



What is it?

You analyse your data to gather information from a single or an integrated dataset. To analyse your data and use GFBio analysis tools, you have to upload your data. Then, they can be statistically explored through **visualisation, overlay, transformation, statistical analysis** or **modelling**. GFBio will also support the handling of large datasets, resulting from data integration. However, the type of software and hardware used for data analysis, as well as the type of data analysis itself, varies according to the discipline. This set of actions should provide new insight into research questions, leading to conclusions that can help decision maker to act upon.

Data analysis types:	Types of IT-infrastructures:
Descriptive, Explorative statistics	Personal computers
Inductive statistics, Parameter estimations	Grid computing
Multivariate statistics	Cloud computing
Model runs, Simulations	
Plotting and other visual representations	

How to do it?

1. **Reproducibility** is a crucial element for verification of results or for data reuse. Therefore, it is important to document the workflows of analysis and visualisation (e.g. computer scripts or text file notes).
2. Be as detailed as possible when writing the process metadata! Which results resulted from which step. Soon, a description tool, supplied by GFBio, can help you with that.
3. Find **appropriate software** that is ideally **open-source** (e.g. [R](#)). Ask yourself what you want to find out and which tools you need. Which software has proven its worth in similar studies and fits best to my expertise and time budget?
4. If **pre-processing** is required, state why. Normalize or transform raw data in meaningful values (algorithms) by using transformation- and analysis tools (soon available at GFBio).
5. Describe how the algorithm is applied.
6. Discover new patterns or outliers (in large datasets) via **data mining** and through plots (statistical or graphical output possible; soon implemented in GFBio).
7. Decide which **statistical model** fits the data best and perform your analysis.

Who does it?

Every kind of researcher dealing with biodiversity related data (data producers and data reusers).

Key elements

- Data upload with a submission tool, soon implemented at the [GFBio-Portal](#).
- Documentation of analysis workflows (process metadata).
- Data analysis with appropriate visualization and analysis tools (soon provided by GFBio) (data mining, descriptive statistics, graphical maps).
- Think about reproducibility!

Useful links

<https://kepler-project.org> (Scientific workflow interface)

http://www.vistrails.org/index.php/Main_Page (Scientific workflow interface)

<http://cran.r-project.org/> (R open-source software for statistics)

<http://www.rstudio.com/products/rstudio/download/> (R Studio open-source tool for R)