

# BioTechMed Kickoff

## HRSM – Integriertes Datenmanagement

30.09.2020

Prof. Dr. Stefanie Lindstaedt

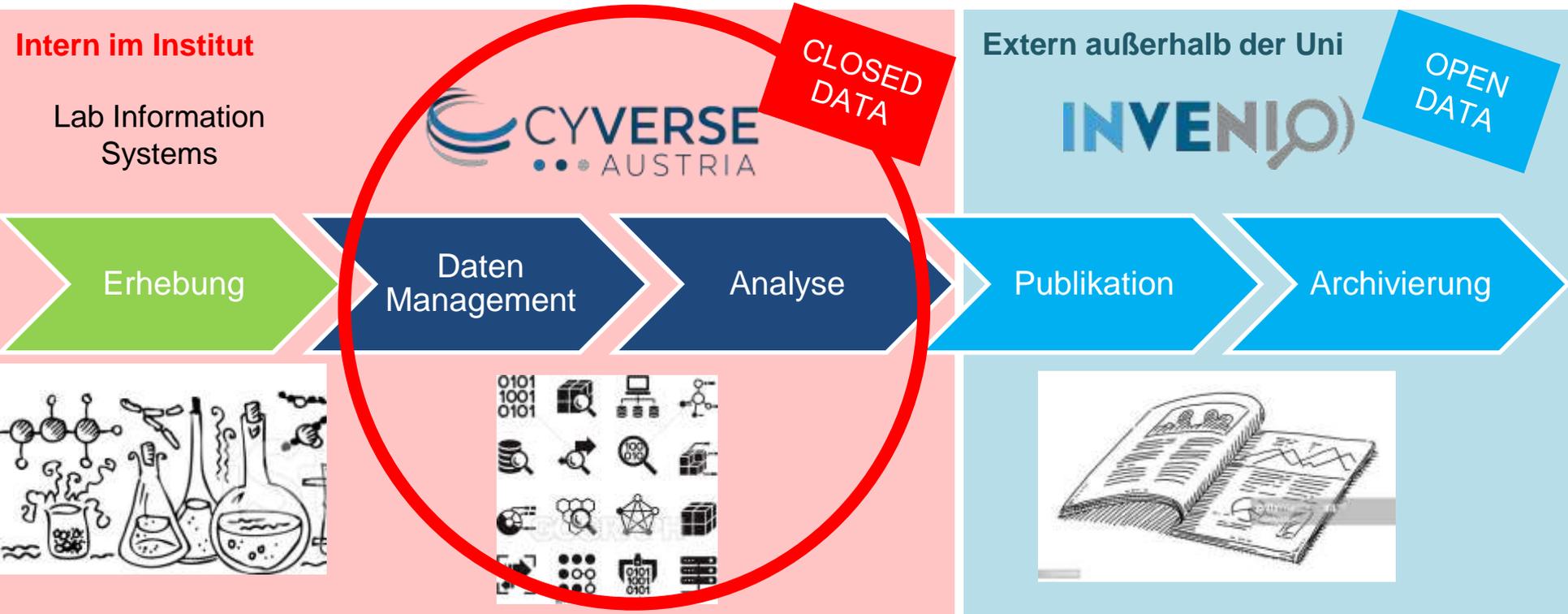
# Das erfolgreiche, intra-Universitäre Team



+ David Bodruzic  
u. Max Matzinger

+ Richard  
Hohensinner

# Daten im Forschungsprozess und die Rolle von CyVerse



Data: raw\_data\_subset

Upload • File • Edit • Download • Share • Metadata • Refresh Trash

- Navigation raw\_data\_subset
- cccc
  - Cyverse\_use\_case
  - raw\_data\_subset
  - UC\_Galaxy
  - UC\_Jupyter\_QIIME
  - UC\_QIIME2
  - analyses
  - data
  - demux\_paired\_end
  - test\_tomi
  - testdata
  - Community Data
  - cyverse\_training
  - example\_data
  - Shared With Me
  - Trash
  - Favorites

Viewing: /TUG/home/cccc/Cyverse\_use\_case/raw\_data\_subset

Name	Last Modified	Size
DBR-3_S35_L001_R1_001.fastq.gz	2019 Nov 12 11:01:08	19.63 MB
DBR-3_S35_L001_R2_001.fastq.gz	2019 Nov 12 11:01:08	24.21 MB
OKC-2_S34_L001_R1_001.fastq.gz	2019 Nov 12 11:01:08	12.02 MB

- Analyses
- Analyses
  - Name
  - JupyterLab\_0
  - JupyterLab\_Qiim
  - JupyterLab\_QIIME2\_R\_2019.10\_analysis1
  - limited\_0.0.2\_analysis1

Apps Manage Tools Switch View

Apps • Workflow • Share • Refresh Search Apps

Categories Topic Operation HPC

My Apps Topic Operation HPC

Apps under development

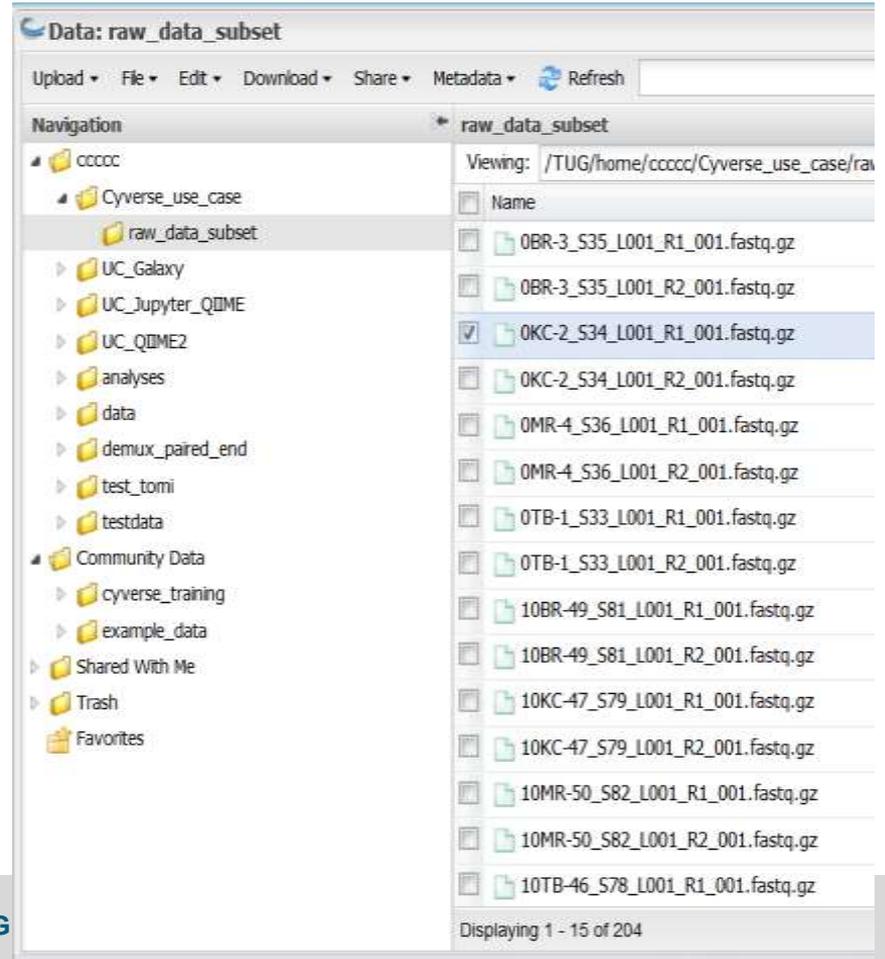
Name	Time
debian-buster-sim-python3... Konrad Lang ★★★★★ (0)	Download OSF Preprints 3.0 Konrad Lang ★★★★★ (0)
fmprepr Konrad Lang ★★★★★ (0)	JupyterLab Qiime2 R 0.1 - ... Konrad Lang ★★★★★ (0)

# Daten Management



## Data Store

- Web based user interface  
Zugriff von überall
- Daten hochladen über
  - iCommand (iRODS)
  - Cyberduck
  - WebDav
  - Drag & drop
- Logging der Zugriffshistorie
- Daten sichtbar machen für Kooperationspartner
- Nicht alle Daten müssen auf CyVerse liegen, können auch über Links eingebunden werden
- Metadaten zusammen speichern mit Daten
- Automatische Metadaten Extraktion (beschränkt)
- Flexible Nutzung von Speicher Ressourcen



The screenshot shows a web-based file manager interface for a directory named 'Data: raw\_data\_subset'. The interface includes a navigation pane on the left and a main file list on the right.

**Navigation Pane:**

- cccccc
  - Cyverse\_use\_case
    - raw\_data\_subset
      - UC\_Galaxy
      - UC\_Jupyter\_QIIME
      - UC\_QIIME2
      - analyses
      - data
      - demux\_paired\_end
      - test\_tomi
      - testdata
    - Community Data
      - cyverse\_training
      - example\_data
    - Shared With Me
    - Trash
    - Favorites

**Main File List:**

Viewing: /TUG/home/cccccc/Cyverse\_use\_case/raw\_data\_subset

Name
<input type="checkbox"/> OBR-3_S35_L001_R1_001.fastq.gz
<input type="checkbox"/> OBR-3_S35_L001_R2_001.fastq.gz
<input checked="" type="checkbox"/> OKC-2_S34_L001_R1_001.fastq.gz
<input type="checkbox"/> OKC-2_S34_L001_R2_001.fastq.gz
<input type="checkbox"/> OMR-4_S36_L001_R1_001.fastq.gz
<input type="checkbox"/> OMR-4_S36_L001_R2_001.fastq.gz
<input type="checkbox"/> OTB-1_S33_L001_R1_001.fastq.gz
<input type="checkbox"/> OTB-1_S33_L001_R2_001.fastq.gz
<input type="checkbox"/> 10BR-49_S81_L001_R1_001.fastq.gz
<input type="checkbox"/> 10BR-49_S81_L001_R2_001.fastq.gz
<input type="checkbox"/> 10KC-47_S79_L001_R1_001.fastq.gz
<input type="checkbox"/> 10KC-47_S79_L001_R2_001.fastq.gz
<input type="checkbox"/> 10MR-50_S82_L001_R1_001.fastq.gz
<input type="checkbox"/> 10MR-50_S82_L001_R2_001.fastq.gz
<input type="checkbox"/> 10TB-46_S78_L001_R1_001.fastq.gz

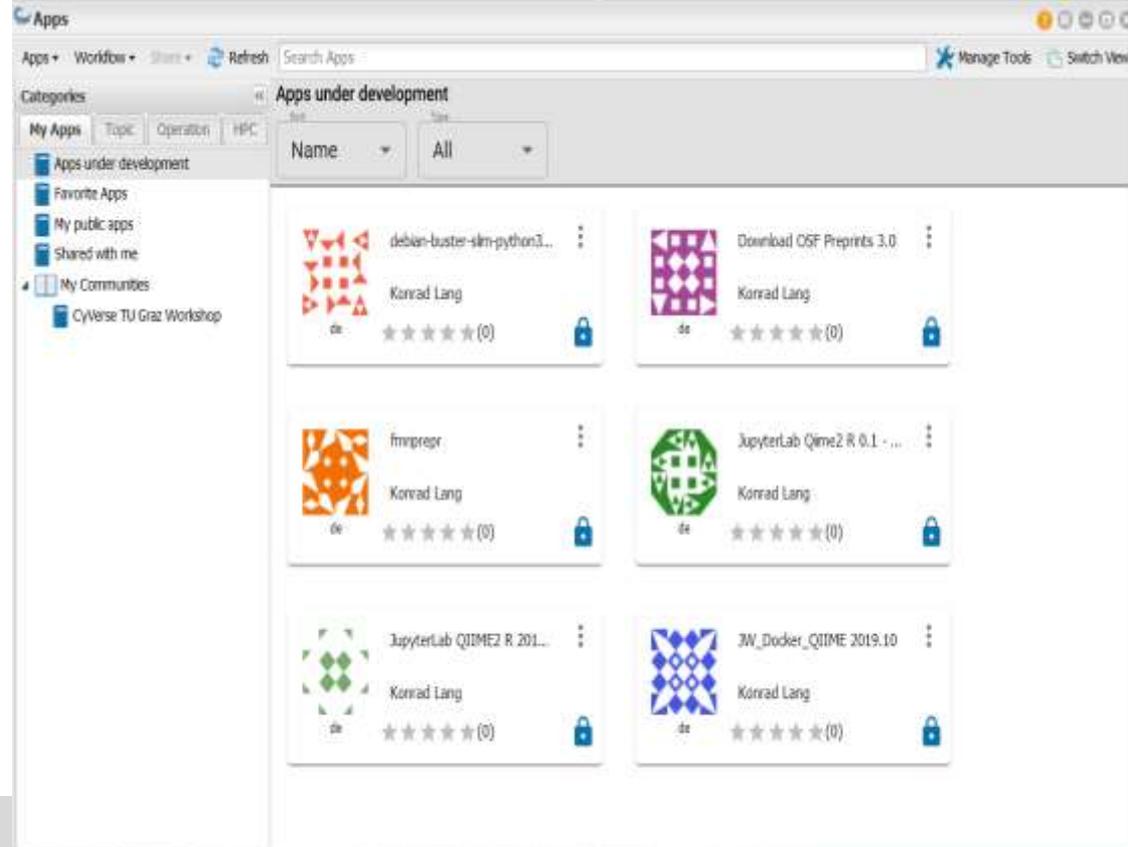
Displaying 1 - 15 of 204

# Daten Analyse



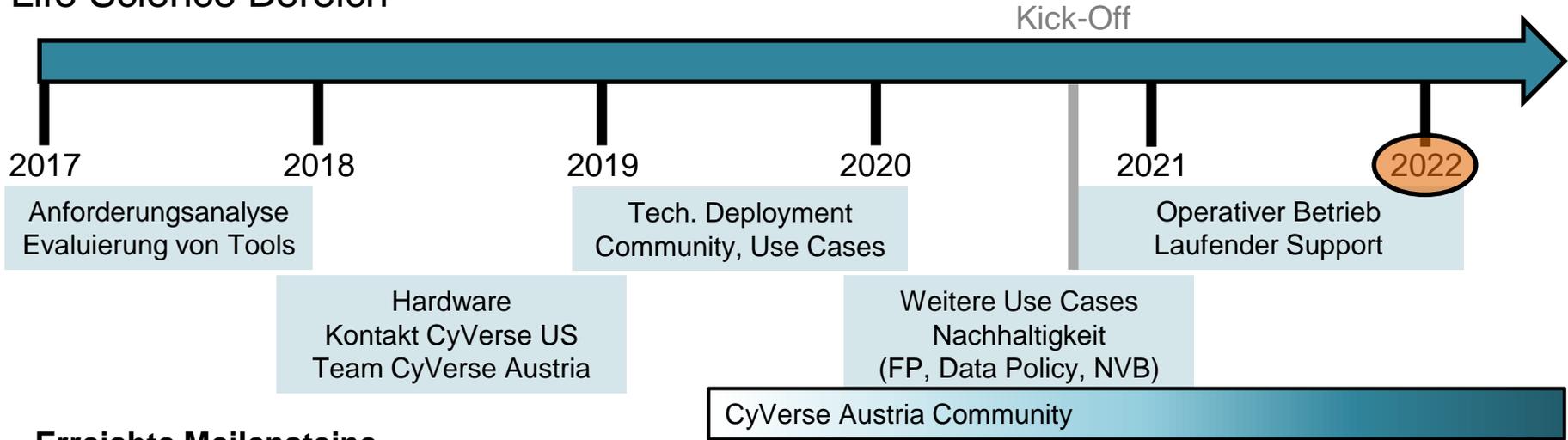
## Discovery Environment - Analytics

- Web-basiertes user Interface  
Zugriff von überall
- > 3000 Analyse Apps stehen zur Verfügung
- Einbinden eigener Analyse Algorithmen möglich
- Speicherung der Analyse zusammen mit den richtigen Versionen der Apps (Reproducibility)
- Einfache Erstellung von Analyse-Pipelines
- Flexible Nutzung von Computational Resources
- VSC kann angesprochen werden
- Am VSC wird kein Speicher benötigt



# Vorgehensweise

**Ziel:** Stärkung von Kollaborationen und Datenanalysen für Grazer ForscherInnen im Life Science Bereich



## Erreichte Meilensteine

(i) Technisches Deployment, (ii) aktive Community, (iii) Trainings und (iv) international Vernetzung  
 Jährliches Reporting zum BMBWF, on-track

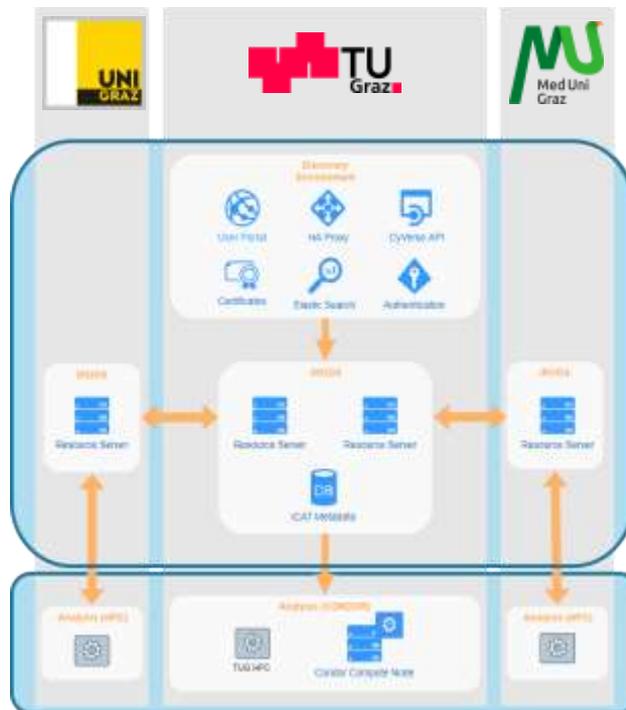
# CyVerse Austria Infrastruktur in Graz

**CYVERSE AUSTRIA**

LABS → DATA → ANALYSIS → PEOPLE

- iRODS**  
Virtualize data storage resources
- Data Store**  
Easy-to-use data management
- Discovery Environment**  
User interface to access/manage

**Data Management**



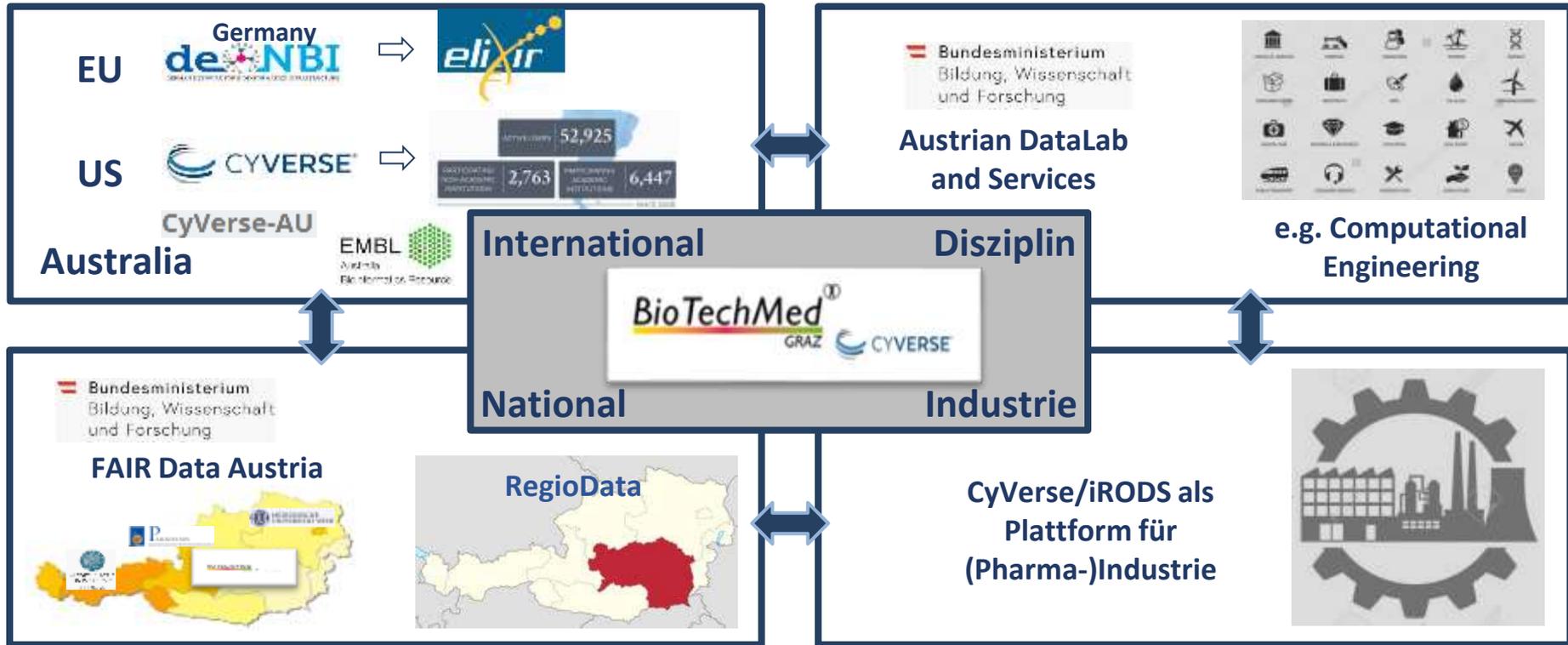
**VICE**  
Visual and Interactive Computing Environment

**HTCondor**  
job scheduling, resource mgmt

**App container**  
Integrated Dockerized tools  
**Data staging-in and out**

**Data Analytics**

# Vision



# Regelmäßige Trainingsaktivitäten

- On-site Training mit CyVerse US
- Carpentries Training
  - **Grazer Universitäten sind jetzt Teil der globalen Carpentries Community**
- Dockertraining
  - **Online Training, um ForscherInnen auch im Homeoffice zu unterstützen**

Feedback Questions	Out of 5
How did this workshop meet your expectations?	4,4
How helpful were the workshop instructors?	4,8
How would you rate the CyVerse infrastructure?	4,2



# CyVerse Austria – Homepage

<https://cyverse.tugraz.at>

## About

CyVerse provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. Our extensible platforms provide data storage, bioinformatics tools, image analysis, APIs, and more.

CyVerse Austria uses open source code from [CyVerse US](#) (University of Arizona) and its deployment is funded by [BioTechMed Graz](#). Within the program [Digitale TU Graz](#) the follow-up project for expansion of CyVerse Austria (Austrian Data Lab and Services) is funded by the [DFG](#).

## CyVerse and the COVID-19 Pandemic

During the COVID-19 Pandemic, CyVerse Austria will strive to maintain full operations. This includes providing help to researchers and educators working remotely. If you have special requests during this time, please [contact us](#).



### Teresita Ceresa (Graz University of Technology, Institute of Environmental Biotechnology)

I am an assistant professor at the Graz University of Technology in the Institute of Environmental Biotechnology. Our scientific work is mainly focused on the fields of life science research, Plant Protection and Bioprocessing, to derive the research areas an employ established biotechnology and molecular biology techniques as well as new niche technologies. With our scientific efforts we are developing new strategies and applications for biotechnology products. We're also very collaborative and have various projects with other teams, where efficient and secure data sharing becomes essential for us. Therefore, I joined the CyVerse Austria workshop, to become more familiar with data management and sharing on CyVerse, as well as shared analysis using Docker technology.

### Clemens Drexler (University of Graz, Institute of Molecular Biosciences)

I am a staff scientist in the core facility for preclinical MRI at the Karl Franzens University of Graz. In my research we preclinical in vivo imaging to investigate a variety of biological research questions such as lipomas and their role in lipid associated disorders, polyploidy in aging and disease and prostate based cancer treatment. Preclinical in vivo imaging is an important step in the evaluation process of biological hypotheses in health and disease of higher organisms (mainly small mammals). It offers a broad in vivo acquisition of quantitative and qualitative (image based evidence) biomarkers, in vivo. Among longitudinal investigations within a subject (e.g. animal) through it is now possible to realize if one of the most powerful tools to take a deep look into physiology, function, structure and organ development in vivo. I think CyVerse Austria can be a very useful tool to support my research. I've acquired imaging dataset as part of an animal experiment. Therefore every image is of high value in the animal ethics and from a financial point of view. MRI acquisition comes along with a huge amount of Metadata. They mainly describe and parameterize the performed image acquisition and post processing steps. Documentation of these steps together with archiving scan data together with the corresponding resultant images, applied segmentation and the quantitative results would be an application in thinking!

# Unsere aktive und motivierte User Community

*“I think everything was perfect. Nice pace, depth, time for hands on work on personal ideas/projects. Thanks!!”*

*“Besides getting impressions and links of many useful tools and trainings I think learning the basics about Docker and git(hub) was the most important for my future tasks. “*

*“more resources and computing capacity!!!”*

*“it was easy to follow. perfect sped, perfect examples chosen. and the web-documentation for reading through it tomorrow again is also veery nice, please keep it online for some time or please send them out by email as one bundled PDF. ”*

*“Thank you, sarah, philipp & christoph! Very informative talks & live demonstrations!”*

*“thanks for the cool training. you made a very good job. looking forward for part ii. ”*



## Kommunikation

- Regelmäßige Treffen (alle zwei Monate)
- Slack Channel mit CyVerse US
- Zammad CyVerse Austria Support

# Nächste Schritte für CyVerse Austria

- **Nachhaltigkeit gewährleisten durch Finanzierungsplan**
  - Entwicklungsplan, Leistungsvereinbarungen der Universitäten
- **Regulierung der Nutzung durch Data Policy**
- **Atmosphäre Modul**
  - OpenStack für Cloud Computing
  - Federated Cloud – in Entwicklung
- **Ausweitung**
  - Mehr Disziplinen und Universitäten (ADLS Projekt)

# Vielen Dank für die Aufmerksamkeit!

Kontakt:

- Stefanie Lindstaedt, [slind@know-center.at](mailto:slind@know-center.at)



Institute of Interactive Systems and Data Science  
Know-Center GmbH

## Weitere Aktivitäten

- Publikationen über CyVerse Austria
  - MDPI MCA Special Issue – CyVerse Austria Infrastruktur
  - Use Case Reports
- CyVerse Austria in Forschungsproposals
  - FWF Einzelprojekt Open Science Badges (TUG, KFU)
  - EC Proposal (MUG, Gernot Plank)
  - FWF Zukunftskolleg (MUG, KFUG, TUG)
  - ...