Introduction:

In this project I will examine how data from a dedicated sampling differs from that found in the GBIF database. The subjects of this study are mosses (Bryophyta). The two other Bryophytes divisions, the  liverworts (Marchantiophyta) and the hornworts (Anthocerotophyta) will not be investigated further. The study area is located in the Caldern Forest near the city of Marburg an der Lahn (Germany). The data from a previously made sampling in the area will be used to examine how different the rates of occurrence and richness of mosses are between the dedicated sampling and the data provided by GBIF.

Methods:

1. Step: Create a polygon of the study area in the Caldern Forest.

2. Step: Use the polygon of the study area to obtain the moss species occurrences from the GBIF database for the same area.

3. Step: Clean the obtained data with the “CoordinateCleaner” to get rid of false entries.

4. Step: Prepare the sampling data to import to the GBIF database -> formatting

5. Step: Add the sampling data of Bryophyte occurrences (Caldern Forest) to the GBIF database.

6. Step: Examine the differences before adding the data and after

7. Step: Statistical evaluation

Discussion:

Regarding the results I will discuss the problems of survey gaps leading to missing information of commonly occurring species. One problem is that mosses never really stood in the focus of environmental or botanical research in the past. Because of their relatively small size compared to other plants it is much harder to locate them and distinguish the species. Therefore GBIF shows significant occurrence gaps for species which are more or less known to be omnipresent in whole central Europe. This could be traced back to their missing importance to botanical research in the past. Luckily the extraordinary properties of mosses pushed them lately more into the focus of environmental research and their importance to ecosystems is now widely acknowledged.

Conclusion:

Nevertheless the survey gaps are obvious enough to question the accuracy of many available datasets from GBIF. Specially for poorly researched species one has to consider the probability that most statements derived from that data are not representative at all.