

**Biodiversity information database systems training and data entry Workshop proceedings held at
Lion's court in Nanyuki on 4th April 2018**



Prayer was led by Luke Lukaria

Members briefly introduced themselves

Introduction of the program of activities by Apollo Kariuki

The GBIF Project gives us an opportunity to organize our datasets. We want to have an information system that makes available and accessible our datasets.

Objectives of the workshop

1. To understand the system and we expect everybody's input in making the system more viable in all the areas involved in research. We would like to come up with something that can be adopted across board
2. To standardize our information which has been collected in different formats. During the workshop the members will be taken through the process of standardizing this datasets
3. To upload this information in our database system.

Workshop agenda

Session I: Workshop objectives

Session II: Biodiversity information system

Session III: Standardizing datasets

Session IV: Uploading datasets

GBIF – enables free and open access to data

Datasets identified by scientists include:

1. Species checklist
2. Climate monitoring
3. Wildlife monitoring
 - aerial census
 - Waterfowl census
 - Ground census
 - Tsetse monitoring
 - Collared animal monitoring
4. Land use land cover monitoring
5. Animal mortality
6. Human wildlife conflict monitoring
7. Invasive species monitoring
8. Water quantity and quality monitoring (ECA, Nakuru)
9. Marine monitoring (SAM Project)
10. Visitor monitoring

Comments by members

Recent development has been that the aerial census is currently not capturing all the mammals as was done in the past. Previously, the census incorporated all other mammals but in the present regime from 2016, we've reverted back to focusing on 2 or 3 species. So we need to check and see how we can monitor the population trends of these other mammals such as the eland, the wildebeest etc

We also need to capture the monitoring of carnivores in the database because their methods of monitoring are different as well as look into other species of special concern such as the roan antelope, the sable antelope.

We need to identify these problems, check the inconsistencies in the datasets and come up with recommendations

We need to focus on key conservation targets such as the special species – roan, sables, grevy's especially the ones that we already have good data on it and organize that dataset.

At any one time there is always a project going in within the organization and the day the project stops the monitoring stops. We need to handle these problem to ensure consistency of the datasets. We can

always maximize on these projects and have the information readily available. With the GBIF, we need to get these datasets and have them on the database system.

Presentation by peter Hongo

Data required by GBIF and the status on KWS Data

What is GBIF – Global Biodiversity Information Facility, aims to make biodiversity data accessible anywhere in the world

What is KWS – mandate to conserve and manage Kenya’s wildlife and its habitat

The relationship between GBIF and KWS – GBIF provides the funds for KWS to manage and make available its datasets

Biodiversity datasets

- Occurrence data – describes the location in time and space
- Checklist
- Sampling events data eg. birds inventories

Biodiversity data standards

Datasets from different origins adapting a similar schema

Data quality – involves data management, data modelling and analysis quality control, quality assurance, data storage, data presentation

Data cleaning and validation - Involves cleaning the data and it’s a delicate process which must be carefully carried out. Fixing the errors must be carefully done

Examples of datasets required by the GBIF Project

- Aerial census datasets (Amboseli-Kilimanjaro, Laikipia-Samburu, Tsavo-Mkomazi)
- Ground census
- Marsabit Forest Ecosystem inventories
- Mt. Kenya Forest Ecosystem inventories
- Zoology
- Kenya bird map
- Botany
- Important Bird Areas datasets

Clean datasets are required to ensure the database system is working properly

Comments

The GBIF project is more localized with specific needs

We have not included data from the veterinary services

We will eventually want to include, information management, security data, visitor data

The data scale- currently organizing the existing data, old data

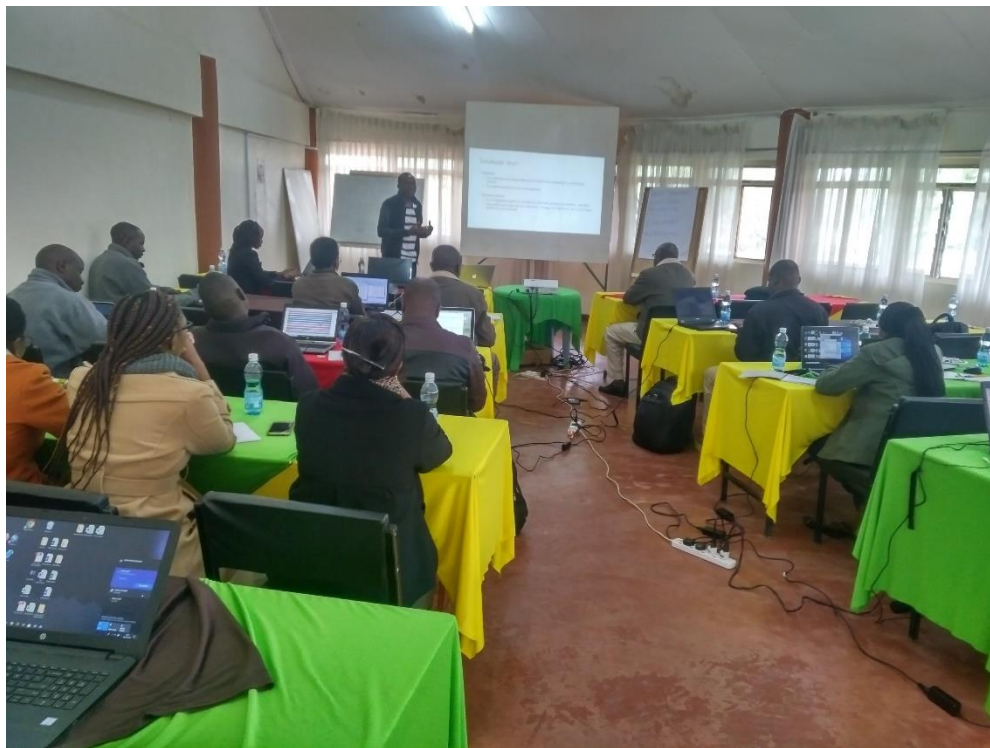
This data if put on the GBIF Website, will be available and accessible and will be known that it is from KWS unlike someone having the data and keeping him to himself

So many people have data and they all have different formats; the GBIF Project can help us standardize these datasets into an agreeable format. – When a scientist is transferred to a different location, this database should be part of the handing over.

The objective of BHLA is similar to the GBIF project. This is then the best forum to make accessible the datasets that we have as an organization. In the near future when BHLA is mentioned will be more aware if the objectives which would then be an opportunity to be visible worldwide

Presentation by Ken Otieno

Database – collection of data put together in a particular format which helps in optimization of the data



Database system – includes the users, the hardware, the software, the data. It defines and regulates the data and the users

We are currently using a file system in the database- drawbacks:

- Data isolation – data scattered in various files
- Duplication of data – everybody has this data and it's not

helpful to someone who doesn't have it, and data redundancy

- Dependency on application programs – if the data is in a particular format such as excel, then it's not accessible to someone who doesn't have that particular program

Hardware, software, people, procedures and standards

The database is user- centered

Functions of the database systems

1. Stores information
2. Avoids redundancy and ensures data integrity

3. It hides the complexity involved
4. It enforces data integrity
5. It implements data security in that we have access levels in the system

The web databased database design

This makes everybody have access to the system

The database system (Ken explained the progress of developing through a demonstration)

Comments by members

About importing data – is the system flexible to accommodate different datasets or does it require a single data format. The system has been developed based on a set standard. That means the datasets have to be reformatted to a standardized format

In census for example, we have different block formats, would that present a problem. **Ans:** the doesn't seem to be an issue

The species list should be arranged alphabetically

The county data may have other datasets that may not be useful to KWS, if you are interested in what data the county has from KNBS, that data might not available in the KWS Database. Secondary data, would be useful can be acquired and recommendations should be made and the data acquired.

The flexibility of the system, the species name is so wrong, if it is possible to classify the species in broad classes first –**ANS:** the long drop down list and we will classify the species In broad classification. Its work in progress

Importing the data, the data has to be copied in a particular folder and with a suggestion that the system has an option of browsing to a particular folder where the data is stored – **ANS:** as it is, there is an option to browse to the system. Once the system is hosted online, it is possible to do that.

The base map; we requested the availability of a base map showing the imagery of the almost the whole country. This would make it possible to overlay the datasets on the particular regions where they are found. **ANS:** Possible and will be done

Is the database privileges of accessing the deleting option – how does it work? It can be very dangerous because the decision is left to one person's privileges and permissions. **ANS:** if someone has no rights to delete then they cannot access that part. The system has logs of who did what in the system. The database has automatic backups and we can set the backups in a different system

Presentation by Peter Maina

He took the members through a data standardization template.

Comments by members

The collared animal monitoring –are we able to get the data in an excel format? **ANS:** The datasets depends on the institutions managing the collared data. Eg. IFAW Gives full access to the data which can be downloaded in excel. Institutions like STE don't give access or give access to limited data.

Coliform needs to be accessed in the lab – if it is not captured when the rest are being done then it can be left out in the database

Land use land cover change classification classes – It's safer to use a classification body that has authority and one which deals with classification of land use land cover. National regional and site specific areas. DRSRS has published the classification classes they use

Checklist – we need to add species, genus, family, order, class and common name – (in that order) the scientific name can be retained

Aerial survey data, - in ground census, the blocks need to be defined. In aerial survey data, blocks do not make much sense in analysis, they are meant to help in navigation and data collection. The PA and CA are more relevant datasets. The blocks change every census and analysis is done according to PAs, CAs, dispersal areas. The blocks should therefore be removed in aerial census or left blank. It could be captured in the metadata.

Data cleaning exercise during the workshop

Lake Nakuru National Park – Timothy, Maina, Ndambuki, Kuloba, Jeniffer, Becky

Marsabit - Jackline, Makau, Joseph, Kimitai, Lala, Anastasia

Mt. Kenya – Faith, Kanyi, Mathenge, Ken Opiti

Aerial Census – Hongo, Maina, Jackuiline Bernard, Wasike, Grace



Final session – recap and progress

Let's think of a catchy name for the database system

Mathenge will follow up with NMK for the rest of the data for Mt. Kenya ecosystems especially records of invertebrates

Hongo to sort out the Marsabit ecosystem dataset

Grace will gather all the data and share

Next steps:

Try to find out how we can publish the data in the GBIF node

Any ideas on how to improve the system can be shared in this forum