

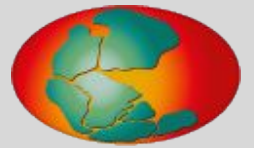
# PANGAEA

# Community Workshop 2022

## Introduction to PangaeaPy

*Kathrin Riemann-Campe and Michael Oellermann*

# Before we start...



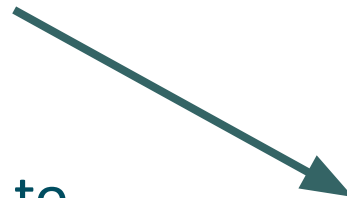
Launch the Python cloud environment!

*GitHub:*

<https://github.com/pangaea-data-publisher/community-workshop-material>

→ click on 2nd link  
“LaunchBinder”

...it may take a short while to  
load...



README.md



**PANGAEA.**  
Community Workshops

## Welcome to our PANGAEA Community Workshop Github page

This repository hosts training material, information and code used in our PANGAEA Community Workshop series.

Try out our practice examples in R or Python online by clicking on the "launch binder" batch



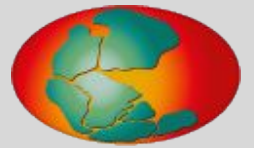
### ONLY DURING THE WORKSHOP

Click on this link to launch a script that prevents the 10 min session timeout when being inactive.



Start the script when you are inactive for more than 10 min and stop it once you resume your practice. Keep time intervals at a maximum to prevent being banned from using the MyBinder services.

# Before we start...



## Run KeepBinderActive.ipynb script to prevent timeout

KeepBinderActive.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
Python	29 minutes ago
R	29 minutes ago
banner.png	29 minutes ago
hello.py	29 minutes ago
hello.R	29 minutes ago
install.R	29 minutes ago
KeepBinder...	7 minutes ago
README.md	29 minutes ago
requiremen...	29 minutes ago
runtime.txt	29 minutes ago

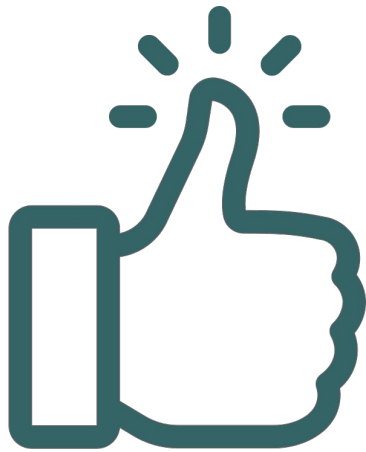
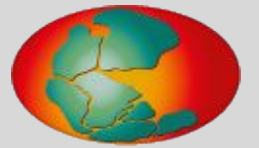
```
[*]: import time

# Function to prevent session timeout by executing code every x minutes
# Make sure to turn this off and keep time intervall as long as possible to avoid being banned from MyBinder
def KeepBinderActive(time_lapse = 9, time_out = 60):
    time_start = time.time()
    while True:
        time.sleep(time_lapse*60)
        time_now = time.time()
        time_diff = round((time_now - time_start)/60, 0)
        print(f'Keeping MyBinder active since {time_diff} minutes')
        if time_diff >= time_out:
            print("This session has timed out")
            break
    return

# Call the timer function
KeepBinderActive()

Keeping MyBinder active since 9.0 minutes
Keeping MyBinder active since 18.0 minutes
```

# Why PangaeaPy?



Credit: Gregor Cesnar

Direct data queries and retrieval via Python

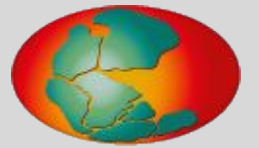
Easy bulk data download

Metadata search and analysis

Enables automated workflows and data mining

Clear record of data retrieval workflow and data sources

# Getting started



## At work/homeoffice

Install PangaeaPy in your environment

“pip install pangaeapy”

Import packages

```
import pangaeapy as pan
```

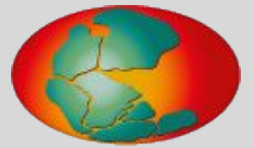
or

```
from pangaeapy.panquery import PanQuery
```

```
from pangaeapy.pandataset import PanDataSet
```

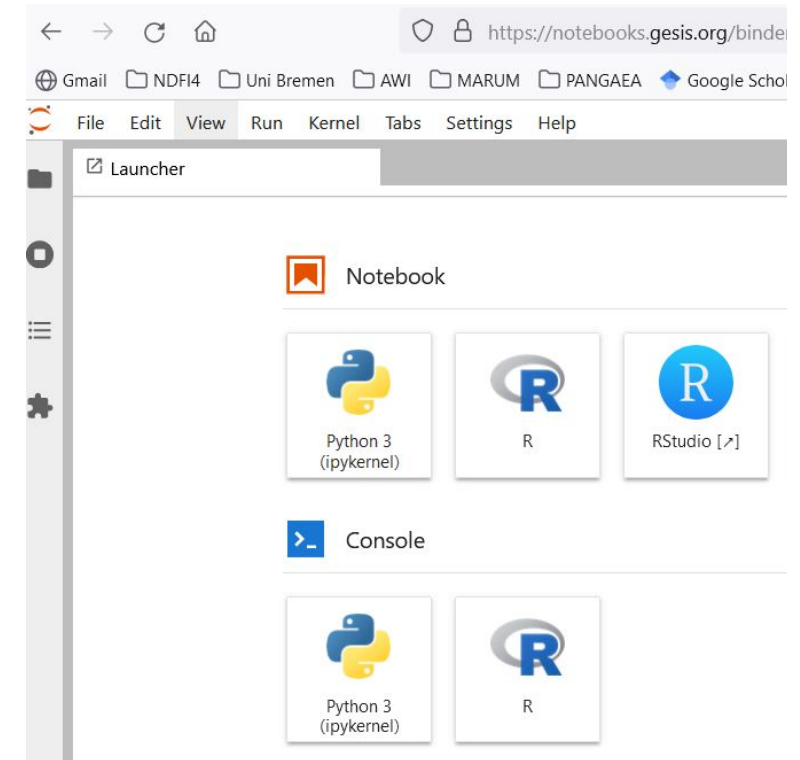


# Getting started

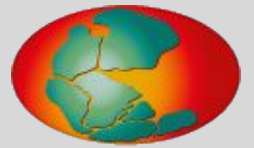


## For our practice today

- Work in cloud environment (MyBinder)
- Linked to PANGAEA's [GitHub account](#)
- JupyterLab, R or RStudio
  - cannot save files but download



# Working in Jupyter notebooks



Run cells with Shift + Enter or clicking play button

The screenshot shows a Jupyter Notebook interface. On the left is a file browser with a search bar and a list of files and folders. On the right is a notebook cell titled 'Basic 1 - load a PANGAEA dataset'. The cell contains text and code for loading a PANGAEA dataset. A play button icon is circled in the notebook toolbar.

File browser contents:

Name	Last Modified
Python	36 minutes ago
R	36 minutes ago
banner.png	36 minutes ago
hello.py	36 minutes ago
hello.R	36 minutes ago
install.R	36 minutes ago
README.md	36 minutes ago
requirements.txt	36 minutes ago
runtime.txt	36 minutes ago

Notebook cell content:

Basic pangaeapy example: how to load a PANGAEA dataset

First we need to import the PanDataSet class from the pangaeapy module:

```
[1]: from pangaeapy import PanDataSet
```

Then we create an PANGAEA dataset object. To initialize the object we need to use either a integer ID or an DOI of the dataset we want to use within the constructor.

Code and Markdown

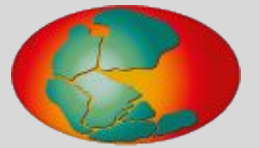
A simple test would be to print the title of the database:

```
[6]: ds.title
```

```
[6]: 'Dipicolinic acid concentration in sediment samples'
```

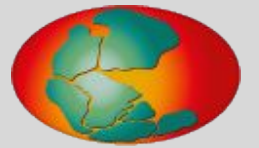
Linked GitHub File repository

Simple 0 \$ 1 Python 3 (ipykernel) | Idle Mem: 206.42 / 8192.00 MB Mode: Command Ln 1, Col 1 Basic 1 - load a PANGAEA dataset.ipynb



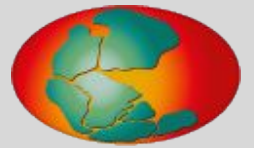
Function	Attributes
<code>pan.PanQuery('search term')</code>	totalcount, result
<code>pan.PanAuthor('author name')</code>	lastname, firstname, ORCID
<code>pan.PanEvent('label')</code>	label, lat, long, elevation, date, device
<code>pan.PanParam</code>	id, name, param_type, unit, PI





Function	Attributes
<code>pan.PanDataSet('PANGAEA ID')</code>	data, id, uri, doi, title, abstract, year, authors, citation, parameters, events, min-/maxtimeextent, children, etc.

# Let`s practice



KeepBinderActive running

Navigate to Python/Examples

Start “getting\_started\_pangaeapy.ipynb”

To switch to RStudio later on:

Turn off simple view →

Launcher Tab → RStudio

The screenshot shows the JupyterLab interface. The left sidebar contains a file browser with a search bar and a list of files and folders. The main area displays the code editor for 'KeepBinderActive.ipynb'. The code in the editor is as follows:

```
[ ]: import time

# Function to prevent session timeout by executing
# Make sure to keep time intervall as long as poss
# if not required anymore to avoid being banned fr
def KeepBinderActive(time_lapse = 9, time_out = 60
    time_start = time.time()
    while True:
        time.sleep(time_lapse*60)
        time_now = time.time()
        time_diff = round((time_now - time_start)/
        print(f'Keeping MyBinder active since {tim
        if time_diff >= time_out:
            print("This session has timed out")
            break
    return
# Call the timer function
KeepBinderActive()
```

At the bottom of the interface, there is a status bar with a 'Simple' toggle switch, memory usage (0 / 8192.00 MB), and other system information.