Wuth R., Theron N., Wilson N., Naude K., Jonas H. 2020. Assessing the Extent of OECMs in South Africa: Final Project Report. BirdLife South Africa, Johannesburg. Available: https://www.birdlife.org.za/wp-content/uploads/2020/11/OECM-Report-2020.pdf [2021, August 26].

Massicotte P, South A (2023). rnaturalearth: World Map Data from Natural Earth. R package version 0.3.4, https://CRAN.R-project.org/package=rnaturalearth.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Secretariat of the Ramsar Convention on Wetlands. 2021. SDG indicator metadata, March 2021. Available: https://unstats.un.org/sdgs/metadata/files/Metadata-06-06-01b.pdf [2021, August 6].

South African National Biodiversity Institute. 2018. *National Biodiversity Assessment (NBA) 2018 flyer.* Available: https://www.sanbi.org/wp-content/uploads/2018/04/20171016-nba-flyer-optimised.pdf [2021, August 6].

Stats SA. 2019. Sustainable Development Goals (SDGs): Country Report 2019. Pretoria. Available: http://www.statssa.gov.za/MDG/SDGs_Country_Report_2019_South_Africa.pdf [2021, August 6].

United Nations Environment Programme. 2021. SDG indicator metadata, July 2021. Available: https://unstats.un.org/sdgs/metadata/files/Metadata-06-06-01a.pdf [2021, August 6].

Van Deventer, H., Smith-Adao, L., Collins, N.B., Grenfell, M., Grundling, A., Grundling, P-L., Impson, D., Job, N., Lötter, M., Ollis, D., Petersen, C., Scherman, P., Sieben, E., Snaddon, K., Tererai, F. & Van der Colff, D. 2019. South African National Biodiversity Assessment 2018: Technical Report. Volume 2b: Inland Aquatic (Freshwater) Realm. CSIR report number

CSIR/NRE/ECOS/IR/2019/0004/A. South African National Biodiversity Institute, Pretoria. http://hdl.handle.net/20.500.12143/6230.

Wetlands International (2023) Migratory waterbird species decline 'canary in the coal mine'. Ede-Wageningen, NL, 13 May 2023. https://www.wetlands.org/migratory-waterbird-species-decline-canary-in-the-coal-mine/

Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16, 446–453. https://doi.org/DOI: http://dx.doi.org/10.1016/S0169-5347(01)02205-4

12 APPENDICES

12.1 APPENDIX A. Conferences, meetings and other engagements

Date	Event	Presenter	Description
24 February 2021	Workshop 1 (User Requirements)	11 project team + 20 key stakeholders - national government; provincial government; NGOs and civil society.	brief stakeholders about the JRS Biodiversity Foundation 'BIRDIE' project; give an overview of different types of indicators this data pipeline might be used for, and provide a conceptual design overview of the pipeline to stakeholders; brainstorm user needs for the freshwater biodiversity data pipeline
1-3 June 2021	Internal working session 'Workshop 2' ('Technical Working Group')	Nancy, Andrew, Res, Francis, Doug, Vernon, Sediqa,	a series of technical discussions and working sessions convened over 3 days with the full team organising into smaller working groups to solve specific tasks, coming together to feed back as a team at the end of each day.
June 2021	External meeting	Andrew, Sediqa + state of Environment Author team	National authors meeting for state of environment reporting (arranged by DFFE), SANBI presented on BIRDIE project development of indicators for wetland birds and wetland condition, we discussed how these will inform RAMSAR and AWEA reporting (also potentially SDG 6).
June 2021	Internal meeting	Andrew + head of the threatened species unit	Discussed the utility of BIRDIE monitoring data for bird species abundance trends that can inform red listing efforts. BirdLife SA will lead this redlisting aiming for 2024 publication (2015 was last version).
2 July 2021	Barberspan stakeholder meeting 'Workshop 3'	Technical leadership of the project + key stakeholders from Barberspan	Main aim was to briefly introduce the project and then dive into a focus on Barberspan-specific user needs for the data pipeline
2021/07/07	Webinar series: Freshwater Bioinformatics in Africa	Nancy Job Francis Strobbe Michael Brooks Francisco Cervantes/Res Altwegg/Vernon Visser	Introduction to the BIRDIE Project Overview of the proposed data pipeline CWAC/SABAP datasets and APIs Overview of statistical analyses for the data pipeline
30 May - 3 June 2022	Brussels trip	Pachi, Francis, Yvan	Pachi travelled to Brussels to work with Francis and Yvan to make back- and front-end meet. Set up dashboard and database structure. Defined model output formats.
27 June - 1 July 2022	ISEC 2022	Res, Pachi, Michael, Vernon, Maphale, Gareth...	Presentation and Poster: Pachi on ABAP and the BIRDIE project
2021/09/14	Webinar series of the Society of Spanish Researchers in South Africa and Embassy of Spain in South Africa	Francisco Cervantes	Statistical Ecology: Hacking and Conservation of Ecological Systems
2021/10/22-24	National Wetlands Indaba	Francisco Cervantes	South African Biodiversity Data Pipeline for Wetlands and Waterbirds (BIRDIE)
23-24 March 2022	Joint Nature Conservation Committee (JNCC) workshop on African Bird Monitoring	Res Altwegg Michael Brooks Doug Harebottle	Statistical modelling aspects of the BIRDIE project BIRDIE project illustrating the value of on-the ground data for decision making Attendee
2022/04/4-8	European Bird Census Council (EBCC) 22nd Conference	Michael Brooks	BIRDIE project overview during round table discussion of the African Bird Atlas Project (ABAP)
2022/04/8-10	African Bird Atlas Project (ABAP) Workshop	Michael Brooks	SABAP 2 component presentation of ABAP, data uses, tools development, and path to policy
15 May 2023	DFFE, BirdLife	Nancy, Pachi, Jackie Jay, Stanley Tshitwamalumoni, Ernst	Meeting held in person in Pretoria to review reporting pages and platform, platform not updated since previous meeting with Ernst, so we focussed on

		Retief, Melissa, Humbu Mafumo	mocking up and visioning what we would like to see included on the reporting pages
16 May 2023	SABAP2 steering committee meeting	SABAP2 steering committee	Updated committee on BIRDIE and briefly discussed need to ensure longevity of SABAP and CWAC
2022/06/07	Ramsar Working Group	Doug Harebottle and Nancy Job	BIRDIE project and Waterbird Conservation Value index
2022/06/27	The International Statistical Ecology Conference	Francisco Cervantes	Delivering policy-relevant indicators: South Africa's Data Pipeline for Wetlands and Waterbirds Poster: ABAP: An R package to access African Bird Atlas Project data" (Francisco Cervantes, Michael Brooks, Dominic Henry and Res Altwegg).
2022/08/15-19	International Ornithological Congress	Francisco Cervantes	Poster: BIRDIE: South Africa's Data Pipeline for Wetlands and Waterbirds
2022/09/5-9	MEDECOS 2022	Nancy Job	BIRDIE: South Africa's Data Pipeline for Wetlands and Waterbirds
2022/11/15-21	Pan African Ornithological Congress	Nancy Job	Workshop on multilateral environmental agreements – the BIRDIE project
2022/11/28-30	Symposium on Biodiversity Informatics	Francisco Cervantes, Francis Strobbe Andrew Skowno	BIRDIE wetlands and waterbirds data pipeline: from data to decision making Attendee
2023/04/17	Department of Statistical Sciences UCT Seminar Series	Francis Strobbe	BIRDIE: a (first) data analysis pipeline for South Africa
2023/05/08	Department of Statistical Sciences UCT Seminar Series	Francisco Cervantes	BIRDIE: streamlined waterbird population models
2023/06/26-30	SANBI - Training course - Analytical techniques in biodiversity Big Data using GBIF: Making an impact	Doug Harebottle	Conservation Planning lecture, using the BIRDIE framework to show how big data from CWAC is ultimately used in conservation outcomes at the end of the pipeline.
2023/07/4-6	National Symposium on Biological Invasions	Vernon Visser	Experiences from other fields of natural science in transforming data into information
2023/07/24-27	International Congress for Conservation Biology	Res Altwegg	Preparing data for decisions: BIRDIE: An online platform for African wetland and waterbird information
2023/09/11	Ramsar National Committee	Doug Harebottle and Nancy Job	BIRDIE project and Waterbird Conservation Value index
21 August 2023	Department of Statistical Sciences UCT Seminar Series	Gareth Adwards	Using state-space time series analysis on wetland bird species to formulate effective bioindicators in the Barberspan wetland
19 September 2023	Ramsar Working Group	Doug Harebottle and Nancy Job	BIRDIE project and Waterbird Conservation Value index
26-28 September 2023	Sol Plaatje University, School of Natural and Applied Sciences, Annual Research Symposium and Postgraduate Expo	Doug Harebottle	Preparing data for decisions: BIRDIE: An online platform for African wetland and waterbird information. Recycled the poster presentation developed by Res for the presentation
14 December 2023	Department of Mathematical Sciences, Norwegian University for Science and Technology, Trondheim, departmental seminar series	Res Altwegg	Mapping Biodiversity in Remote Areas
4 December 2023	Multidisciplinary Workshop - Green Transition & Biodiversity	Francis Strobbe	BIRDIE: A South Africa biodiversity data pipeline for wetlands and waterbirds Decision making in the biodiversity sector is only as good as the data that underpins the science.
6-7 February 2024	Enhancing waterbird and wetland priority actions for South Africa 'Workshop 4'	See workshop proceedings report	The final workshop of Phase 1 of the 'BIRDIE Project' (see report)

16 February 2024	5th State of Environment Information Community of Practice meeting (SEI-CoP 2024)	Nancy Job and Sediqa Khatieb	"BIRDIE: an online platform for African wetland and waterbird data". Note: The SEI-CoP is an annual meeting that allows experts to share their knowledge and experience on the state of the environment information provision and associated systems. The SEI-CoP represented a great opportunity to showcase the innovative work that the BIRDIE has done with regards to making citizen science data discoverable for informed decision making.
16 February 2024	Meeting with BirdLife and DFFE on information outputs	Jackie Jay (DFFE), Melissa Lewis (Birdlife SA), Carol Poole (SANBI), Emily Botts (Independent Editor), Kyle Lloyd (Birdlife SA), Jessica Wilmot (Birdlife SA), Carina Pienaar (Birdlife SA)	Brainstorm and share content for use in potential site posters and general infosheets (wetlands and waterbirds, Ramsar, migratory waterbirds, wetland site management)
19 February 2024	Department of Statistical Sciences UCT Seminar Series	Res Altwegg	SEEC Research Highlights
1 March 2024	An introduction BIRDIE project for BirdLife SA staff	Hosted by Alan Lee.	AI meeting report generated
6 March 2024	National Ramsar Committee Meeting	Hosted by DFFE, 40+ ramsar site managers and stakeholders	Discussed progress of the BIRDIE project and
7 March 2024	BIRDIE-Kenya	Colin Jackson, Nancy Job, Pachi Cervantes, Michael Brooks, Res Altwegg	Meeting to discuss the current BIRDIE functionality with Colin (Kenya lead) and to discuss how the pipeline could be adapted for Kenya
8 March 2024	National Ramsar Report meeting	Nancy Job, Jackie Jay, Kristen Muthady, Stanley Tshitwamulomoni	The contribution of the BIRDIE project was discussed during this larger project to prepare for the Country Report, to be authorised by the Minister of Forestry, Fisheries and the Environment in August 2024 then delivered to Ramsar in October

12.2 APPENDIX B. Key project outputs

Data Product	Key Users	Use	Links
BIRDIE **web platform**	Wetland site manager Government official Environmental professional Birdwatcher or tour guide Researcher or academic	Bird indicators such as abundance, distribution and species richness. Ancillary data such as temperature, precipitation and surface water. Trends of how species abundance has changed over time. Summaries of species and condition at many monitored wetland sites. Raw data from the source datasets and models to integrate into individual user workflows and analyses. See also site reports, species reports, Ramsar page.	https://birdie.sanbi.org.za
ABAP R **package**	Researchers Scientists Students Anyone wanting access to raw data	It allows users to download data from the African Bird Atlas Project directly from R. It also provides functionality to annotate these detection/non-detection data with environmental variables from the Google Earth Engine repository, making their analyses more accessible and reproducible. As an example, James Swigler used the ABAP package to work with ABAP data for his Masters thesis at UCT: https://open.uct.ac.za/bitstream/handle/11427/37377/thesis_sci_2022_swingler%20james.pdf?sequence=1&isAllowed=y Other postgraduate students and researchers are using the package. It will also be an essential tool for facilitating the red-listing assessment of the birds of South Africa.	https://github.com/AfricaBirdData/ABAP
CWAC R **package**	Researchers, scientists, students and anyone wanting access to raw data	It allows users to download data from the Coordinated Waterbird Counts project directly from R. Postgraduate students and researchers are using this package to access CWAC data.	https://github.com/AfricaBirdData/CWAC
Site report page (BIRDIE web platform)	Researchers Scientists Students Anyone wanting access to prepared report pages on CWAC sites	This page is a resource for site managers, researcher, and bird enthusiasts interested in wetland ecosystems. Here, you will find detailed summaries of various wetland sites, including their condition, the bird species they support together with their rate of change over 5 and 10 years, a total species count and a list of threatened species per site. With access to ancillary data, such as catchment area and management information, and the Waterbird Conservation Value (WCV) at the site and species level, you can gain a comprehensive understanding of these habitats and the trends affecting their health. Users can download the contents of tables within the report page via csv	https://biodiversityadvisor.sanbi.org/contentmanagement/index?guid=7613ac09-0b2f-46b8-adb5-0d441795e709
Sites **master list**		This excel spreadsheet supplements the automation of data pulled from CWAC	
Species report page (BIRDIE web platform)	Researchers Scientists Students Anyone wanting access to report pages on species	Similar to the site reports page, this page is a resource for site managers, researcher, and bird enthusiasts interested in wetland ecosystems. Here, you will find detailed summaries of all the analysed species, in which CWAC sites they occur, their rate of change over 5 and 10 years, and what is their area of occupancy over the years.	https://biodiversityadvisor.sanbi.org/contentmanagement/index?guid=42305260-12b0-4a92-912e-7d801d15fdac

		Users can download the contents of tables within the report page via csv	
Species **master list**		This excel spreadsheet supplements the automation of data pulled from CWAC	
Ramsar reports (BIRDIE web platform)	Researchers, Scientists, Students, Ramsar stakeholders and anyone wanting access to a waterbird focussed Ramsar report for South Africa	This page has outputs and functionality that can be used by our national government entity (DFFE) for their international reporting needs. It is a resource for anyone interested in a complete view of the most current status of a Ramsar site in South Africa, including a lot of parameters, variables and descriptions, and for each Ramsar site also being a CWAC site the latest WCV at the site level.	https://biodiversityadvisor.sanbi.org/contentmanagement/index?guid=8830c563-511a-4d76-80c2-9336e6054d44
Ramsar sites **master list**	Allows the team to view, edit, visualise and manage future updates to the information.	This excel spreadsheet supplements the automation of selected species and site information, in support of Ramsar sites	Available from the Resources section of the BIRDIE Platform.
Raw data behind the dashboard visualisation (BIRDIE web platform) *See Appendix C / 12.3 for a list of available APIs	Researchers / scientists / students and anyone wanting access to raw JSON data	This UI allows users to dig deeper into the raw analysed data or summary and overview information on sites and species. Through a simple interface, a user can retrieve all the data that is presented on the website or map viewer and use it for his/her own analysis or reporting. The FAQ explains how to query the API and download the results as a .json file, GeoJSON file or a URL. The FAQ section also explains which browser is the best to use for viewing a URL, and how to view the GeoJSON files. The FAQ section contains descriptive text (instructions), along with images.	https://birdieapp.sanbi.org.za/birdie/swagger-ui/
FAQ Frequently asked questions	Anyone using the web platform	Documented frequently asked questions, initially prepared by the team, supplemented by user feedback	https://biodiversityadvisor.sanbi.org/contentmanagement/index?guid=1250f663-1b8a-4851-9387-93b92d4795da
Journal paper	Researchers, scientists and scientifically minded public	"BIRDIE: A data pipeline to inform wetland and waterbird conservation at multiple scales" Total views: 1834, downloads 147 (per journal site accessed 4 Sept 2023). This scientific publication captures the essence of the project and it is a fundamental reference for anyone using the BIRDIE systems. It does need to be complemented with all the documentation produced for the website, though, because BIRDIE has seen technical development since the paper was published.	https://www.frontiersin.org/articles/10.3389/fevo.2023.1131120/full
Journal paper	Researchers / scientists / scientifically minded public / anyone using the ABAP data	"The African Bird Atlas Project: a description of the project and BirdMap data-collection protocol" Total views: 719, CrossRef citations 5, Altmetric 4 (according to journal web site accessed 28 Feb 2024), 16 citations according to Google Scholar. This paper presents the ABAP data set and is envisioned to be the primary reference for anyone using ABAP data.	https://www.tandfonline.com/doi/abs/10.2989/00306525.2022.2125097
Note: also see 7.3 for links to communication and awareness raising outputs from the project			

12.3 APPENDIX C. APIs

API	Explanation
Abundance	**Access Occupancy model data**
GET/openapi/v1/ratio_est_sites/{sites}/{species}/{ratio}/{start_year}/{end_year}	Get the proportion or slope modeled information on species at a site, over a range of years
GET/openapi/v1/ratio_estcilower_sites/{sites}/{species}/{ratio}/{start_year}/{end_year}	Get the proportion or slope lower confidence intervals modeled information on species at a site, over a range of years
GET/openapi/v1/ratio_estciupper_sites/{sites}/{species}/{ratio}/{start_year}/{end_year}	Get the proportion or slope upper confidence intervals modeled information on species at a site, over a range of years
GET/openapi/v1/season_count_sites/{sites}/{species}/{season}/{start_year}/{end_year}	Get the counts on species at a site, over a range of years
GET/openapi/v1/season_est_sites/{sites}/{species}/{season}/{start_year}/{end_year}	Get modelled estimates on species at site, over a range of years
GET/openapi/v1/season_estcilower_sites/{sites}/{species}/{season}/{start_year}/{end_year}	Get modelled lower confidence intervals on species at site, over a range of years
GET/openapi/v1/season_estciupper_sites/{sites}/{species}/{season}/{start_year}/{end_year}	Get the modelled upper confidence interval on species at site, over a range of years
Distribution	Access State-Space model data
GET/openapi/v1/psi/{pentad}/{species}	Get the predicted occurrence of a species in a pentad (estimate 'est', median 'med', upper bound 'up', lower bound 'lb')
Species	**Access species information data**
GET/openapi/v1/guild	Get the guilds and species under that guild
GET/openapi/v1/list_species	Get the list of species used
Variables	**Access ancillary data and additional variables**
GET/openapi/v1/rec/geojson/pentad/{parameter}/{year}	Get data for ancillary variables
GET/openapi/v1/year_list	Get the list of years
API	**API service up**
GET/openapi/v1/ok	
Dev	**API in development**
GET/openapi/v1/{site_convention}/{site_type}/{site_group}	Get the list of sites filtered by convention (AEWA...), type (wetland, dam...) and province
GET/openapi/v1/realoccu/{pentad}/{species}	Get the estimated occurrence of a species in a pentad
GET/openapi/v1/ssm/{sites}/{species}/{season}/{start_year}/{end_year}	Get the state-space model output on species at a site, over a range of years
GET/openapi/v1/ssm/{site_group}/{species}/{season}/{start_year}/{end_year}	Get the state-space model output on species for a province, over a range of years
Schemas	
site_group {province}	string example: ALL, Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North West, Northern Cape, Western Cape
year_list {yearlist}	integer example: 2008