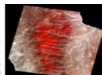
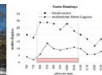


# **GEOCOMPUTATION AND MACHINE LEARNING FOR ENVIRONMENTAL APPLICATIONS**

**Date: 06 April 2021, 2.00 PM - 03 June 2021, 4.30 PM**

**Dr. Giuseppe Amatulli  
&  
Dr. Longzhu Shen**



# Learning objectives

**With continuous practice through the weeks, students will become familiar with new command lines and cover numerous topics, including:**

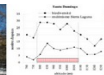
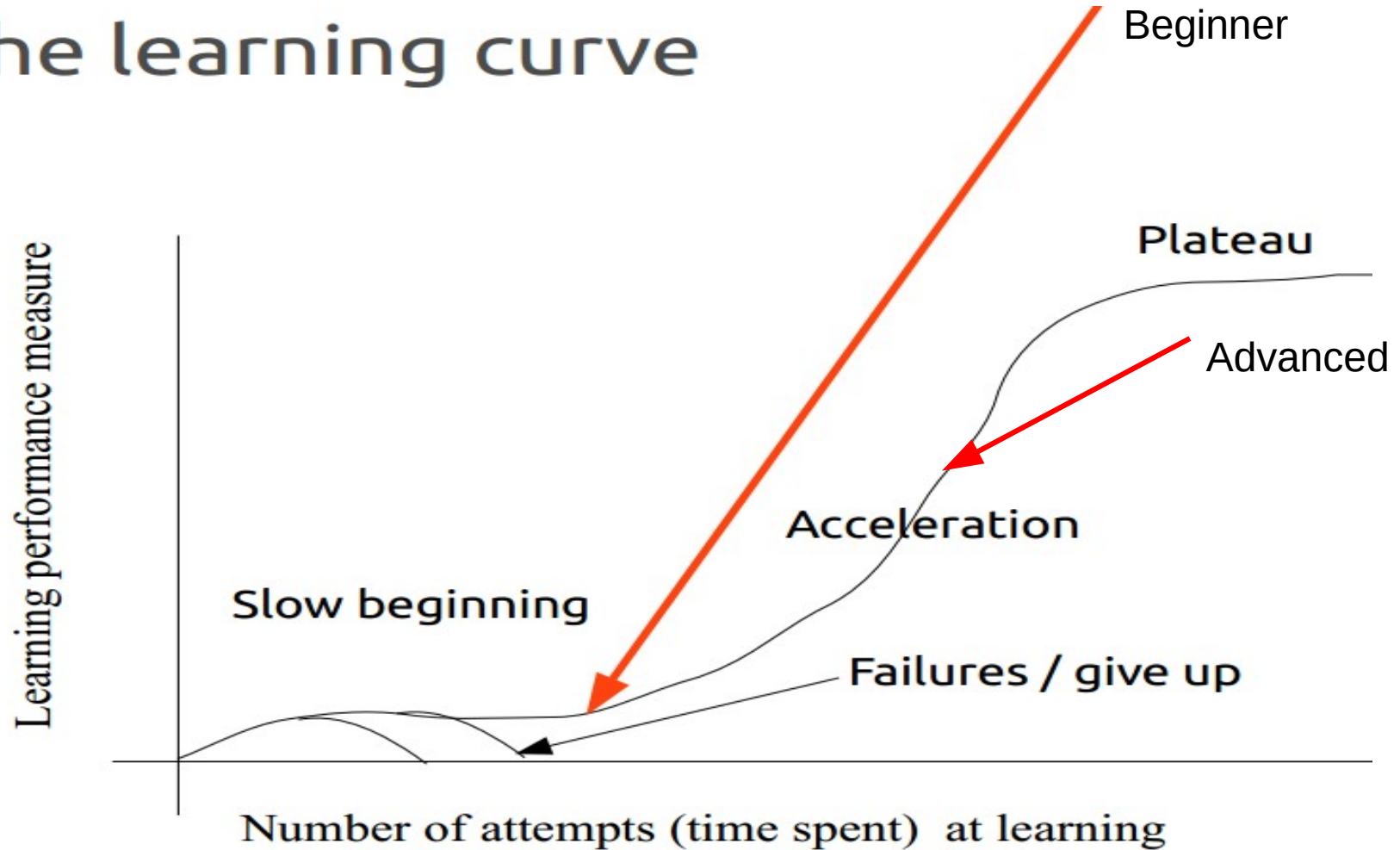
- **Learning open source tools for GIS and RS applications.**
- **Acquiring command line utilities for spatial/temporal data under Linux OS.**
- **Acquiring command line utilities and ML theoretical foundation for environmental application.**
- **Developing data processing skills.**
- **Independent learning, critical thinking, problem solving.**

**Upon completion of the course, you should be able to:**

- **Apply the process of science, by conducting, analyzing, and interpreting findings related to GIS & RS project in the framework of ML applications**
- **Use quantitative reasoning for statistical/spatial analysis**
- **Convey your understanding of environmental phenomenons**



# The learning curve



# Scientific knowledge

- **Spatio/temporal analysis**
- **Spatio/temporal data integration**
- **Spatio/temporal modeling**
- **Geostatistic**
- **Machine Learning**



# Tools

**Grass & Qgis** Geographic Information Systems

**Python:** GIS, statistic, modeling, text manipulation

**LINUX Bash** shell programming

**AWK** for processing text-based data

**GDAL/OGR/PKTOOLS** geotools library for the manipulation of geospatial data

**CDO** geotools library for the manipulation of netCDF

<http://spatial-ecology.net/docs/build/html/index.html>

<http://spatial-ecology.net/dokuwiki/doku.php>



# Knowing each other (3 min)

- **Name, where are you coming from....**
- **What is your background and personal interest?**
- **Final project / PhD thesis objectives / keywords?**
  - **What data are you going to analyses?**
  - **Not sure yet... no problem**
- **Do you have an experience on Linux OS or other open source software?**
- **Do you currently use any programming language?**
- **What are your interests and expectations on this training?**



# Coding knowledge

- **Beginners / Inter mediated / Advanced**

[https://docs.google.com/forms/d/1i4FqM4xlwMBVVZzTViuDDqwL-BsaPs2m8NObB2\\_oTK0/edit#responses](https://docs.google.com/forms/d/1i4FqM4xlwMBVVZzTViuDDqwL-BsaPs2m8NObB2_oTK0/edit#responses)

## Covered in the course

- **AWK, GDAL, PKTOOLS, CDO, Python, R, ML (keras, TensorFlow, PyTorch)**
- **Parallel processing in bash and python environment**
- **Machine Learning: theoretical foundation and application in the environmental field**
- **Supervised and unsupervised classification application**
- **Image processing / raster processing / large data-set processing in a proper way**
- **Hydrological modeling**



## What is possible to acquire with the course

- I wish to expand my knowledge on geo-computation in open-source environment
- I would like to see a bit of everything, to get a glimpse of what is available
- Able to run extensive spatial data stuff quickly and without "pain"

## What is NOT possible to acquire with the course

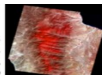
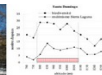
- Implement a machine learning optimizer for an inverse modeling system to quantify CO2 emissions.





# Syllabus clarification

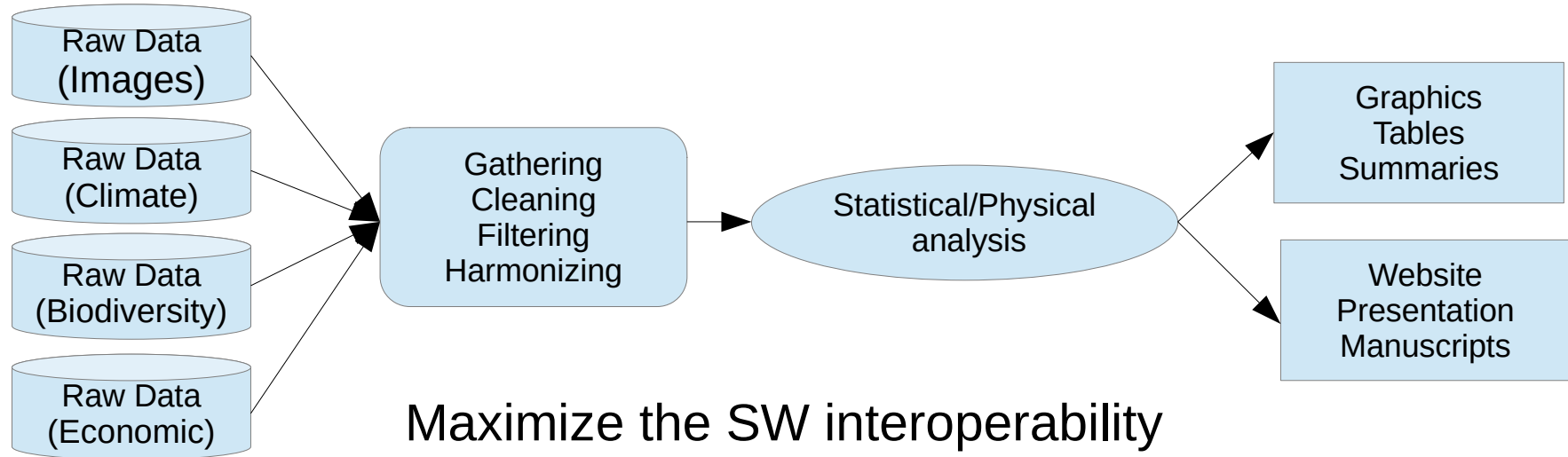
- **Material** <http://spatial-ecology.net/docs/build/html/index.html>
- **Data via github** [https://github.com/selvaje/SE\\_data](https://github.com/selvaje/SE_data)
- **Online recording video lecture later shared on Spatial Ecology web**
- **Handling script and data via github (code)**
- **Community support among us for trouble shouting via slack**



# Reproducible research & “big data” processing

**Codes that are easily published** > no license constraints

**Complex work-flows** > integrate different data analysis methods



Maximize the SW interoperability  
in a stable Operating System



# Why use Linux/OpenSource?

**Security:** extremely stable and reliable, no viruses,  
interoperable: Unix, Windows, Mac, Android, ...

**Applications:** thousands of free programs,  
programming languages, server services

**Versatility:** minimum HW requirements,  
extremely portable, very fast performance

**Freedom:** free to download/test/install/modify/  
configure/develop/distribute/... it's fun!



# Freedom? and why it's fun?

## **Code:**

- Understating the code beyond a process
- Be able to modify the code
- Build up your own algorithm.
- Use all the SW that I want without license constraints

## **Help:**

- Get help from mailing list
- Keep in touch with the developers for code adjustment and improvement

## **Process:**

- Job priority processing
- Job scheduling
- RAM management

## **Remote server:**

- Automatic connection to remote servers
- Overpassing quota issues in remote servers, by creating a folder linked to your PC

## **Hardware resources:**

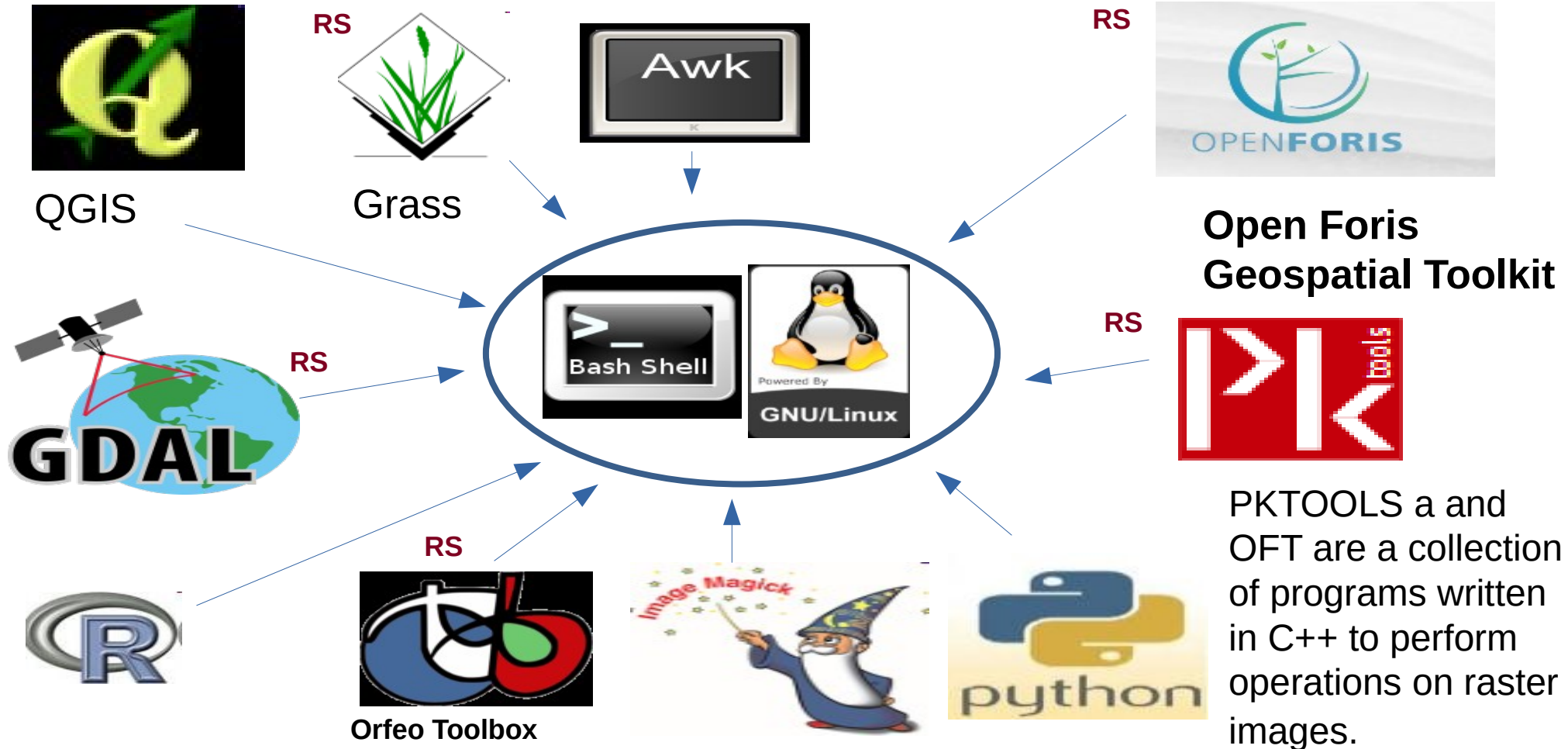
- Storing temporal file in ram rather in the hard-disk, by creating a folder in the ram
- Get the best of different programming languages and create a unique work flow.

## **Last but not least:**

- Enjoy the life in the meantime the PC is working for you!

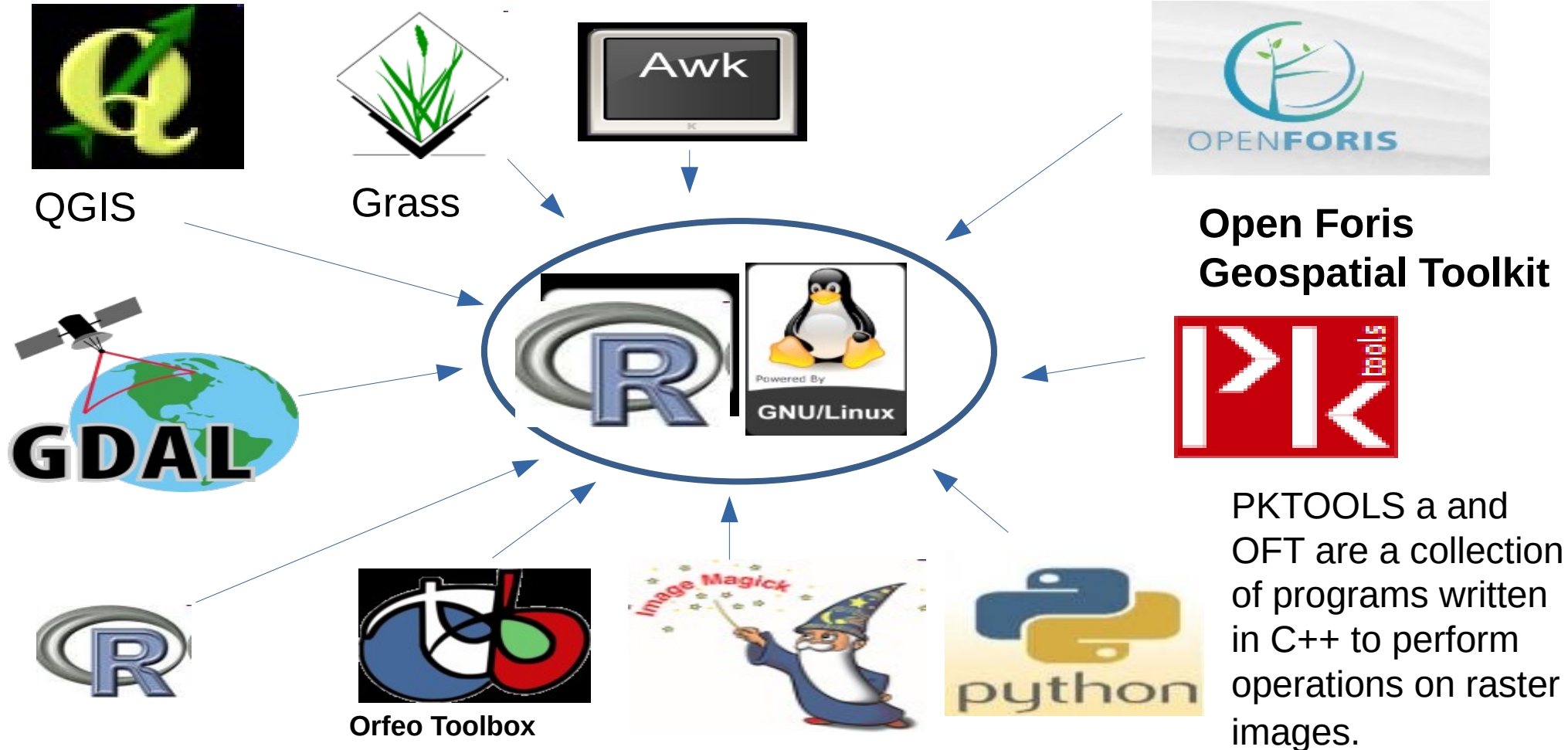
# Ubuntu Linux operating system

## Programming languages interaction



# Ubuntu Linux operating system

## Programming languages interaction



# Ubuntu Linux operating system

## Programming languages interaction

