FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION (ITC)

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

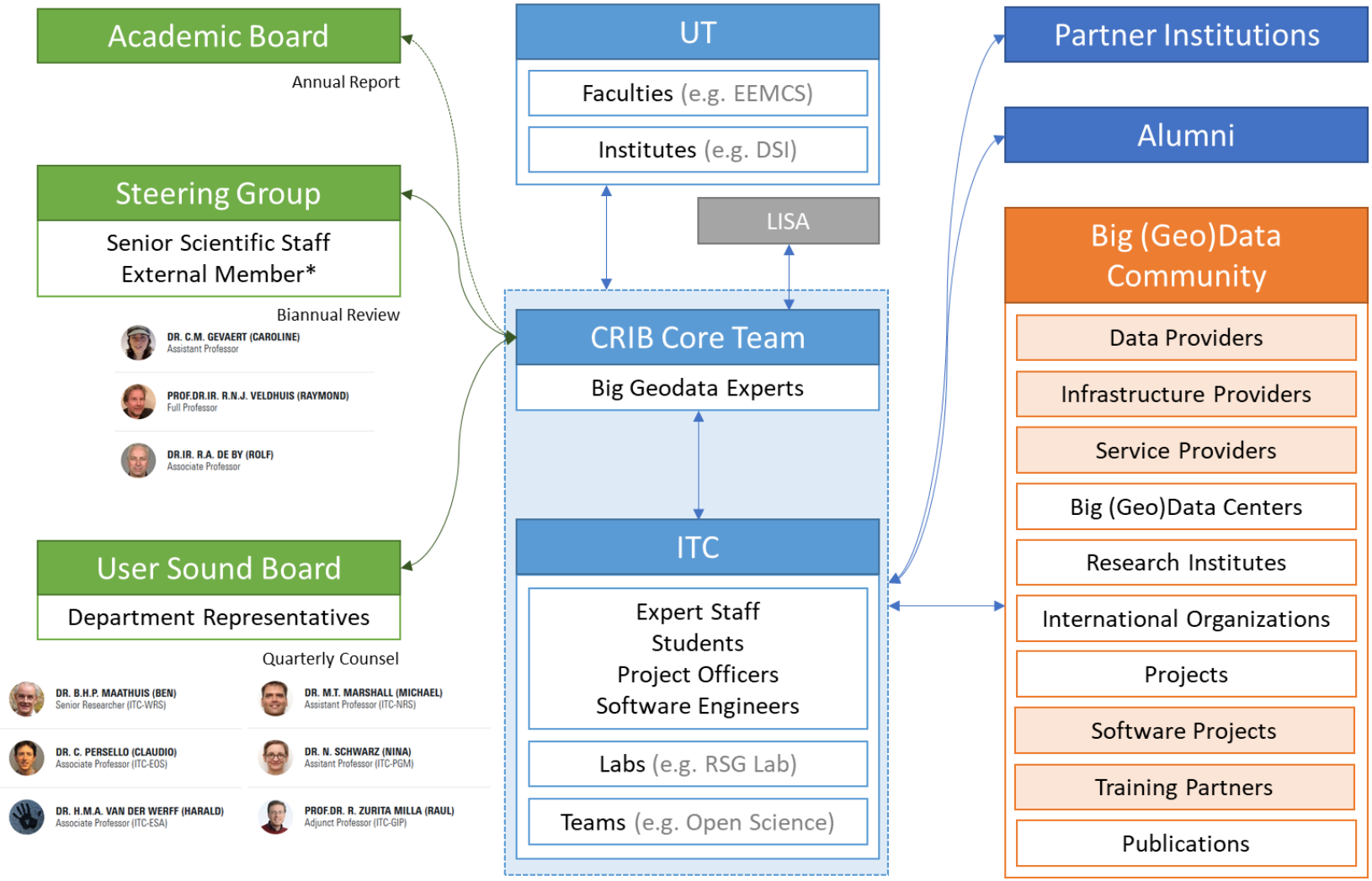
Collect, develop, and share **operational know-how** on cloud computing and big data technologies to solve large-scale geospatial problems.

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


Academic Board
Annual Report
Steering Group
Senior Scientific Staff
External Member*
Biannual Review
DR. C.M. GEVAERT (CAROLINE)
Assistant Professor
PROF.DR.IR. R.N.J. VELDHUIS (RAYMOND)
Full Professor
DR.IR. R.A. DE BY (ROLF)
Associate Professor
User Sound Board
Department Representatives
Quarterly Counsel
DR. B.H.P. MAATHUIS (BEN)
Senior Researcher (ITC-WRS)
DR. M.T. MARSHALL (MICHAEL)
Assistant Professor (ITC-NRS)
DR. C. PERSELLO (CLAUDIO)
Associate Professor (ITC-EOS)
DR. N. SCHWARZ (NINA)
Assitant Professor (ITC-PGM)
DR. H.M.A. VAN DER WERFF (HARALD)
Associate Professor (ITC-ESA)
PROF.DR. R. ZURITA MILLA (RAUL)
Adjunct Professor (ITC-GIP)
UT
Faculties (e.g. EEMCS)
Institutes (e.g. DSI)
LISA
CRIB Core Team
Big Geodata Experts
ITC
Expert Staff
Students
Project Officers
Software Engineers
Labs (e.g. RSG Lab)
Teams (e.g. Open Science)
Partner Institutions
Alumni
Big (Geo)Data Community
Data Providers
Infrastructure Providers
Service Providers
Big (Geo)Data Centers
Research Institutes
International Organizations
Projects
Software Projects
Training Partners
Publications

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

Contact

https://itc.nl/big-geodata

@BigGeodata

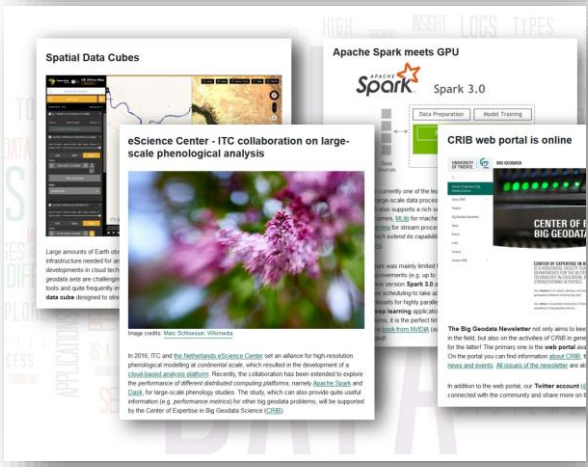
320 followers*

Big Geodata Newsletter

194 subscribers* (50% outside UT)

Big Geodata Channel

* 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)

ⓘ https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit
ⓘ 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)

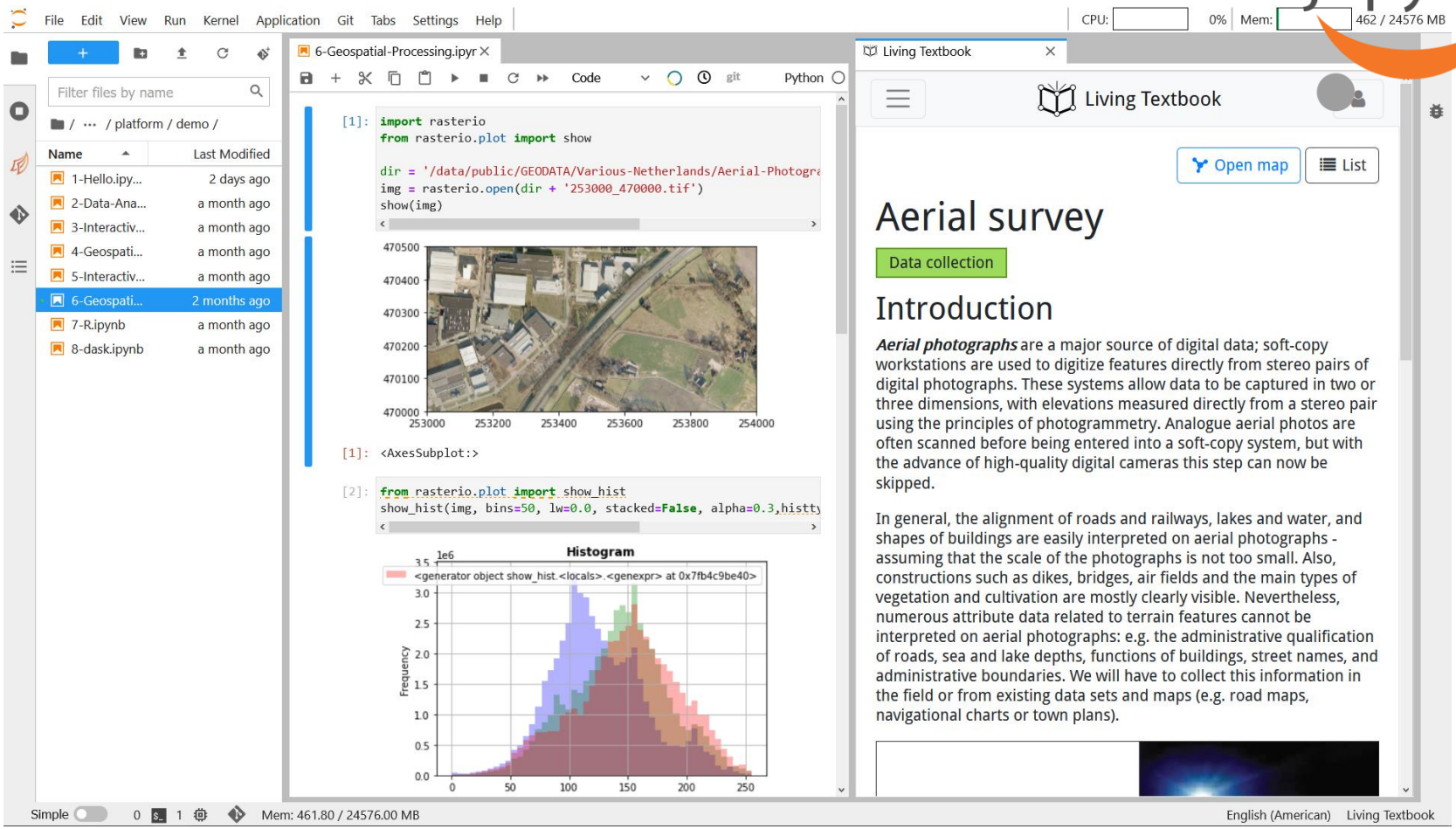



Current Usage

- 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



ⓘ JupyterLab documentation: https://jupyterlab.readthedocs.io/en/stable/
ⓘ **CRIB JupyterLab Training:** https://itc.nl/big-geodata/training/jupyterlab/

Available Software

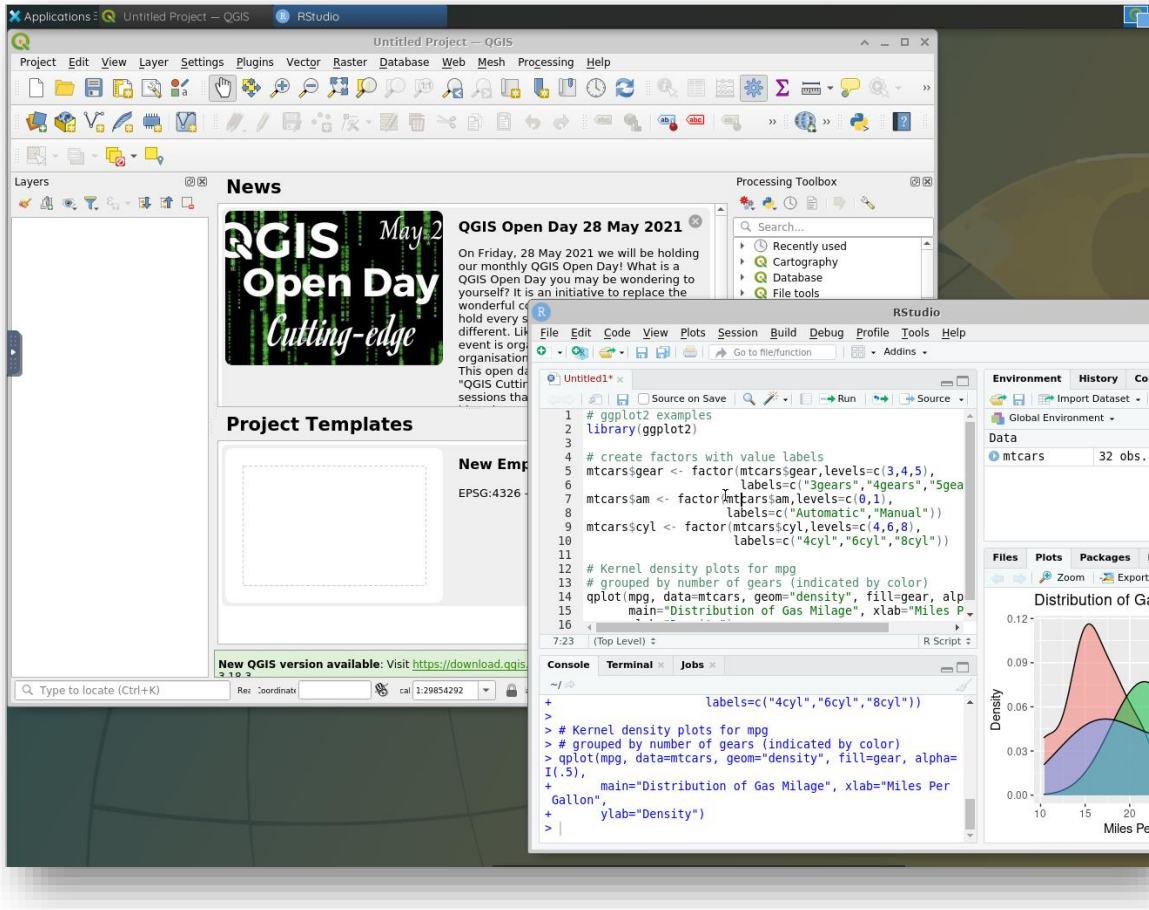
https://crib.utwente.nl

… and many more!

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

Available Desktop Applications

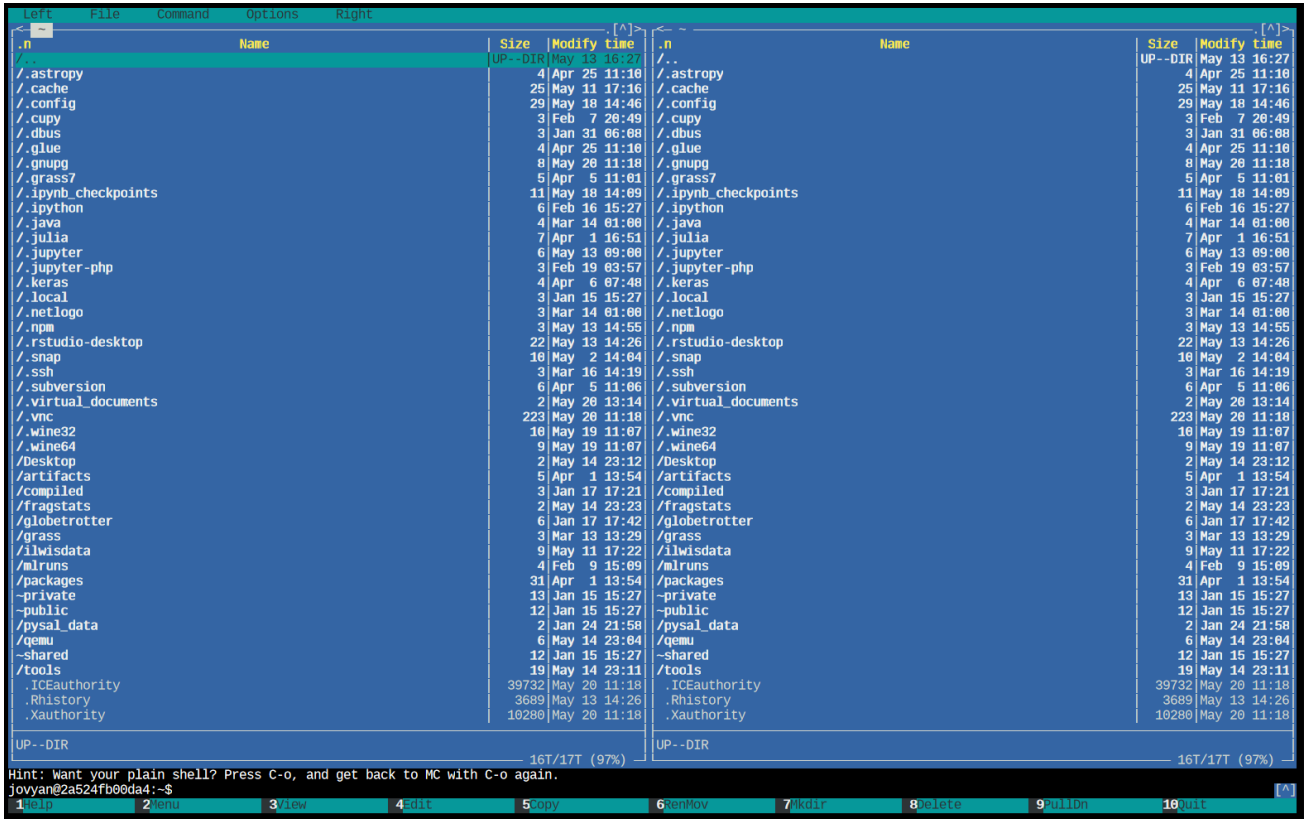
- QGIS
- GRASS GIS
- SAGA GIS
- SNAP
- **ILWIS 3***
- **ILWIS 4***
- VS Code
- PyCharm
- R Studio
- Netlogo
- GNU Octave
- **MATLAB***
- Glueviz
- Orange Data Mining
- Firefox



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

ⓘ XFCE Desktop Environment: https://www.xfce.org/

Terminal Access



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

ⓘ 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

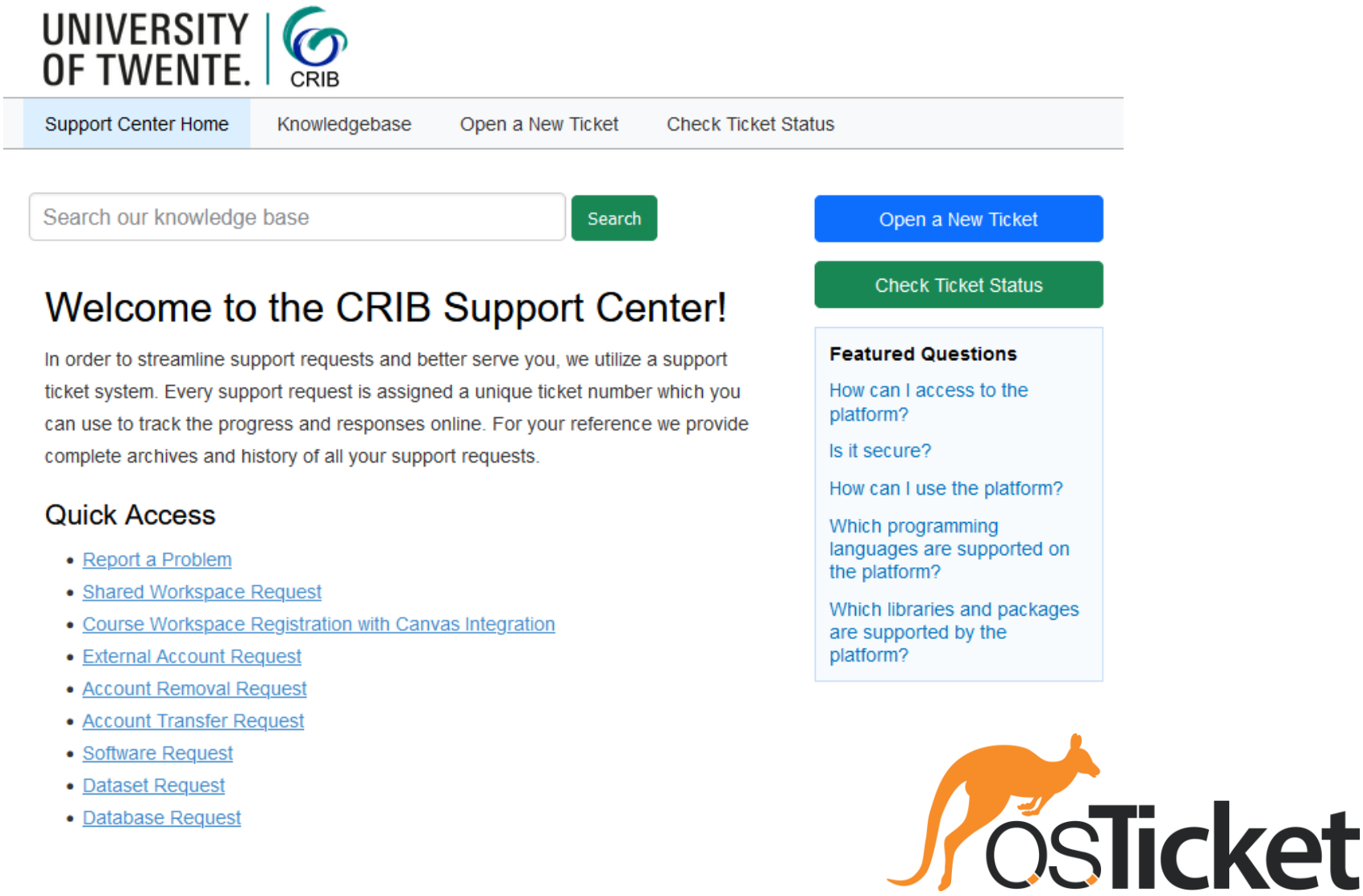
Open source lightweight code hosting solution

Open Data Kit

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

Support Center

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



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

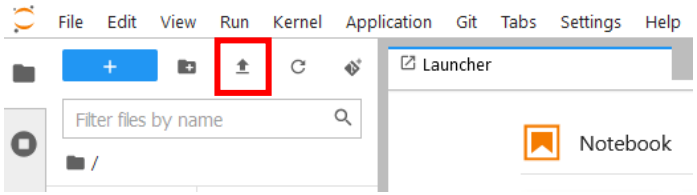
Local storage (fast)

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

Big Geodata Newsletter