FBIP training on Data analysis

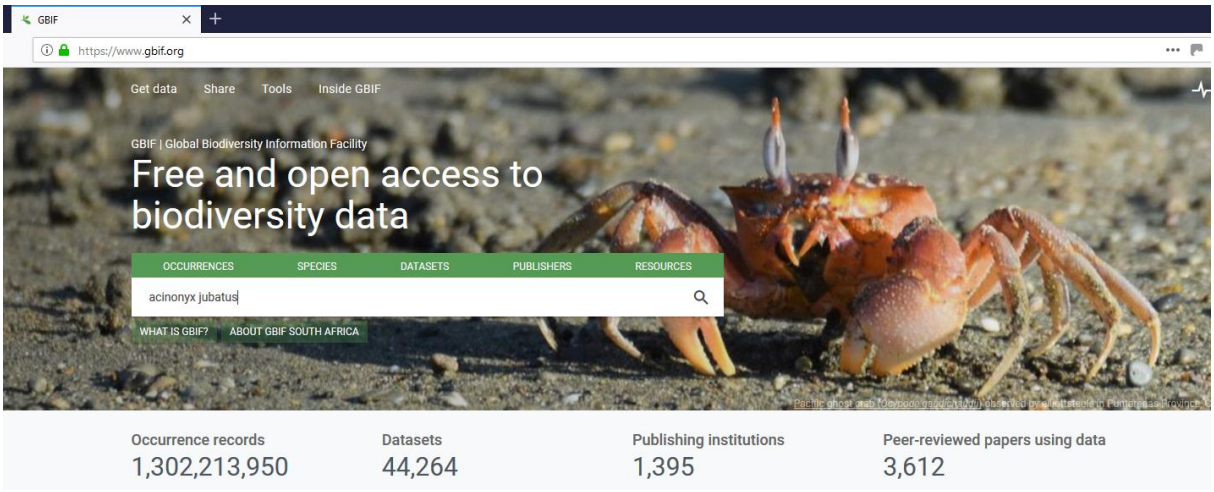
October 2019

Practical exercises on GBIF

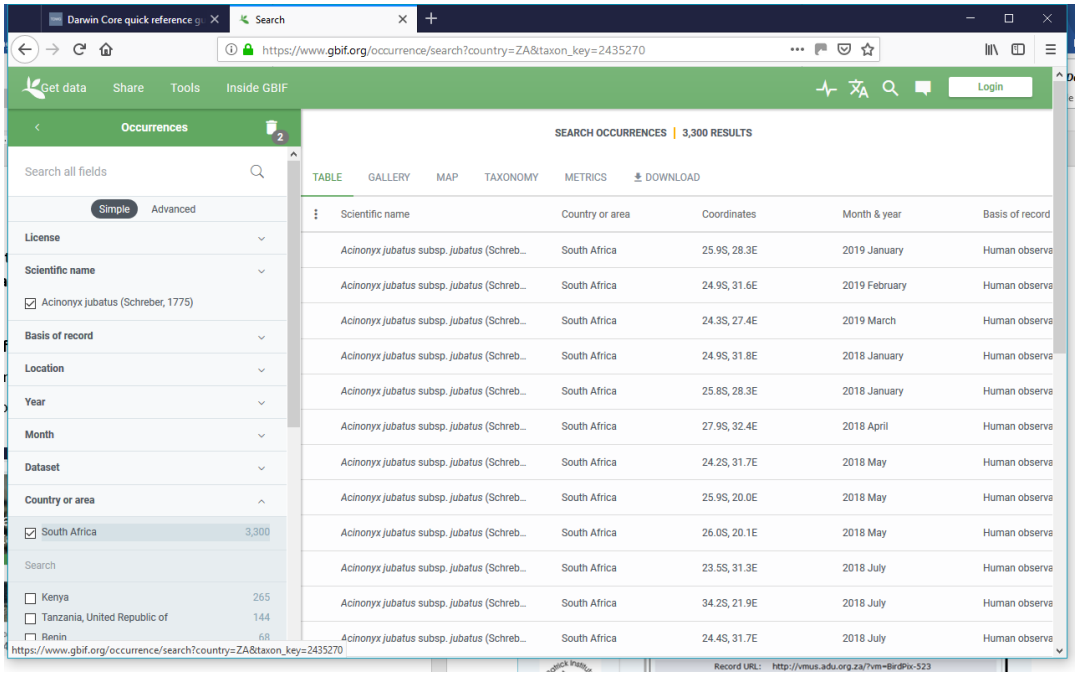
Exercise 1. Searching for data on GBIF

a) In your web browser, open www.gbif.org

b) Type in a scientific name of a species (e.g. *Acinonyx jubatus*) and search for occurrences



c) Apply various filters, e.g. country, year, or basisOfRecord, using the tickboxes on the left of the screen.



d) Continue to explore the data and the website

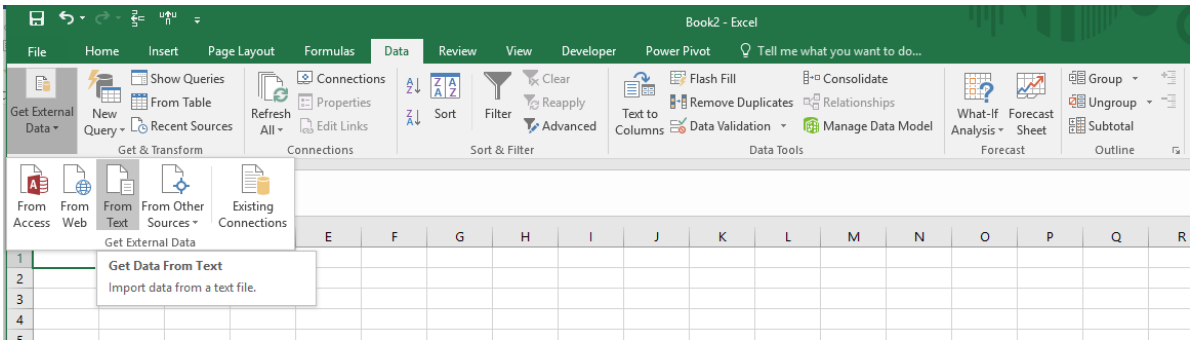
Note: You will need an account to be able to download data, so we'll now explore a dataset that has already been downloaded for you.

Exercise 2. Opening datasets that are downloaded from GBIF

a) The following file: 0001780-190415153152247.zip has been downloaded directly from GBIF.

b) **Unzip** the file (you can usually do this by right-clicking on the file name and selecting “Extract files” from the menu)

c) To open the file, open a new blank spreadsheet in Excel. Go to **Data** - **Get external data** - From Text (see image below)



d) Open 0001780-190415153152247.csv in Excel (using a **Tab** as the delimiter – if this does not work, please try other delimiter options until the data is formatted correctly)

e) Explore the dataset.

f) Answer the following questions:

What species is the dataset about?

Which institutions and countries does the data come from?

What type of data are these – are they based on observations of live animals or of museum specimens (hint: check the basisOfRecord column)?

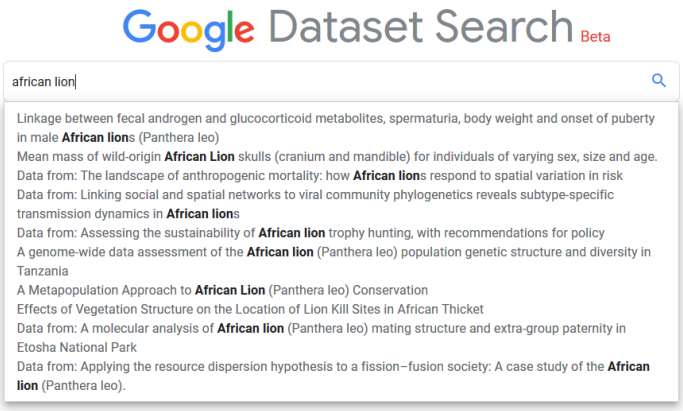
Are there any media associated with the occurrence records?

Do all the records have geographic coordinates so that they can be plotted on a map?

Exercise 3. Google Dataset Search

a) In your web browser, copy and paste https://toolbox.google.com/datasetsearch

b) In the search box, type in **African Lion**, as shown below. You will notice that a list of dataset appears. Click on the **search button**, to show all datasets (*but note that you choose a single dataset from the drop-down list if you wish*).



c) Notice that there are more than 100 datasets available for African Lion. Take note of the fact that most datasets have a unique identifier (a DOI), authors and a licence. We'll talk more about licences in the next session.

d) Now go back to the dataset search landing page and enter a species that you are interested in. Are there any datasets available for this species?

Exercise 4. GBIF Metadata

In this exercise, we will look at the metadata provided with a published dataset on GBIF.

a) Open the following page in your web browser https://www.gbif.org/
b) Click on **DATASETS**
c) In the search box, type in **EWT**
d) Select the first dataset: **EWT: African Crane Conservation Programme Sightings**
e) At the bottom of the first paragraph, click on **MORE**
f) Scroll down through, and familiarise yourself with, the different metadata sections.

Exercise 5. Data papers and data paper template

Two data papers have been provided for you to look at (Marnewick_etal_2017_.. and Smith_etal_2016_...). Both of these describe datasets that have been published to GBIF.

a) Locate both datasets on GBIF, and compare the metadata published on GBIF with the data paper content. You will notice that they match closely, with the data paper providing some additional content.
b) View the DataPaperTemplate_v1.docx file. This is a metadata template that is based on GBIF metadata and also on data papers published through Pensoft. It can be used to draft a data paper.

Exercise 6. Metadata spreadsheet

It is good practice, when storing your datasets long term (or when storing your students or datasets from other researchers), to store it with metadata. Here are two examples of metadata sheets prepared by EWT and SANBI for their datasets.

a) View the metadata templates (EWT_Metadata_Spreadsheet.xlsx and SANBI_Resources metadata_template.xlsx). These are institutional templates that are broadly based upon GBIF metadata, and can be used to capture metadata about your own datasets.