

CyVerse Austria: Cloud-based scientific data management in research & education

Introduction to CyVerse

Research Data Management @ TU Graz

- DMP Support
- RDM Optimization Institute/Group: Workflows/Pipelines, Infrastructure
- RDM Individual support & support in Research Projects (Collaborations)
- RDM guidelines & policies
- Tool-dependent support
(workshops, introduction, training, assistance, development)

RDM-Tools

- TU Graz DMP Tool (<https://www.tugraz.at/sites/rdm/tools/tu-graz-dmp-tool>)
- GitLab (<https://www.tugraz.at/sites/rdm/tools/gitlab>)
- eLabFTW (<https://www.tugraz.at/sites/rdm/tools/elabftw>)
- TU Graz Repository (<https://www.tugraz.at/sites/rdm/tools/tu-graz-repository>)
- CyVerse (<https://www.tugraz.at/sites/rdm/tools/cat-cyverse-austria>)

What is CyVerse?

- A computational infrastructure for data-driven discovery
- The cyberinfrastructure includes:
 - Data Storage
 - Data Science Workbench
 - Education and Training

CyVerse US

- Created in 2008 by the National Science Foundation
- Led by University of Arizona
- Partners:
 - Texas Advanced Computing Center
 - Cold Spring harbor Laboratory

Data platforms for Analytics next to CyVerse?

- Google (Colab, Drive, Sheets a.o.)
- Microsoft (365, Onedrive, Power BI)
- Nextcloud (In-Apps)
- SAS (Viya),
- Cohesity,
- Mendeley Data Monitor, a.o.

The logo for Google Colab, featuring the word "Google" in its multi-colored font above the word "colab" in a bold, orange, lowercase sans-serif font.

Erich Hohenstein, 2022-04-14,
Intro to Goolge Colab on
<https://levelup.gitconnected.com/google-colab-what-is-it-how-to-use-it-and-why-should-i-care-721bb8a88c90>

The Nextcloud logo, consisting of three blue circles of varying sizes arranged in a triangular pattern, with the word "Nextcloud" in a blue sans-serif font below them.

<https://nextcloud.com>

The Cohesity logo, with the word "COHESITY" in a bold, uppercase sans-serif font. The letter "S" is highlighted in green.

<https://www.cohesity.com/>



Andrea Negri, 2017-01-13, SAS
Viya is perfect for IT Nerds at
<https://blogs.sas.com/content/brightdata/2017/01/13/sas-viya-perfect-for-it-needs/>

Brand-new: European Open Science Cloud - EU Node

<https://open-science-cloud.ec.europa.eu/>

Login with your university credentials and use some resources for free

Services

<p>File Sync & Share 10 credits consumed in this period View Service ></p>	<p>Interactive Notebooks Access enabled View Service ></p>	<p>Large File Transfer Access enabled View Service ></p>
<p>Virtual Machines Not available Please read our User Access Policy for more information View Service ></p> <p><small>Credits renewed 2025-02-16</small></p>	<p>Cloud Container Platform Not available Please read our User Access Policy for more information View Service ></p>	<p>Bulk Data Transfer Not available Please read our User Access Policy for more information View Service ></p>

CyVerse Austria

- Uses open source code from CyVerse US
- Started in 2016 within a project funded by the ministry
- Hardware:
 - Storage: Total: 100 TB; used: 3.2 TB
 - CONDOR resources (non-interactive): 10 CPUs, 20 GB RAM
 - VICE resources (interactive): 200 CPUs, 412 RAM

Moving Data

- **Discovery Environment**
 - CyVerse web interface
- **CyberDuck**
 - 3rd party software tool to move large and multiple files; runs on Windows and macOS
- **iCommands**
 - Collection of commandline tools by iRODS (this is the technology supporting the CyVerse Data Storage); runs only on linux
- **GoCommands**
 - Commandline tool; runs on any modern OS

Data infrastructures: benefits compared to local working

- No installation on local machines,
Data access from everywhere
- Usage via web browser or command line
- HPC & shared resources
- Container-based applications (reliability)
- Independent from own system



<https://www.docker.com>

Data infrastructures: disadvantages compared to local working

- Internet connection is needed
- Long procedure to start (VPN, login, start application...)
- Installation of new apps need expertise (e.g. Docker)
- Bug fixing runs over a ticket system



<https://www.docker.com>

CyVerse Help/Courses/Tutorials

- Online Course: <https://imoox.at/course/catmoox>
- Learning Center: <https://learning.cyverse.org/>
- CyVerse-Mooc: https://cyverse-learning-materials.github.io/cyverse_mooc/
- CyVerse Videos: <https://www.youtube.com/@CyverseOrgProject>

CyVerse Austria Registration

Go to registration page: user.cyverse.tugraz.at



Welcome to CyVerse!

New to CyVerse?

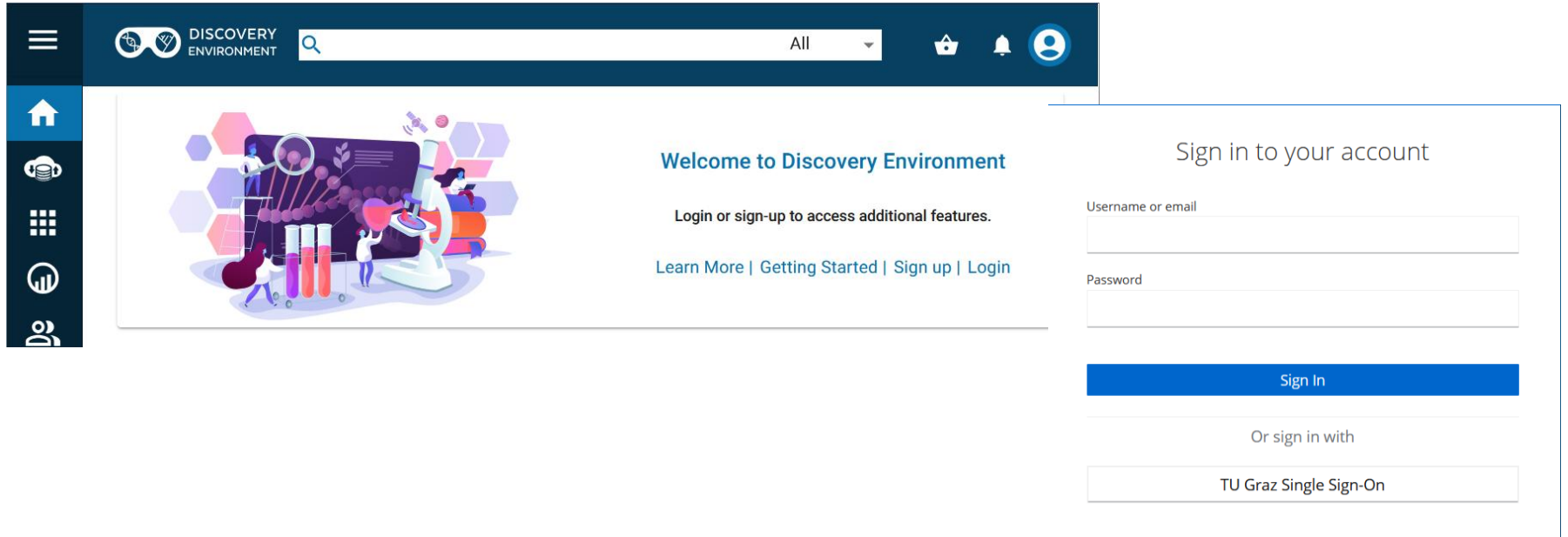
[SIGN UP](#)

[LOGIN](#)

[FORGOT PASSWORD](#)

CyVerse Austria Login

Go to registration page: de.cyverse.tugraz.at



The screenshot shows the Discovery Environment login page. At the top, there is a dark blue navigation bar with a hamburger menu icon, the 'DISCOVERY ENVIRONMENT' logo, a search bar with the text 'All', and icons for a shopping cart, notifications, and a user profile. Below the navigation bar is a vertical sidebar with icons for home, a microscope, a grid, a magnifying glass, and a person. The main content area features a large illustration of a laboratory scene with a microscope, test tubes, and a person. To the right of the illustration, the text reads: 'Welcome to Discovery Environment', 'Login or sign-up to access additional features.', and a list of links: 'Learn More | Getting Started | Sign up | Login'. On the far right, there is a white login form titled 'Sign in to your account' with fields for 'Username or email' and 'Password', a blue 'Sign In' button, and an option to 'Or sign in with TU Graz Single Sign-On'.

A white line-art sketch of a large, classical-style building with a central dome and multiple arches, set against a dark background.

CyVerse Austria: Demo

Speech-to-text Analyse mit Whisper



Was ist Whisper?

- Automatisches Spracherkennungssystem (ASR)
- Trainiert auf 680.000 Stunden mehrsprachiger und multitasking-überwachten Daten
- Ermöglicht die Transkription in mehrere Sprachen sowie die Übersetzung aus diesen Sprachen ins Englische
- Open Source Software
(<https://github.com/openai/whisper?tab=readme-ov-file>)

Wofür verwende ich Whisper?

- [Shared RDM Services & Infrastructure](#)
- Knowledge Transfer – [Webinarreihe „RDM in Österreich“](#)
 - Über 40 Episoden momentan
 - Abrufbar auf [YouTube](#) & [TU Graz Repository](#)
- Whisper wird zur Erstellung von Untertiteln für die Aufzeichnungen verwendet



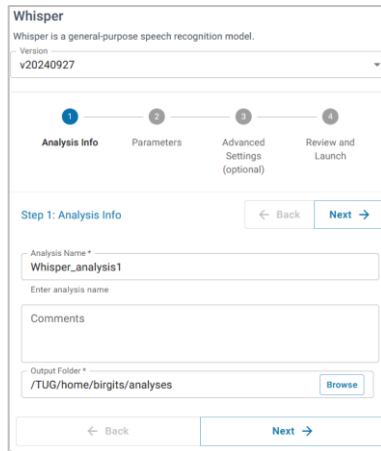
Warum verwendet ich Whisper über CyVerse?

- Läuft online, es werden keine lokalen Ressourcen benötigt
- Läuft immer, es kann beispielsweise auch über Nacht oder am Wochenende laufen
 - Kürzere Wartezeiten für den Abschluss der Analyse
- Es können mehrere Vorgänge eingereicht werden, die nacheinander bearbeitet werden
- Ist schneller im Vergleich zur lokalen Nutzung

Wie verwendet man Whisper auf CyVerse?

1

Analyseinformationen definieren



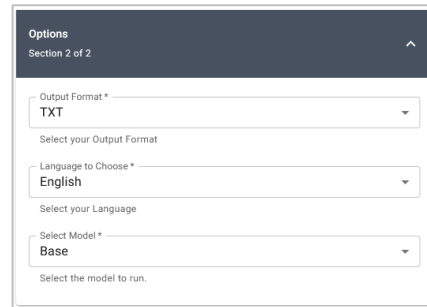
The screenshot shows the 'Whisper' interface with a progress bar indicating four steps: 1. Analysis Info (active), 2. Parameters, 3. Advanced Settings (optional), and 4. Review and Launch. Below the progress bar, the 'Step 1: Analysis Info' section contains the following fields:

- Version:** v20240927
- Analysis Name *:** Whisper_analysis1
- Comments:** (empty text area)
- Output Folder *:** /TUG/home/birgits/analyses (with a 'Browse' button)

Navigation buttons include 'Back' and 'Next' at the bottom of the form.

2

Sprachauswahl & Modellauswahl



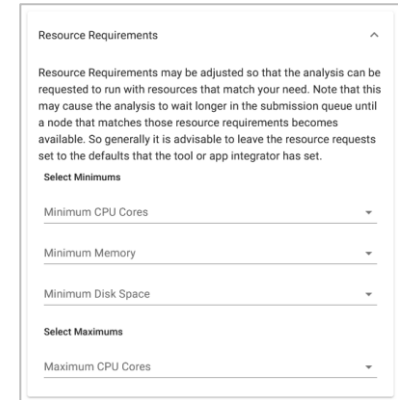
The screenshot shows the 'Options' section, labeled 'Section 2 of 2'. It contains the following configuration options:

- Output Format *:** TXT
- Language to Choose *:** English
- Select Model *:** Base

Each option is accompanied by a dropdown menu. A 'Next' button is visible at the bottom right of the section.

3

Definition der Ressourcen

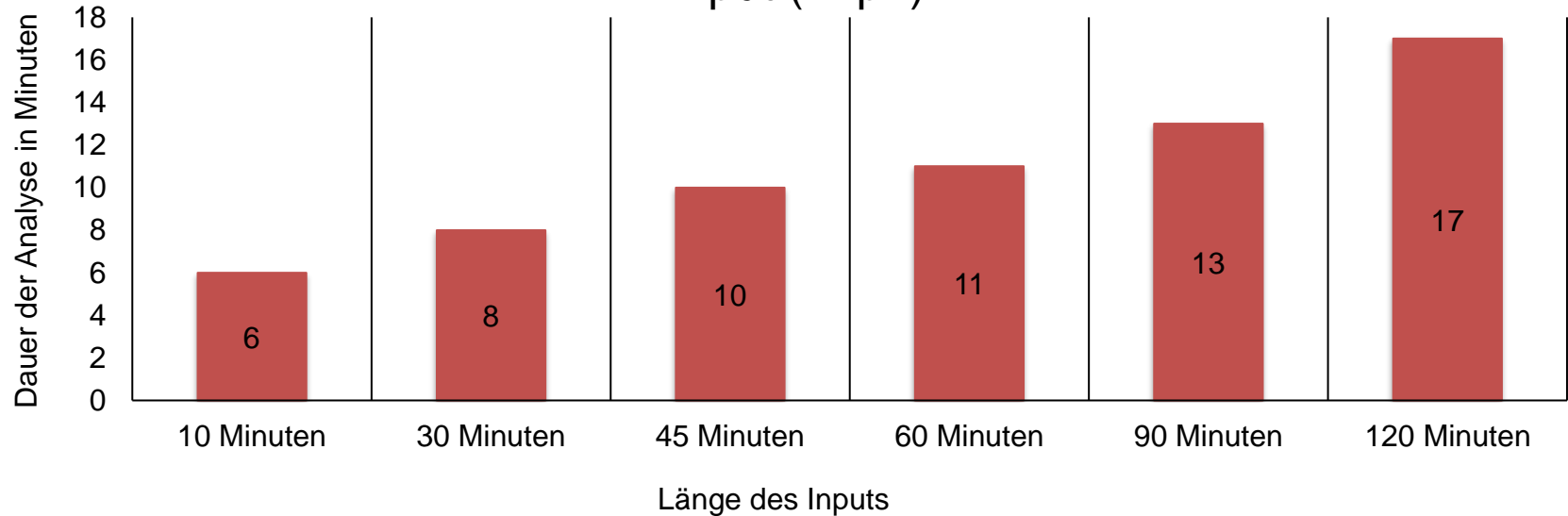


The screenshot shows the 'Resource Requirements' section. It includes a warning message and several adjustable settings:

- Resource Requirements:** Resource Requirements may be adjusted so that the analysis can be requested to run with resources that match your need. Note that this may cause the analysis to wait longer in the submission queue until a node that matches those resource requirements becomes available. So generally it is advisable to leave the resource requests set to the defaults that the tool or app integrator has set.
- Select Minimums:**
 - Minimum CPU Cores
 - Minimum Memory
 - Minimum Disk Space
- Select Maximums:**
 - Maximum CPU Cores

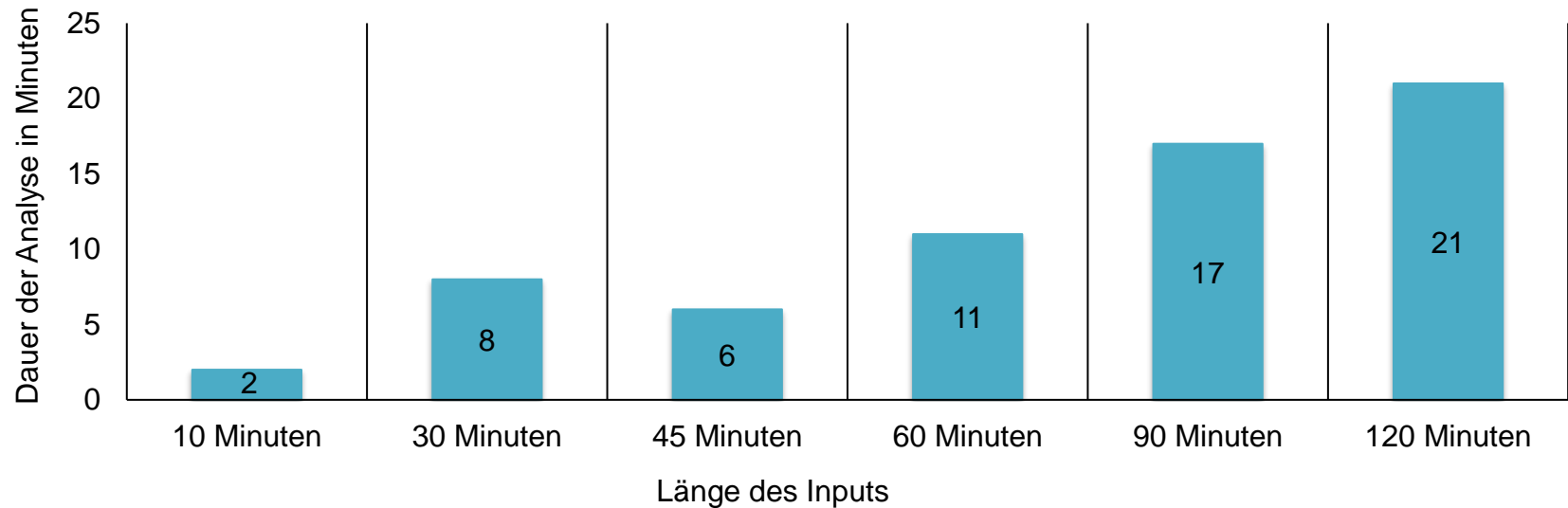
Wie sieht die Performance von CyVerse aus?

Analysedauer auf CyVerse (in Minuten) mit englischem Input (.mp4)



Wie sieht die Performance von CyVerse aus?

Analysedauer auf CyVerse (in Minuten) mit deutschem Input (.mp4)



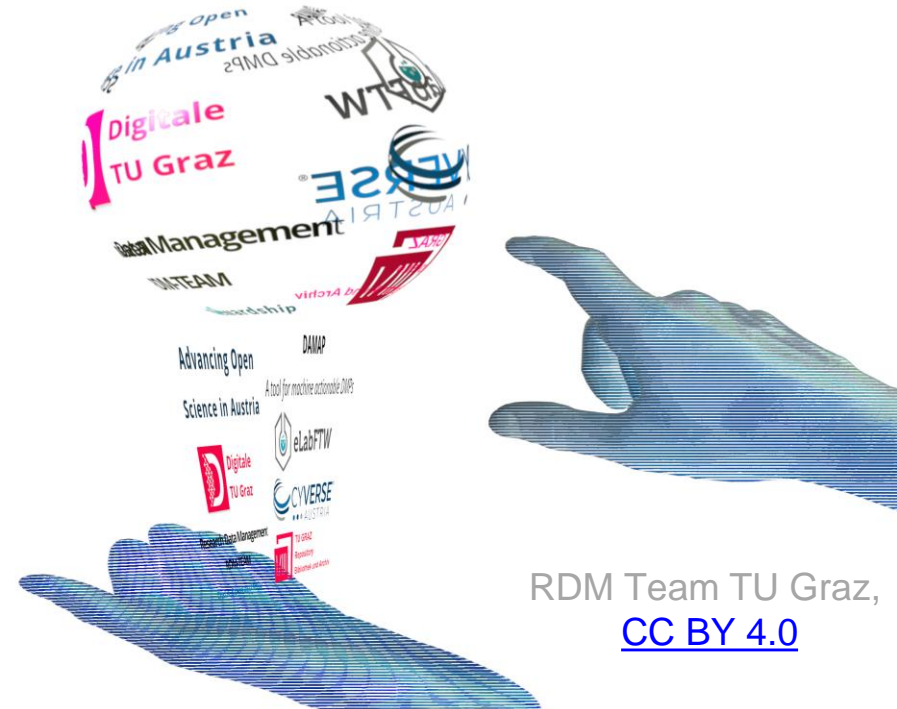
rdmteam@tugraz.at



<https://rdm.tugraz.at>



Supporting reproducible research



RDM Team TU Graz,
[CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)