Introduction to Text Processing and Data Analysis with Python

**Objective**

The processing and analysis of large datasets has become a regular task in the Natural- and Life sciences. This introductory course into the scripting language Python lays the foundation for designing rapid, reproducible and scalable solutions to this problem.

**Description**

High throughput analyses, producing thousands to million data points, are meanwhile common in many areas of the Natural- and Life Sciences. A manual processing and analysis of such data is tedious, and even simple tasks, such as changing the format of output files, can severely impair a timely and meaningful interpretation of the data. Unfortunately, the diversity of the tasks that come along when handling such data makes the use of pre-existing out-of-the-box solutions typically impractical, if they exist at all.

The scripting language Python is an intuitive and powerful tool for developing custom-tailored solutions for problems ranging from basic data handling and management up to the design of complex workflows and novel algorithms for data analysis. In this course we will introduce you into the basic concepts of Python, making you familiar with the various data types and the general structure of Python scripts, but also with the basic concepts of a structured and standardized data analysis. Based on specific examples from the Natural- and Life Sciences we will guide you through the implementation of first algorithms in Python aiding in the solution of your particular data analysis problems.

**Methodology**

We will start with a general introduction into algorithm design followed by a presentation of the various data types and functions available in Python. We will then proceed to writing simple scripts for basic data handling and information extraction from text files.

Towards the end of the course we will begin to add modularity to Python scripts by introducing sub-routines – reusable parts of the code with a dedicated function – which later will allow structuring larger scripts.

All exercise in the course will be based on real-world example data. Participants are encouraged to bring their own data analysis problems to work on.

The workshop will take place on Campus Riedberg.

**Organizational Information**

<table>
<thead>
<tr>
<th>Language</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target group</td>
<td>Doctoral Candidates at all stages and Postdocs from Natural- and Life Sciences</td>
</tr>
<tr>
<td>Date</td>
<td>Monday-Tuesday, 6-7 January 2020, 9:00 – 17:00</td>
</tr>
<tr>
<td>Registration</td>
<td>For registration click here</td>
</tr>
</tbody>
</table>