



UNIVERSITY OF TWENTE.

CENTER OF EXPERTISE IN BIG GEODATA SCIENCE
GEOSPATIAL COMPUTING PLATFORM

dr. ing. Serkan Girgin MSc
s.girgin@utwente.nl

Center of Expertise in Big Geodata Science (CRIB) is a horizontal facility established in **March 2020** to **enable** the better use of **geospatial cloud computing and big data technologies** in education, research, and institutional strengthening activities at ITC.

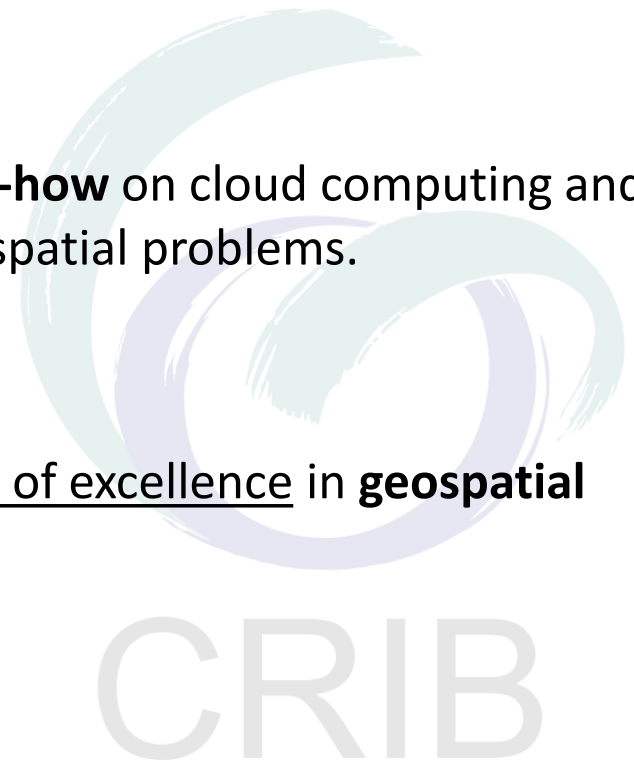
Mission

Collect, develop, and share **operational know-how** on cloud computing and big data technologies to solve large-scale geospatial problems.

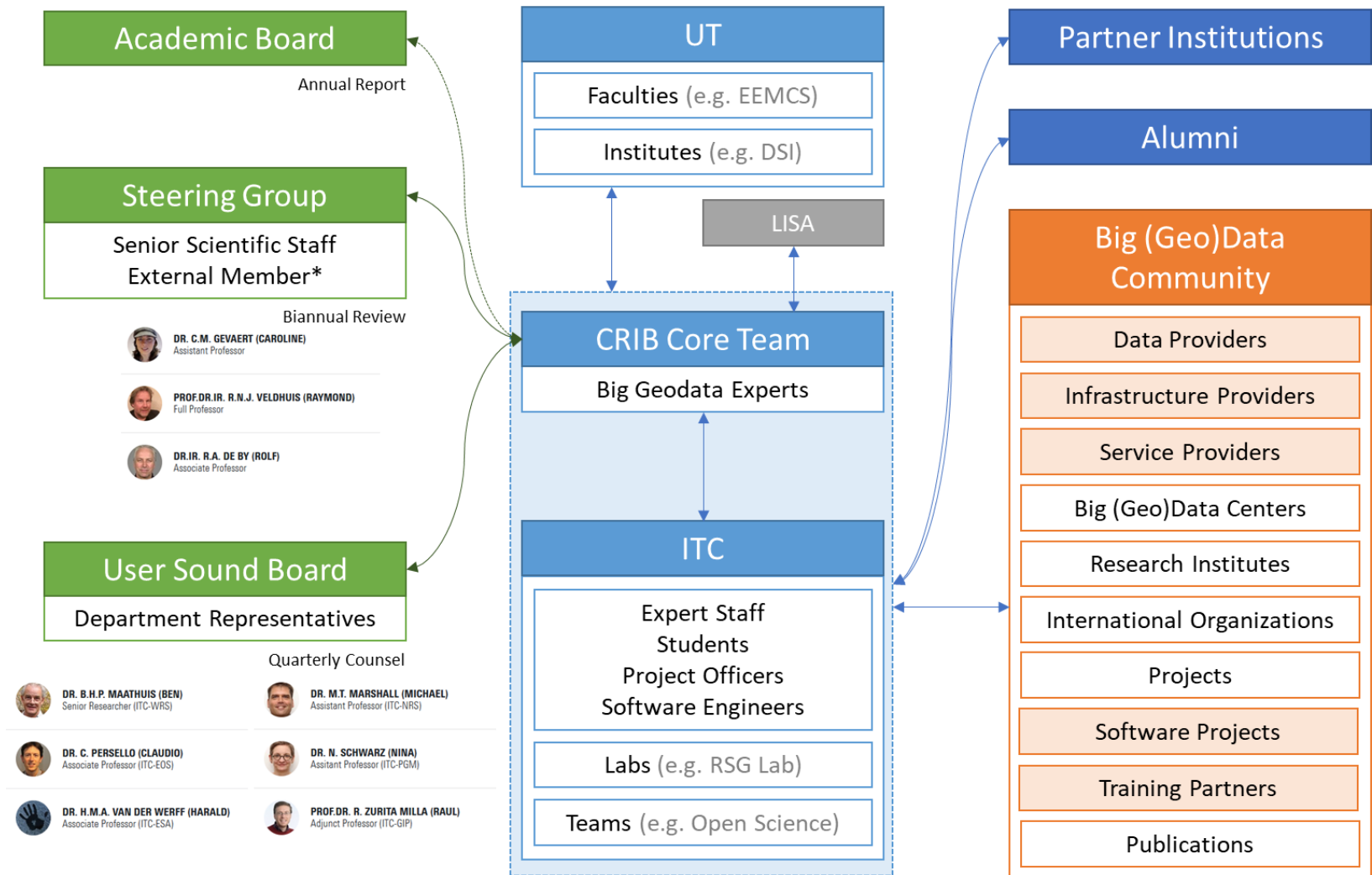
Vision

Position UT/ITC as a globally renowned center of excellence in **geospatial cloud computing and big data** science.

<https://itc.nl/big-geodata>



Structure



Activities

- Capacity and Knowledge Development
 - Improving expert knowledge, providing theoretical and hands-on training, facilitating community of practice
- Infrastructure Development
 - Local and cloud-based infrastructure for big geodata computing
- Project Services
 - Consultancy and advisory services for integration and better use of big data technology, active support to research projects
- Monitoring and Networking
 - Monitoring recent developments in big (geo)data technology, networking with data providers, developers, research institutions
- Visibility
 - Ensuring high visibility of big-data related activities

CENTER OF EXPERTISE IN BIG GEODATA SCIENCE

Home ITC / Research / Research facilities / Labs & resources / Center of Expertise in Big Geodata Science

CENTER OF EXPERTISE IN BIG GEODATA SCIENCE (CRIB) IS A HORIZONTAL FACILITY THAT SUPPORTS ALL ITC DEPARTMENTS FOR THE BETTER USE OF BIG GEODATA TECHNOLOGY IN EDUCATION, RESEARCH, AND INSTITUTIONAL STRENGTHENING ACTIVITIES.

Our mission is to collect, develop, and share operational know-how to solve large-scale geospatial problems involving big data.

Our vision is to position University of Twente and ITC as a globally renowned center of excellence in big geodata science.

Center of Expertise in Big Geodata Science
43 Tweets

Center of Expertise in Big Geodata Science
@BigGeodata

Center of Expertise in Big Geodata Science (CRIB) supports the use of geospatial big data technology in education, research, and capacity development activities

📍 Enschede, the Netherlands 🔗 itc.nl/big-geodata 📅 Joined June 2020

148 Following 172 Followers

Spatial Data Cubes

Large amounts of Earth observation data are needed for geoscientific research. However, the combination of these data presents a challenge for data storage and processing. In this project, we are developing a data cube designed to store and process large amounts of Earth observation data.

Apache Spark meets GPU

Spark 3.0

eScience Center - ITC collaboration on large-scale phenological analysis

In 2016, ITC and the Netherlands eScience Center set an alliance for high-resolution phenological monitoring at continental scale, which resulted in the development of a cloud-based analysis platform. Recently, the collaboration has been extended to explore the performance of different computing platforms, namely Apache Spark and Dask, for large-scale phenology studies. The study, which can also provide useful information (e.g. performance metrics) for other big geodata problems, will be supported by the Center of Expertise in Big Geodata Science (CRIB).

CRIB web portal is online

The Big Geodata Newsletter not only aims to inform the community about the activities of CRIB in general, but also on the activities of CRIB in general. The primary aim is to provide information on the portal you can find information about CRIB news and events. All issues of the newsletter are available in the web portal. Our Twitter account is connected with the community and share more news.



Big Geodata Newsletter

194 subscribers* (50% outside UT)



* As of 27 January 2022

ITC Geospatial Computing Platform

- Designed to serve primary activities identified by the needs assessment:
 - **Self learning**
 - **Exploratory research**
 - **Education**
- Design criteria
 - **Highly available** 24/7, no queue
 - **Ready to use** Pre-installed software
 - **User friendly** Interactive user interface
 - **GPU enabled** GPU for each user
 - **Distributed-computing friendly** Computing cluster
 - **Low cost** Feasible investment

NVIDIA Jetson AGX Xavier

- **8-core CPU**
(NVIDIA Carmel **ARMv8.2**, 2.26GHz, **NVIDIA L4T**)
- **512-core GPU**
(**Volta Architecture** with 64 Tensor Cores)
- **32GB memory**
(256-bit LPDDR4x, 2133MHz, 137GB/s, **Unified**)
- **32GB storage**
(eMMC 5.1)
- **Dual Deep Learning Accelerator***
- **Vision Accelerator***
- **4x 4Kp60 video encoder**
(H.264/H.265)
- **2x 8Kp30 / 6x 4Kp60 video decoder**
(H.265)
- **Gigabit Ethernet**
(RJ45)
- **500 GB / 1 TB M.2 NVMe SSD**
(Samsung EVO 970 Plus, 3GB/s)



① <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>

① https://elinux.org/Jetson_AGX_Xavier

Computing Resources

- **Service Units (2 x Dell PowerEdge T320)**
 - 6-core CPU (Intel Xeon E5-2420 v2, 12 threads, 2.70 GHz)
 - **40 TB** local storage (4 x 10 TB 3.5" 7.2K SAS HDD, RAID 2+1)
 - **192 GB** memory
- **Computing Units (16 x NVIDIA Jetson AGX Xavier*) (128 cores, 512 GB)**
- **Big Data Computing Units (Dell PowerEdge R730)**
 - 2 x 18-core CPU (Intel Xeon E5-2695, 72 threads, 2.10 GHz, max. 3.3 GHz)
 - **NVIDIA RTX A4000** GPU (Pascal architecture, 3840 CUDA cores, 12 GB GDDR5X memory)
 - **768 GB** memory
- **Big Data Computing Unit with Local Storage (Dell PowerEdge R730xd)**
 - 2 x 8-core CPU (Intel Xeon E5-2640, 32 threads, 2.60 GHz, max. 3.4 GHz)
 - **24 TB** local storage (20 x 1.2 TB 2.5" 10K SAS 12 GB/s HDD, RAID 20+2)
 - **NVIDIA RTX A4000** GPU (Pascal architecture, 3840 CUDA cores, 12 GB GDDR5X memory)
 - **768 GB** memory
- **Rapid Data Computing Unit (Dell Optiplex 9020)**
 - 4-core CPU (Intel Core i7-4770 CPU, 8 threads, 3.40 GHz)
 - **32 GB** memory
- **120 TB** external storage (0.2 PB total)

Platform as a Service

<https://crib.utwente.nl>

- Based on **open-source** software ([Ubuntu](#), [Docker](#), [JupyterHub](#), [JupyterLab](#), ...)
- Accessible through a **web browser** ([No software installation or VPN are required](#))
- **No registration** is required ([Login with UT credentials](#))
- Each user has an individual and isolated **working environment**
- Each user has access to all available* **unit resources**, including **GPU**
- Each user has access to all available* **cluster resources**
- **Replicated storage** with minimum two copies ([Hardware failure protection](#))
- **Distributed storage** for big data processing ([HDFS](#))
- Automatically scales and **balances workload** among the units
- Low energy footprint ([10-30W per unit](#))

* Resource availability depends on resource usage of other active user



Key Features

<https://crib.utwente.nl>

- **Interactive notebook, terminal and remote desktop** access are available
- Multiple interactive languages are supported (Python, R, Julia, Octave, Go, ...)
- **Up-to-date and optimized software packages** are **ready to use** (No setup required)
- Users can install additional packages (e.g., Python, R packages)
- Different **architectures** and **OS-specific applications** are supported (e.g., Windows)
- Distributed computing clusters are **ready to use** (Dask, Apache Spark)
- **Public** assets are shared by all users (e.g., OSM Planet Data)
- **Shared workspaces** allow assets to be shared by selected users
- Access can be granted to **external users**
- **User support** is available
- Provided and maintained by **CRIB** at no extra cost (i.e., free PaaS)



Current Usage

- Operational since **January 2021**
- **704*** registered users
- **72*** shared workspaces for projects and courses
- **5-30*** concurrent users at a time
- Provided approximately **62,500+*** hours of multi-core/GPU computation
- Overall, quite **positive feedback** from a wide-range of use cases
- Several **project proposals** consider to utilize the platform
- Other **UT units** (e.g., DCC, BDSI) are interested in having similar platforms
- LISA build a similar platform for **UT-wide use**
Co-developed by CRIB, available at <https://jupyter.utwente.nl> (VPN)
- 4TUResearchData, TU Delft, FAO are interested to **have similar platforms**

* As of 15 February 2022

Interactive Access



The screenshot displays a JupyterLab environment. On the left is a file browser showing a directory structure with files like '1-Hello.ipynb', '2-Data-Ana...', '3-Interactiv...', '4-Geospati...', '5-Interactiv...', '6-Geospati...' (selected), '7-R.ipynb', and '8-dask.ipynb'. The main area is split into two panes. The top pane shows a code editor with the following Python code:

```
[1]: import rasterio
from rasterio.plot import show

dir = '/data/public/GEODATA/Various-Netherlands/Aerial-Photogr...'
img = rasterio.open(dir + '253000_470000.tif')
show(img)
```

Below the code is a plot of an aerial photograph showing a residential area with buildings and roads. The axes are labeled with coordinates: x-axis from 253000 to 254000 and y-axis from 470000 to 470500.

The bottom pane shows a histogram plot titled 'Histogram'. The x-axis is labeled 'Frequency' and ranges from 0 to 250. The y-axis is labeled 'Frequency' and ranges from 0.0 to 3.5. The plot shows a distribution of data with a peak around 150. The legend indicates the data is from a generator object.

On the right side of the interface, there is a 'Living Textbook' panel. It has a search bar and a list of items. The selected item is 'Aerial survey', which has a 'Data collection' tag. Below the title is an 'Introduction' section with the following text:

Aerial photographs are a major source of digital data; soft-copy workstations are used to digitize features directly from stereo pairs of digital photographs. These systems allow data to be captured in two or three dimensions, with elevations measured directly from a stereo pair using the principles of photogrammetry. Analogue aerial photos are often scanned before being entered into a soft-copy system, but with the advance of high-quality digital cameras this step can now be skipped.

In general, the alignment of roads and railways, lakes and water, and shapes of buildings are easily interpreted on aerial photographs - assuming that the scale of the photographs is not too small. Also, constructions such as dikes, bridges, air fields and the main types of vegetation and cultivation are mostly clearly visible. Nevertheless, numerous attribute data related to terrain features cannot be interpreted on aerial photographs: e.g. the administrative qualification of roads, sea and lake depths, functions of buildings, street names, and administrative boundaries. We will have to collect this information in the field or from existing data sets and maps (e.g. road maps, navigational charts or town plans).

① JupyterLab documentation: <https://jupyterlab.readthedocs.io/en/stable/>

① CRIB JupyterLab Training: <https://itc.nl/big-geodata/training/jupyterlab/>

Available Software

<https://crib.utwente.nl>

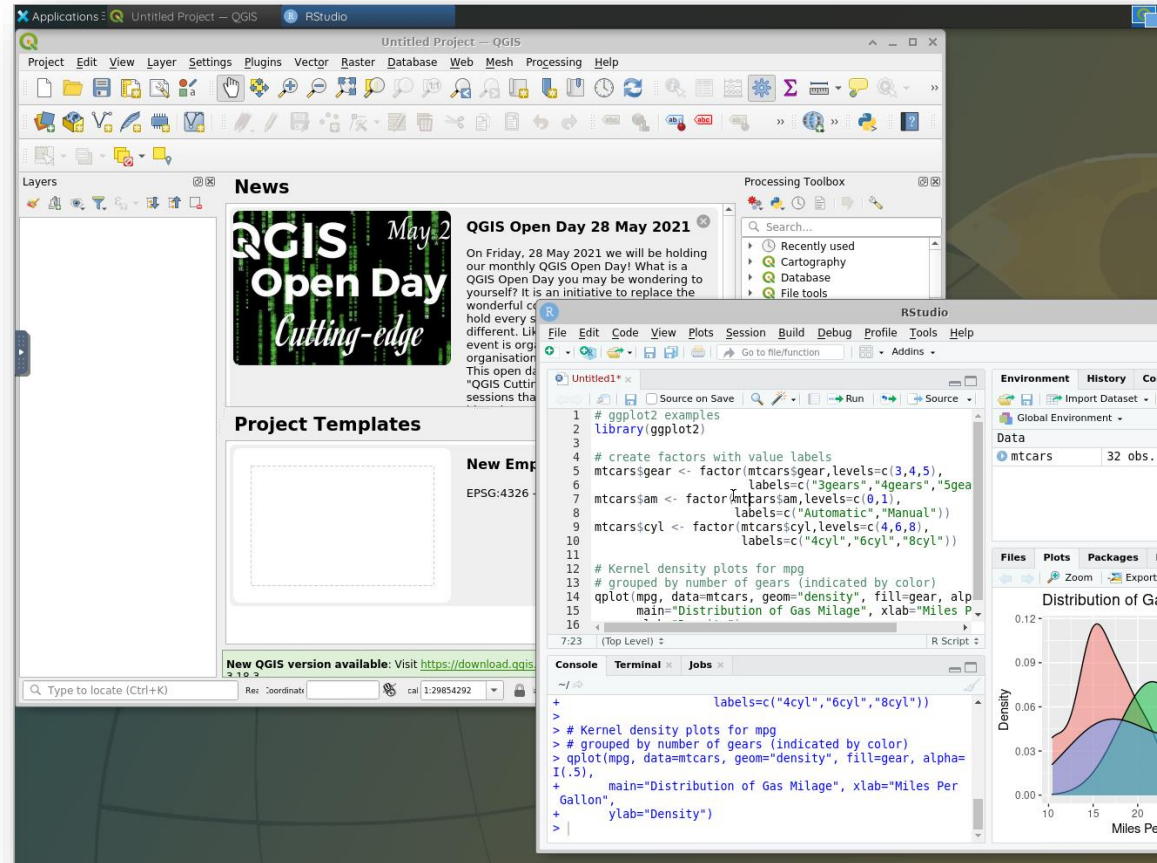


... and many more!

① Complete lists of more than **800+ Python** and **400+ R** packages are available at `/public/platform`

Available Desktop Applications

- QGIS
- GRASS GIS
- SAGA GIS
- SNAP
- ILWIS 3*
- ILWIS 4*
- VS Code
- PyCharm
- R Studio
- Netlogo
- GNU Octave
- **MATLAB***
- Glueviz
- Orange Data Mining
- Firefox



- **XFCE Desktop Environment** is available through:
 - Launcher > Remote Desktop

① XFCE Desktop Environment: <https://www.xfce.org/>

Terminal Access

```
.n Name Size Modify time
./ .astrophy 4 Apr 25 11:10
./ .cache 25 May 11 17:16
./ .config 29 May 18 14:46
./ .cupy 3 Feb 7 20:49
./ .dbus 3 Jan 31 06:08
./ .glue 4 Apr 25 11:10
./ .gnupg 8 May 28 11:18
./ .grass7 5 Apr 5 11:01
./ .ipython_checkpoints 11 May 18 14:09
./ .ipython 6 Feb 16 15:27
./ .java 4 Mar 14 01:00
./ .julia 7 Apr 1 16:51
./ .jupyter 6 May 13 09:00
./ .jupyter-php 3 Feb 19 03:57
./ .keras 4 Apr 6 07:48
./ .local 3 Jan 15 15:27
./ .netlogo 3 Mar 14 01:00
./ .npm 3 May 13 14:55
./ .rstudio-desktop 22 May 13 14:26
./ .snap 18 May 2 14:04
./ .ssh 3 Mar 16 14:19
./ .subversion 6 Apr 5 11:06
./ .virtual_documents 2 May 20 13:14
./ .vnc 223 May 20 11:10
./ .wine32 10 May 19 11:07
./ .wine64 9 May 19 11:07
./ Desktop 2 May 14 23:12
./ artifacts 5 Apr 1 13:54
./ compiled 3 Jan 17 17:21
./ fragstats 2 May 14 23:23
./ globetrotter 6 Jan 17 17:42
./ grass 3 Mar 13 13:29
./ ilwisdata 9 May 11 17:22
./ mlruns 4 Feb 9 15:09
./ packages 31 Apr 1 13:54
./ -private 13 Jan 15 15:27
./ -public 12 Jan 15 15:27
./ pysal_data 2 Jan 24 21:58
./ qemu 6 May 14 23:04
./ -shared 12 Jan 15 15:27
./ tools 19 May 14 23:11
./ ICEauthority 39732 May 20 11:18
./ .Rhistory 3689 May 13 14:26
./ .Xauthority 10280 May 20 11:18
UP--DIR 16T/17T (97%)

Hint: Want your plain shell? Press C-o, and get back to MC with C-o again.
joyvan@2a524fb09da4:~$
```

- bash shell is available through:
 - JupyterLab: Launcher > Terminal or File > New > Terminal
 - Remote Desktop: Applications > Terminal Emulator (use this for UI scripts)

① bash Tutorial for Beginners: <https://linuxconfig.org/bash-scripting-tutorial-for-beginners>

Available Services

<https://crib.utwente.nl>



GeoServer

Open source server for sharing
geospatial data



MapServer

Open source platform for
publishing spatial data



PostgreSQL

Open source relational database



MariaDB

Open source relational database



GeoNode

Open source geospatial content
management system



Dataverse

Open source research data
repository software



Gitea

Open source lightweight code
hosting solution



Open Data Kit

Open source platform to collect
data quickly, accurately, offline, and
at scale

Support Center

<https://crib.utwente.nl/support/>

UNIVERSITY
OF TWENTE.



CRIB

[Support Center Home](#)

[Knowledgebase](#)

[Open a New Ticket](#)

[Check Ticket Status](#)

Search our knowledge base

Search

[Open a New Ticket](#)

[Check Ticket Status](#)

Welcome to the CRIB Support Center!

In order to streamline support requests and better serve you, we utilize a support ticket system. Every support request is assigned a unique ticket number which you can use to track the progress and responses online. For your reference we provide complete archives and history of all your support requests.

Quick Access

- [Report a Problem](#)
- [Shared Workspace Request](#)
- [Course Workspace Registration with Canvas Integration](#)
- [External Account Request](#)
- [Account Removal Request](#)
- [Account Transfer Request](#)
- [Software Request](#)
- [Dataset Request](#)
- [Database Request](#)

Featured Questions

[How can I access to the platform?](#)

[Is it secure?](#)

[How can I use the platform?](#)

[Which programming languages are supported on the platform?](#)

[Which libraries and packages are supported by the platform?](#)



Important Directories

- Home directory (full-access, also used by the system)
/home/jovyan or **~**
- Private directory (full-access)
~/private or **/data/private**
- Public directory (read-only, maintained by CRIB)
~/public or **/data/public**
- Shared directories (read-only or full-access)
~/shared/<directory> or **/data/shared/<directory>**

Network
storage

- Local directory (full-access, PowerEdge R730xd only)
~/shared/<directory> or **/data/shared/<directory>**
- Temporary directory (full-access, not permanent)
/tmp
Useful for intermediate files during computation

Local storage
(fast)

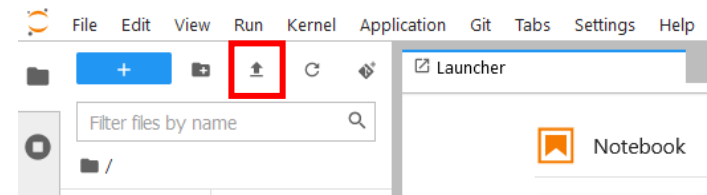
Additional Information

- Up-to-date information on the platform is available in `public/platform`
 - `benchmark/` : Performance benchmarks
 - `config/` : Configurations of custom-built software packages
 - `demo/` : Example interactive notebooks for demonstration purposes
 - `test/` : Functional tests of selected packages and libraries
 - `languages` : List of available programming languages
 - `apt-packages` : List of installed system packages
 - `python-packages` : List of installed Python packages
 - `r-packages` : List of installed R packages
 - `custom-built` : List of custom-built software packages
 - `whats-new` : List of changes and platform updates
 - `work-in-progress` : Features under development
 - `faq` : Frequently asked questions

FAQ

- How can I upload files to the platform?

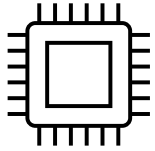
- Use the **Upload** button
- You can upload multiple files
- Files are uploaded to the active folder
- You cannot upload directories
- To **upload directories** with multiple files or sub-directories:
 - Create an archive file of the directory (e.g. zip, tar.gz)
 - Upload archive file
 - Extract archive file
 - On the terminal, `unzip <archive.zip>` or `tar xzvf <archive.tar.gz>`
 - On the remote desktop, Applications > Accessories > Archive Manager
- We will provide better options soon



FAQ

- How can I install R, Python, Julia, etc. packages?
 - For **Python**, open a terminal and enter the command:
`pip install <package name> or pip install <package name>==<version>`
 - For **R**, enter the command:
`install.packages('<package name>', repos='https://cloud.r-project.org')`
 - For other languages, please refer to the user documentation
 - Packages are installed to your **home directory** (they are permanent)
 - They are not updated automatically (you should keep them up to date)
 - You may encounter installation errors if the package requires additional system libraries or it is not compatible with the selected architecture
 - Please [contact us](#) if you have difficulties, **we will install it for you**, which will make it also available to other users
 - **Warning:** Local package dependencies are not guaranteed for platform updates
 - **Warning:** Local packages are architecture dependent
 - [Conda](#) is not supported, but you can use [virtual environments](#), if necessary

Contact



<https://crib.utwente.nl>
<https://crib.utwente.nl/support/>



<https://itc.nl/big-geodata>



crib-itc@utwente.nl
s.girgin@utwente.nl



[@BigGeodata](https://twitter.com/BigGeodata)



[Big Geodata Newsletter](#)

