Biodiversity Data Cleaning Workshop proceedings held on Tuesday 20/03/2018 to Wednesday 22/03/2018 at the Yellow Green Hotel in Naivasha



Figure 1: Participants in the Data Cleaning Exercise

Day 1

Objective I

Introduction of participants, workshop objectives and opening remarks. This session was led by Mr. Apollo Kariuki – Ag. Head Planning and Environmental Compliance

Since KWS was formed, the organization has been attempting to organize data and this project is key to assist in achieving that objective. Over time, the organization realized that there is no much data and the little data that exists is not so organized and centralized. This project comes in handy in providing for this.

Once organized, the data will support a lot of research activities both within KWS and even abroad. The database will also be key in supporting management decisions. In future, the organization would also want to organize the spatial database and make the data readily available.

The database has to be shareable, timely, available, and accessible which can be a model that can be replicated at the ecosystem level. So, we were all encouraged to put all the effort needed to achieve the objectives in this workshop.

Objective II

Presentation on the Biodiversity Database Information System and its contents. This session was presented by Ken Otieno – The database Developer

The database is an online based database system which can be accessed from anywhere with internet connection. The database has some security measures.

Security measures

- Main access
- Log in place to access the privileges

Access levels - User is created and defined depending on what you want to do. E.g. data entry forms access for a data entry person

The Access is – user name and password

Datasets - need to be made available at the end of the workshop

Presentation on how to input data in the database was made and the key points to note were:

- The fields /columns required for each of the data entry targets
- The system has a way of checking for data duplication
- During data entry, You can only handle one dataset type at a time
- There will be a 'invitation to locate the folder where the data is stored' before it is uploaded
- The import button is for the bulk datasets
- There's a provision for uploading a photo in the checklist
- The system will also be able to generate reports of lists and graphs
- The system will be able to use some 'artificial intelligence' to combine two datasets to observe trends e.g. Animal location with reference to rainfall distribution
- The system in future will be able to incorporate the spatial analysis aspect

Comments to the presentation

- The system is good work
- The mandatory fields are useful when it comes to data entry
- The GPS Coordinates should be in a standardized format X,Y or lat, long
- The issues of the base maps either on open street map or any other and the possibility of overlaying any other layers such as vegetation would be an added advantage for the database
- We need to think of other ideas we can input in the database

Objective III



Unveiling of the data format spread sheet by Ken Otieno

Figure 2: Participants working on the data cleaning excersie

Objective IV

Presentation on the data required by GBIF project and status of the existing datasets. This was done by Peter Hongo

He explained the relationship between GBIF and KWS

KWS undertakes management and conservation of wildlife resources within the country while GBIF (Global Biodiversity Information Facility) focuses on making biodiversity information data available to support sustainable development

GBIF provides funds to organize and mobilize Biodiversity Information from KWS

Examples of data sets required

- Aerial census (Amboseli, laikipia Samburu, tsavo mkomazi, mara ecosystem) 10,000 records
- Ground census (Lake Nakuru national park, Nairobi national park) 2000 records at least
- Inventories (Marsabit forest ecosystem 718 records (plants 298, large mammals 4, small mammals 8, and invertebrates 199) Mt. Kenya ecosystem - 637 Records, avifauna, herpetofauna, invertebrates)

The kind of datasets the organization currently has requires some cleaning to a standardized format to make it easier for input into the database and that brought us to the main objective for this workshop.

Objective V

Data cleaning -Roles divided into the following themes;

- Aerial census Tsavo-Mkomazi, Laikipia-Samburu and Amboseli-West Kilimanjaro ecosystems (Peter Hongo, Grace and Zachary)
- Ground census NNP and LNNP (Peter Maina and Gideon)
- Inventories Marsabit and Mt. Kenya (Jackline Mutwiri and Faith)



Figure 3: Faith Muchiri presenting her progress report in the data cleaning Exercise

Day 2

The data cleaning exercise continued and at the end of the day, all participants presented their progress report.

Way forward

- The data cleaning exercise was quite involving and successful.
- Some changes were made in the data format depending on the existing data and this will be forwarded to Ken for changes in the database.
- Participants were asked to finalize the data cleaning exercise in the shortest time possible.