Bio-IT, BioNT and Data Science

Supporting computational biology at EMBL

Renato Alves

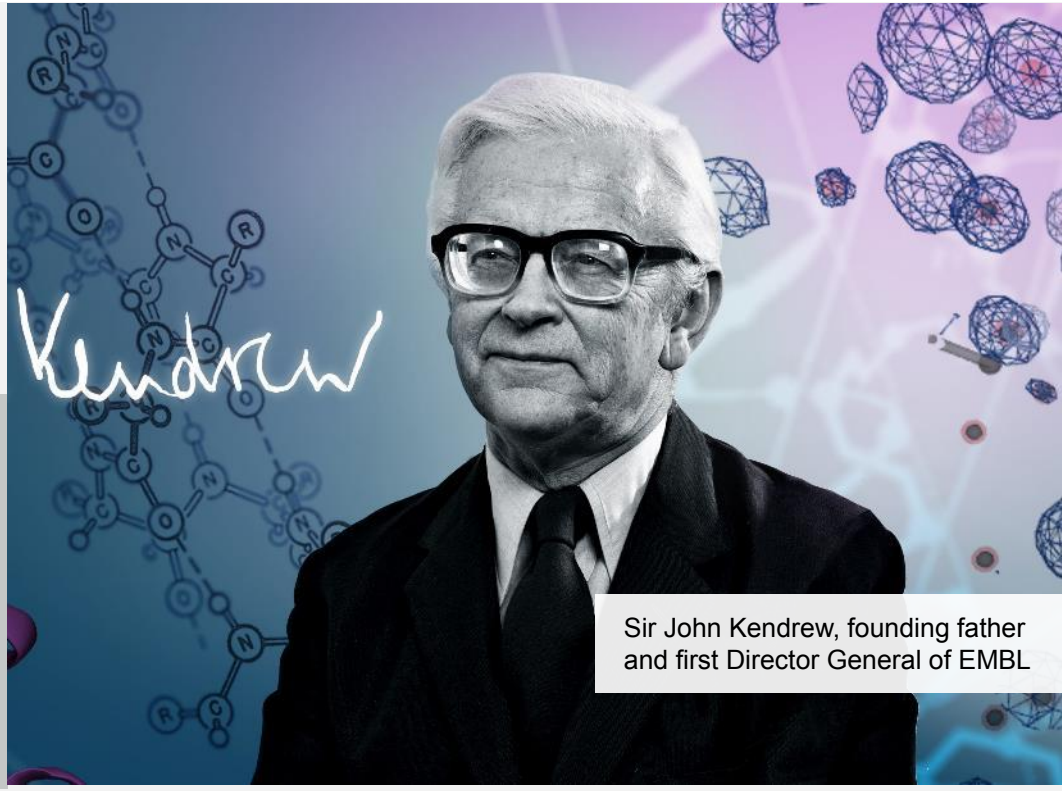
EIROforum - Workshop on knowledge management, 6th June 2024

EMBL 50 YEARS | 1974–2024

What is EMBL?

Europe’s centre of excellence in life science research, services and training

Founded in 1974 by 10 states (now 29 + 1 prospect) as an intergovernmental organisation to promote the molecular life sciences in Europe and beyond.



Sir John Kendrew, founding father and first Director General of EMBL

EMBL sites – over **1700** people and more than **80** nationalities

EMBL-EBI
Bioinformatics

Grenoble
Structural biology

Barcelona
Tissue biology and disease modelling

Hamburg
Structural biology

Heidelberg
Life sciences

Rome
Epigenetics and neurobiology

EMBL 50 YEARS | 1974–2024

European Molecular Biology Laboratory

Knowledge exchange challenges

pan-EMBL activities

Hybrid/digital

Diverse needs

Multidisciplinarity

Rotation of personnel

Jargon

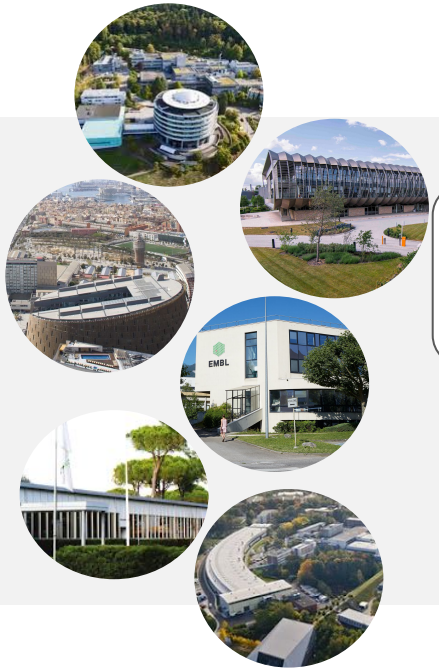
Overload of options

COMMUNITY

TRAINING

TOOLS

INFORMATION



The EMBL Bio-IT project

European Molecular Biology Laboratory

Community initiative to build, support and promote **computational biology research** at EMBL

EMBL Bio-IT

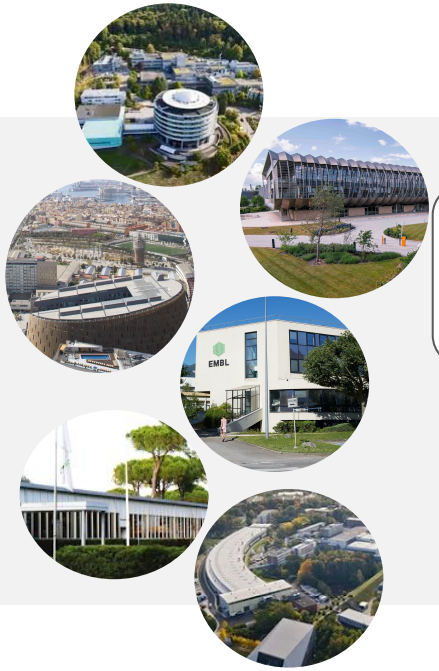
The Project Managers:

Renato Alves

Lisanna Paladin

+ *BioNT project manager (Isabela Paredes Cisneros)*

EMBL 50 YEARS | 1974–2024



The EMBL Bio-IT project

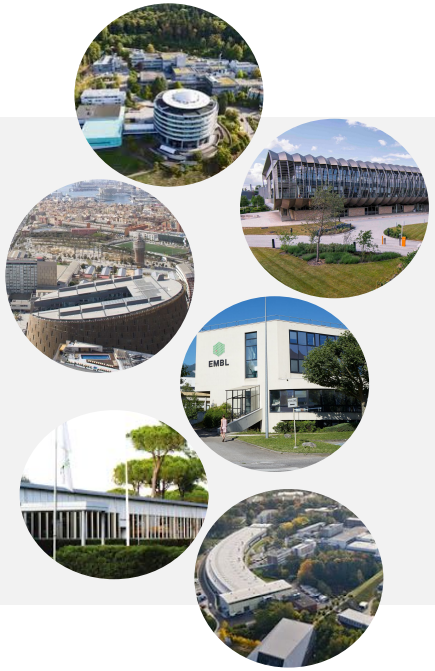
European Molecular Biology Laboratory

Community initiative to build, support and promote **computational biology research** at EMBL

EMBL Bio-IT

#community

EMBL 50 YEARS | 1974–2024



Data Science Centre

Managing, analysing and sharing data at a scale

Analysis Methods and Standards

- Artificial Intelligence
- Curated reference biodata sets
- Shared data analysis workflows

Integrated Data Management

- Data life cycles

Provision of Public Data Services

- Data resources
- EMBL tools

Technical Infrastructure

- Data processing compute infrastructure
- Network
- Storage

People and training

- Internal training and career development
- External training

EMBL 50 YEARS | 1974–2024

Bio-IT website: information hub

bio-it.embl.de

Training

- Courses and events
- Course materials
- Consulting sessions

Community

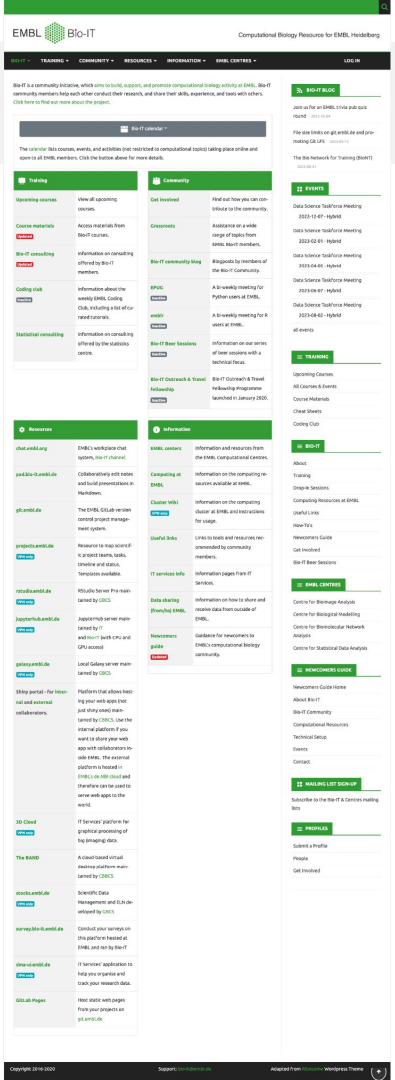
- Coding clubs
- Grassroots
- Community blog

Resources

- EMBL chat
- GitLab
- Coding platforms
- Notes & survey tools
- Other resources

Information

- EMBL centers
- IT services info
- WikiHows
- Newcomers guide



EMBL 50 YEARS | 1974–2024

COMMUNITY

The World of
Molecular Biology
Die Welt der Molekularbiologie

Taskforce & community meetings

Hybrid format

CHANGED 19 HOURS AGO

EDITABLE

Bio-IT Taskforce Meetings

Click here to join Zoom
Meeting ID: 963 1820 3177
Password: 604455

Online shared timer

Next taskforce meetings

- 2024-01-10 14:00
- 2024-03-06 14:00
- 2024-05-08 14:00
- 2024-07-03 14:00
- 2024-09-04 14:00
- 2024-11-06 14:00

Past taskforce meetings

- 2023 and prior years

2024-01-10 14:00 (German-time) - hybrid

Agenda

- Introductions & newcomers
- The Lean Coffee model

Todo from past meetings

▸ Expand list

Attending

Please add your name here.
Create new bullet points if necessary:

CHANGED A FEW SECONDS AGO

OWNED THIS NOTE

EDITABLE

Data Science Community Meeting

click here to join the zoom meeting

```
Meeting ID: 823 807 3039
Passcode: r9V9NE
```

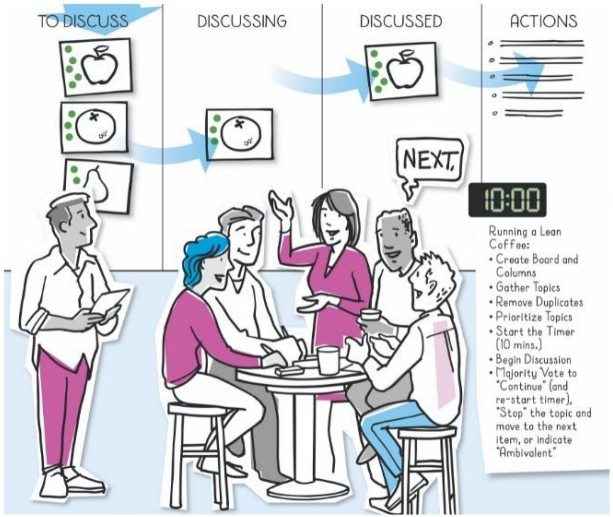
Next DS community meetings

- 2024-02-07 14:00
- 2024-04-03 14:00
- 2024-06-05 14:00
- 2024-08-07 14:00
- 2024-10-02 14:00
- 2024-12-04 14:00

Past DS community meetings

- 2023-12-06 14:00
- 2023-10-04 14:00

Lean Coffee model



Coding clubs

Hybrid format

EPUG	A bi-weekly meeting for Python users at EMBL.
emblr	A bi-weekly meeting for R users at EMBL.

Programming Clubs: EPUG and emblr

- Every Tuesday, 15:00 Germany time
- Alternating Python / R
- Introduce new topics?
  - Image analysis

Workflow Management Systems Club

New

- Monthly, first Thursday, 10:00 Germany time

BioinfoRome

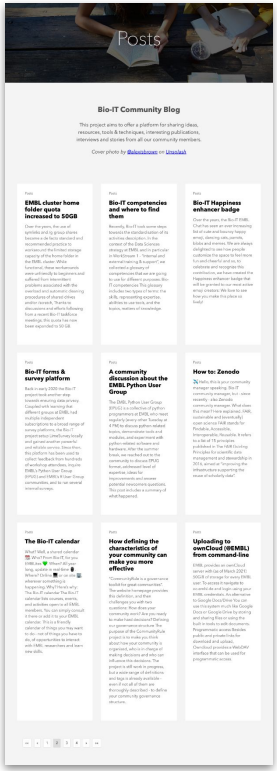
- Flexible schedule, discussion about technical questions

Community blog and Newsletter

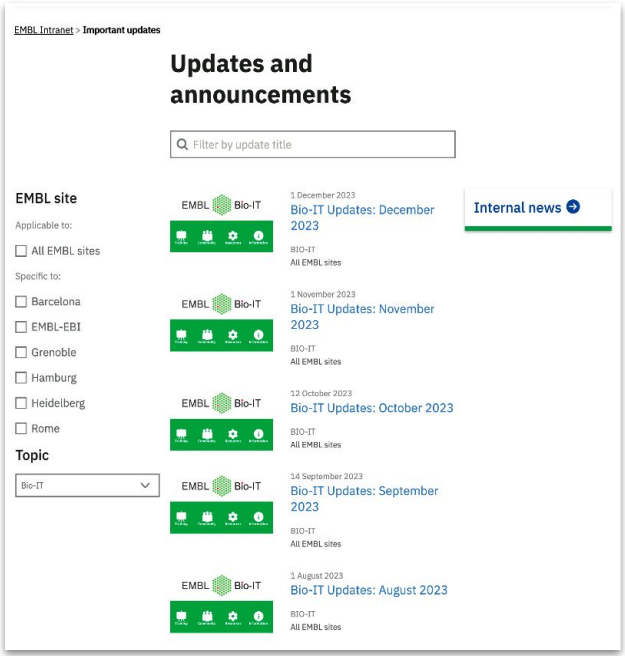
grp-bio-it.embl-community.io/blog

Bio-IT Community Blog

This project aims to offer a platform for sharing ideas, resources, tools & techniques, interesting publications, interviews and stories from all our community members.



embl.org/internal-information

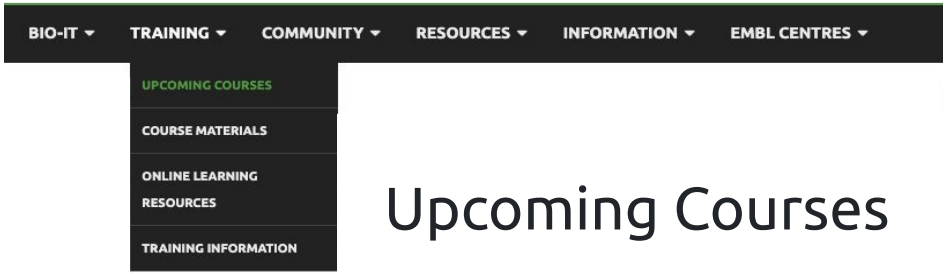


EMBL 50 YEARS | 1974–2024

TRAINING

Rome Building 14 - Meeting
Matthias
Kristina
Puzna Gadre
Niko
Nadine Tochter
Basilia Aicardi
Saravanan Devendran
LATER.

Courses and events



Upcoming Courses

List of course and events that are planned, date/time and place (in person / hybrid / online)

Hybrid format

Special mention: Bio-IT and Data Science Taskforce Meetings

Training plan 2024

List of courses that we plan to organise, including estimated dates, status, links to materials/notes, lead and team

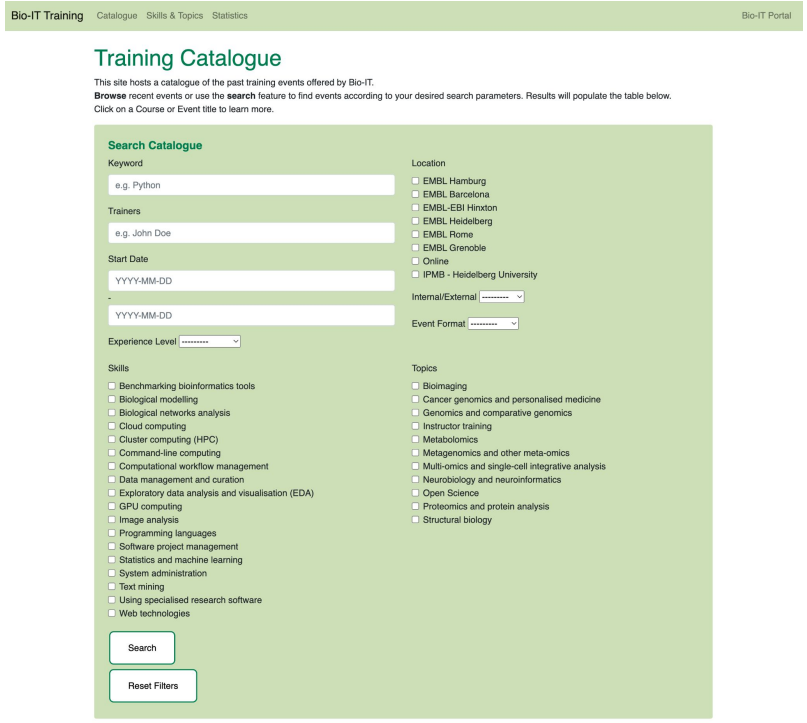
Invite to express training needs through the Bio-IT catalogue

VPN only

EMBL 50 YEARS | 1974–2024

Courses and events

catalogue.bio-it.embl.de



Course: Version Control with Git

Course Overview

VPN only

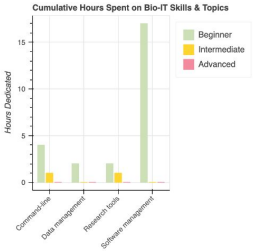
Description

Do you struggle to keep track of the changing versions of your scripts/programs? Have you ever made changes to a file, only to wish shortly afterward that you could easily get back the old version? Do you want to learn how to harness the awesome features of GitLab for collaboration, issue tracking, software publication, etc? This course aims to teach good practice in version control, using git in conjunction with the EMBL GitLab system.
Participants will learn: how to maintain a history of changes with Git and how to compare differences between versions; how to restore old versions of files; how to synchronise local versions with a remote repository; how multiple developers can collaborate effectively on a project, and how to resolve conflicting changes; how to manage a project through the GitLab interface; how to choose the right license for a project; how to work with Git directly within your development environment (e.g. RStudio). The course will provide an introduction to Git via three different interfaces: the command line, a graphical client, and GitLab.

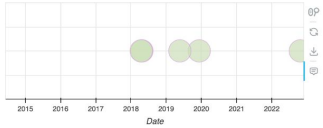
By the Numbers

Events Offered		5
Course Attendance	Minimum	11
	Maximum	37
	Total	124
Course Duration (Days)	Minimum	1
	Maximum	2
	Total	6

Curious how your time will be spent?



Timeline of Course Offerings (2018 to Present)



Interested in attending this course?

Click the button below to indicate your interest and join the mailing list for upcoming instances of the course.

I'm Interested!

Latest Course Resources:

Bio-IT Record (2022-10-24):
https://bio-it.embl.de/events/version-control-with-git-git-embl-de-git-week-basic-module/
Git Repository (2022-10-24):
https://github.com/swcarpentry/git-novice
Reference materials (2022-10-24):
https://pad.bio-it.embl.de/sd2Mk8YVS8uMg3KI-nL6yA

Events Offered

Version Control with Git & git.embl.de I Git week, basic module

Bio-IT Record: https://bio-it.embl.de/events/version-control-with-git-git-embl-de-git-week-basic-module/
Git materials: https://github.com/swcarpentry/git-novice
Reference materials: https://pad.bio-it.embl.de/sd2Mk8YVS8uMg3KI-nL6yA

Start Date	2022-10-24
Course Duration (Days)	2
Course Attendance	27
Instructor(s)	Lisanna Paladin, Renato Alves
Location(s)	EMBL Heidelberg Online
Reach	Internal
Course Level	Beginner
Hours per Skill	Beginner Command-line computing: 3 Beginner Software project management: 3 Beginner Using specialised research software: 2

EMBL 30 YEARS | 1974–2024

Consortium project

Partners: 6 academic, 3 SME, 4 countries

BioNT

TRAINING AND PROFESSIONALISATION IN COMPUTATIONAL LIFE SCIENCES

biont-training.eu

Coordinated by:

Isabela Paredes Cisneros

1. Bioinformatics introduction
2. Introduction to Python
3. Command line and cluster computing
4. Open and FAIR, data management

Workshops in the Basic Curriculum

Community event!

Long-term sustainability: Online training material for self-paced learning with translations

June 2023 | December 2023 | June 2024 | December 2024 | June 2025 | December 2025 | March 2026

Workshops in the Advanced Curriculum

9. Instructors training
10. Software best practices
11. Machine learning of biological data
12. System administration

Co-funded by the European Union

EMBL 50 YEARS | 1974–2024

Course materials

Course Materials

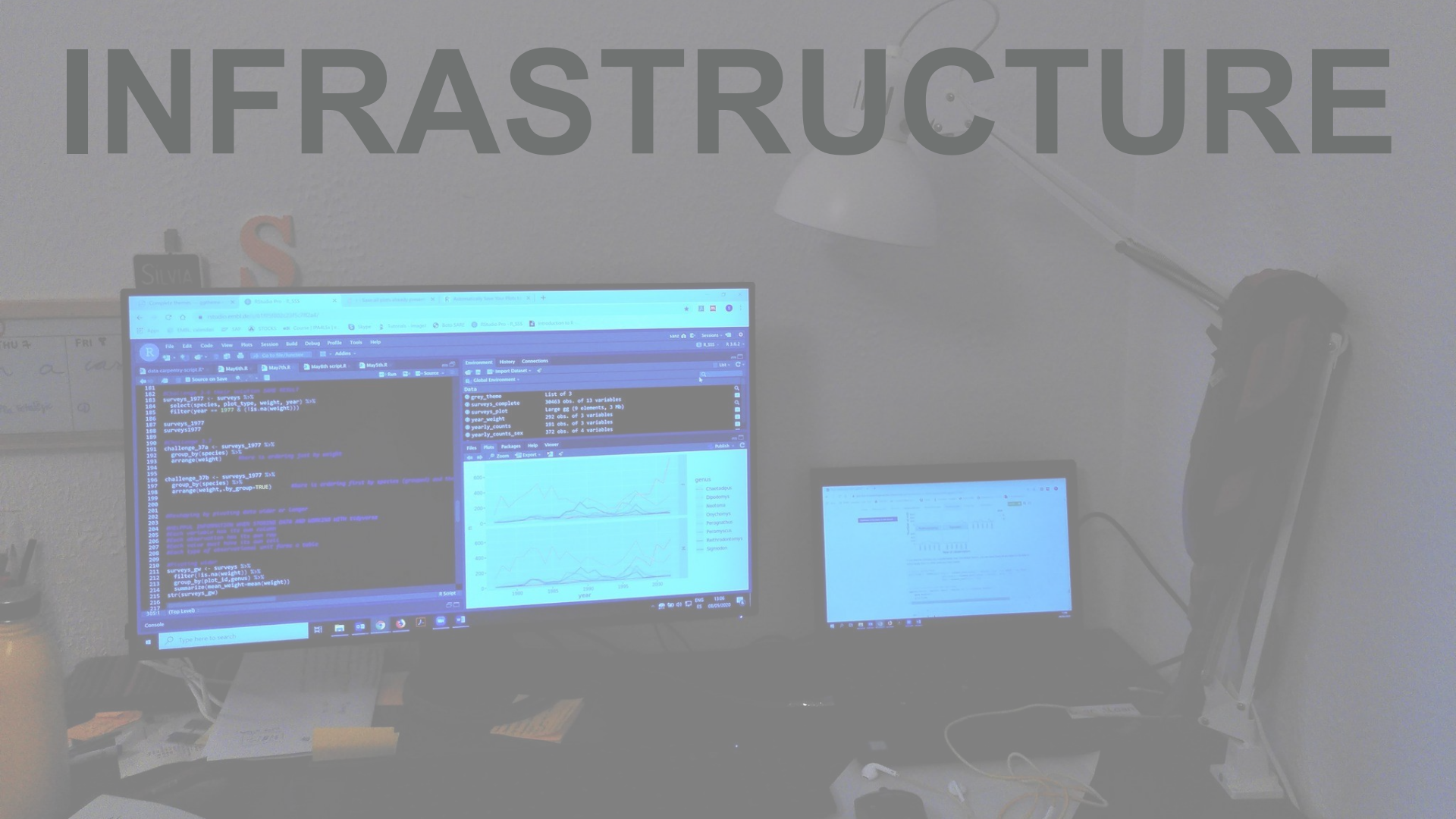
Links to training materials, in different formats when available and contact of the main maintainer.

Sections:

- Bio-IT courses
- Cluster training
- Singularity training
- CBNA courses
- CBM courses
- Predoc course

Recommended Resources for Online Learning

INFRASTRUCTURE



EMBL chat

chat.embl.org

Welcomebot BOT 3:06 PM

Awesome! I've added you to the following channels:

~Bio-IT 👉 Bio-IT channel for general bioinformatics, computational support
~Bio-IT-announce 👉 Bio-IT channel for course/workshop and Bio-IT activity announcements
~cluster 👉 EMBL Heidelberg HPC cluster support
~GBCS 👉 Genome Biology Computational Support channel
~Mathematical Modelling 👉 Genome Biology Computational Support channel
~STOCKS 👉 Support channel for STOCKS EMBL's ~GBCS electronic lab and data management system

You may also be interested in the following (click their name to join):
~Folding / AlphaFold 👉 A channel to discuss protein folding with AlphaFold and related technologies
~How could be done better 👉 A channel to discuss scientific figures and other kinds of scientific output in a constructive way

Awesome! I've added you to the following discussion channels for different programming related technologies:

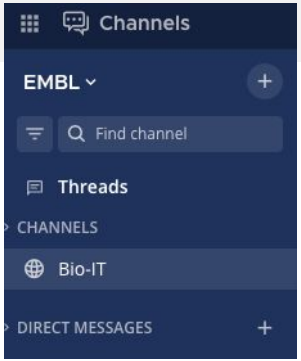
~Git 👉 Git and EMBL's GitLab
~Julia 👉 Julia
~MachineLearning 👉 For classic and deep learning discussions
~Python 👉 Python
~R_and_statistics 👉 R and statistics
~RStudio-Server 👉 RStudio EMBL server (by ~GBCS)

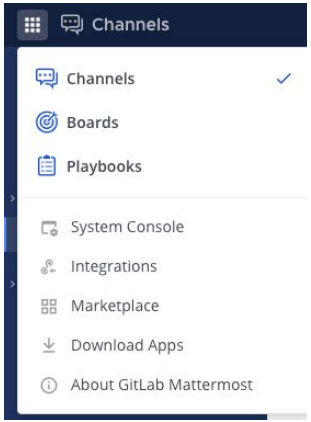
Awesome! I've added you to the following support channels:

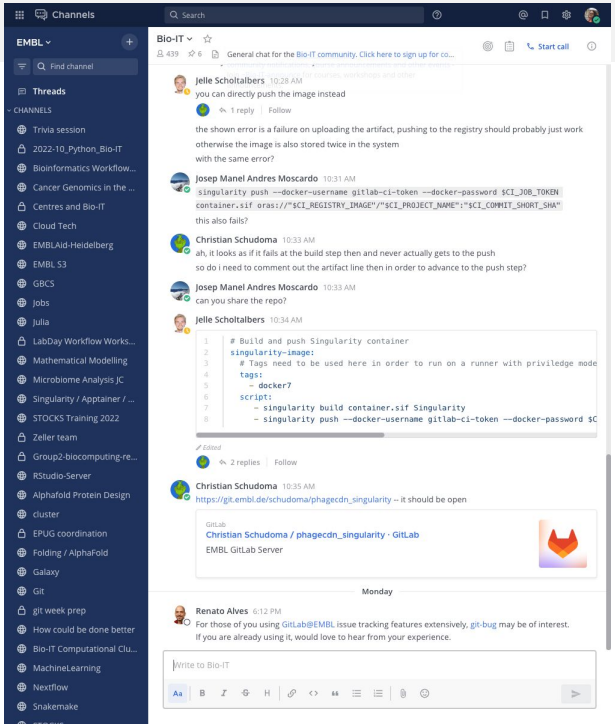
~Galaxy 👉 Galaxy@EMBL (by ~GBCS)
~Nextflow 👉 Nextflow workflow manager
~Snakemake 👉 Snakemake workflow manager
~Singularity / Apptainer / Docker 👉 Singularity containers

Awesome! I've added you to the following channels:

~Jobs 👉 Announce jobs and other opportunities in or outside of EMBL.
~marketplace 👉 General trading, buy&sell, giveaways, lost&found, ...
~EMBLAid-Heidelberg 👉 EMBLAid channel for Heidelberg (see also channels for other sites).







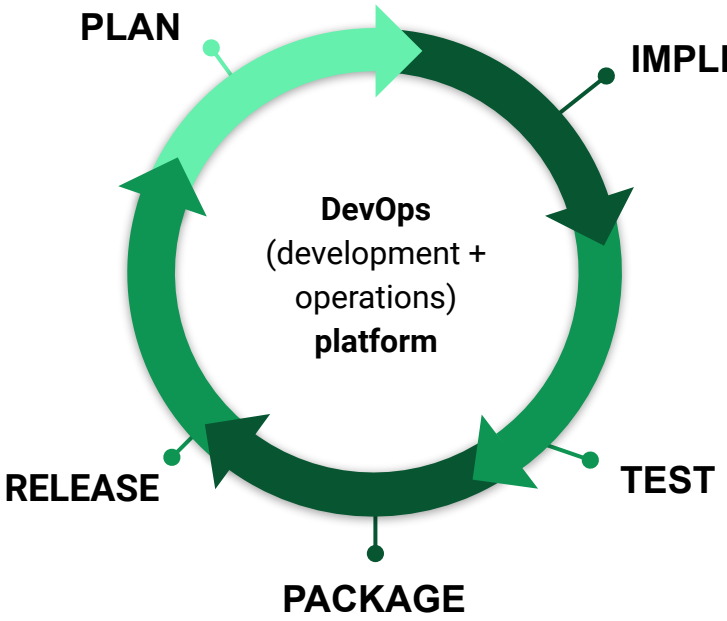
EMBL 50 YEARS | 1974–2024

GitLab

git.embl.de

Git = Version control system

- System that records changes made to a project (group of files) over time
  - Keeps versions in a smart way
  - Allows reverting back to a previous state
- Essential to (programming) projects involving several people
  - Allows branching (working separately on different features)
  - Allows to compare/merge changes made
  - Allows synchronization of repositories in different places (distributed version control)

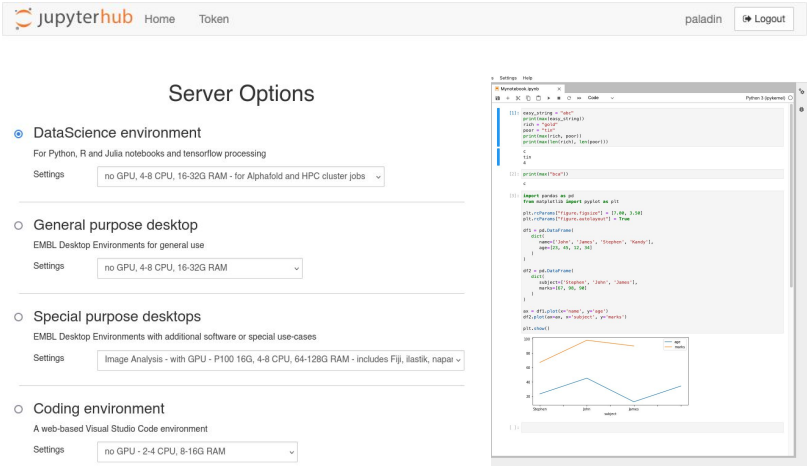


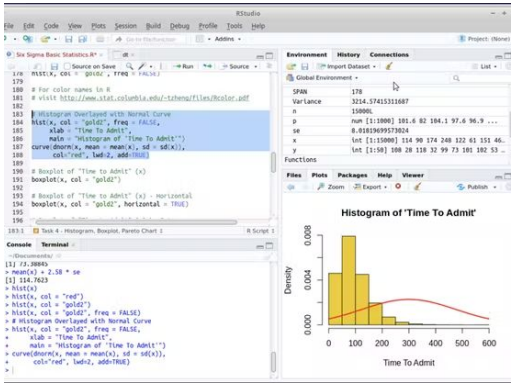
Coding platforms

jupyterhub.embl.de

jupyterhub

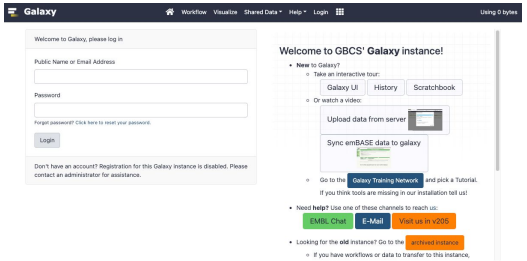
GBCS Genome Biology Computational Support





RStudio

rstudio.embl.de



Galaxy

galaxy.embl.de

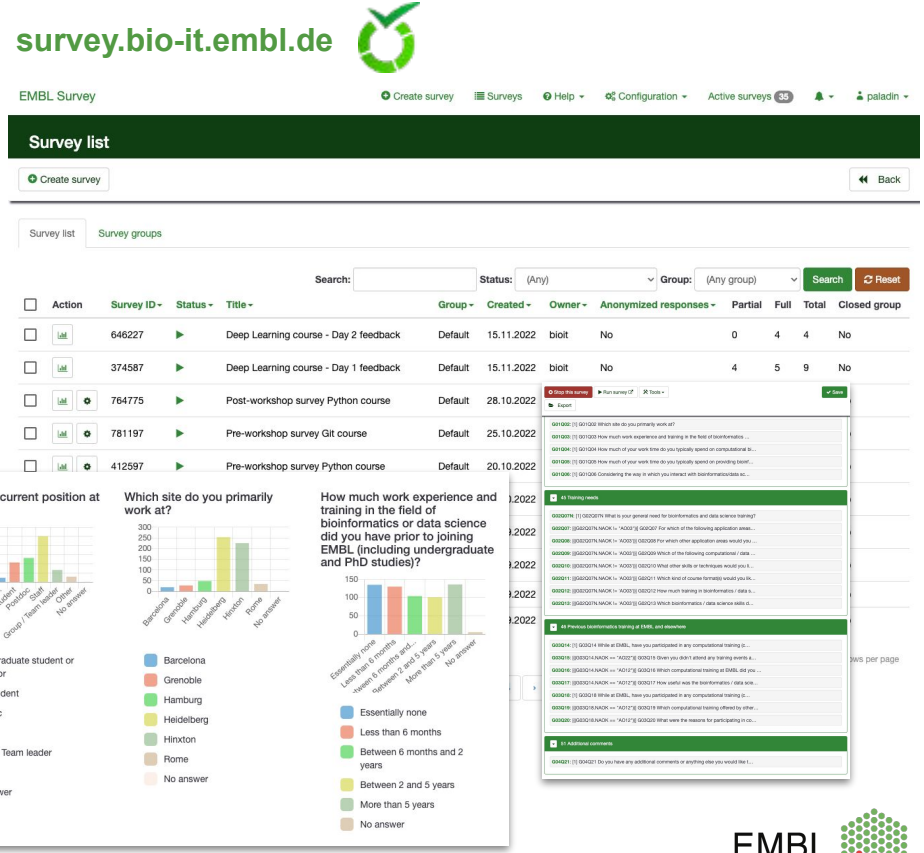
EMBL 50 YEARS | 1974–2024

Notes/Presentation and survey tools

pad.bio-it.embl.de

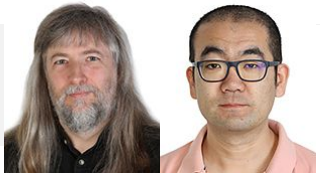
CodiMD

survey.bio-it.embl.de



EMBL 50 YEARS | 1974–2024

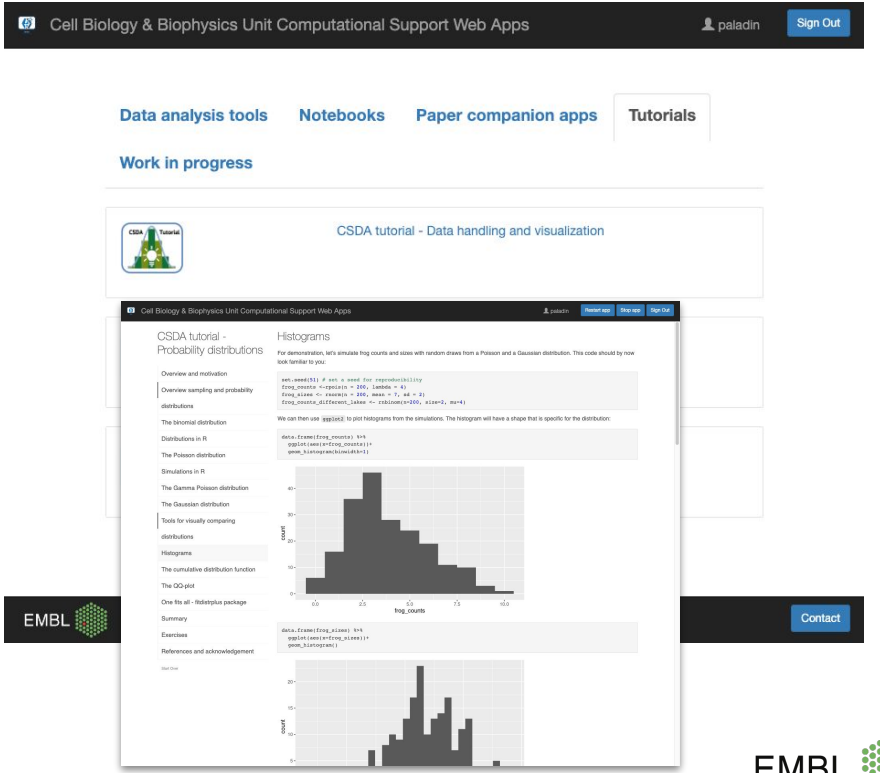
Other resources



Cell Biology and Biophysics Computational Support

band.embl.de

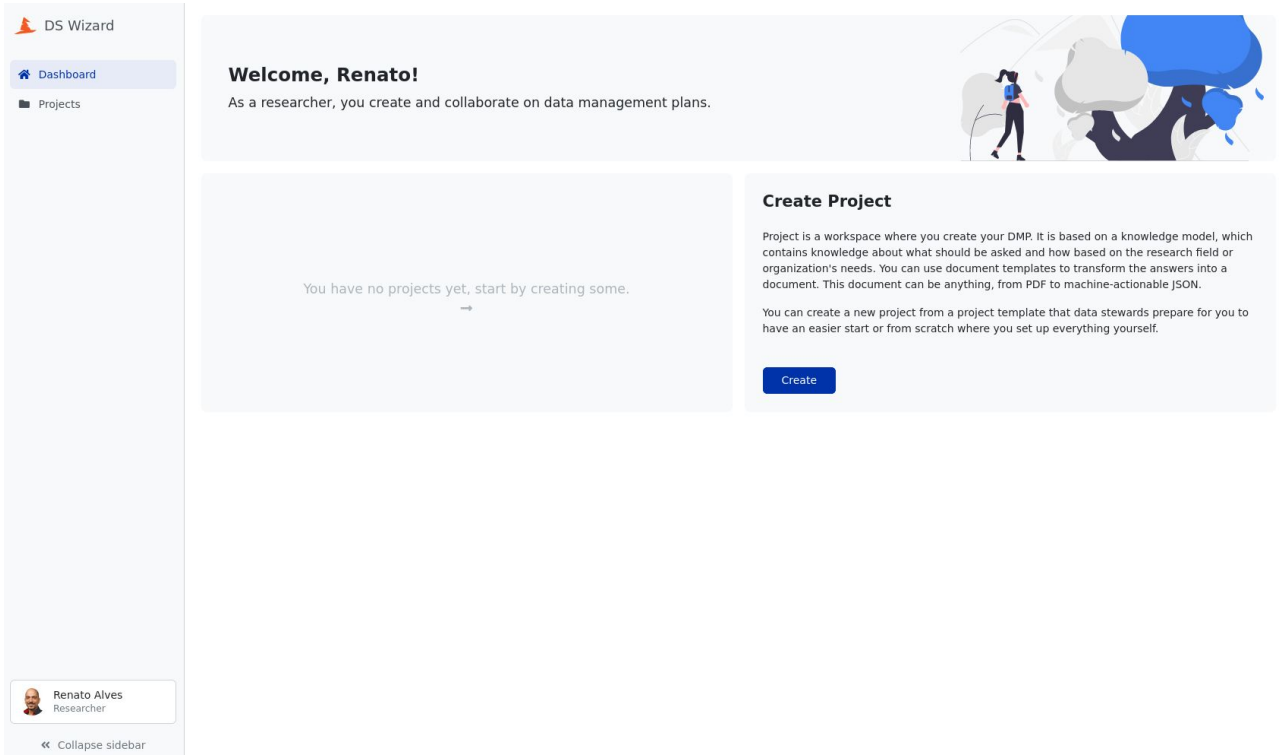
cbbcs-01.embl.de/shinyapps/



EMBL 50 YEARS | 1974–2024

Other resources

ds-wizard.embl.org



INFORMATION

Consulting sessions

Bio-IT Drop-In Sessions

Drop-in sessions with Renato Alves and Lisanna Paladin run **10:00-12:00 every Tuesday** in the cafeteria/staff lounge at EMBL Heidelberg and/or online via Zoom. Check the individual listings on the online events calendar for Zoom connection details.

For advice & help with:

- General bioinformatics
- Command Line
- Scripting
- Task/analysis automation
- Biocomputing at EMBL Heidelberg

Newcomers guide

Table of contents

- About Bio-IT Project
- Members of the Bio-IT Community
- Computational Resources
- Technical Setup
- Events
- Contacts

Bio-IT Newcomers Guide

Note that several linked pages in this document can be accessed only from inside the EMBL network

Table of contents

- About Bio-IT Project
- Members of the Bio-IT Community
- Computational Resources
- Technical Setup
- Events
- Contacts

About this document

This document is developed to help newcomers to learn all about computational systems at EMBL-HD, Bio-IT, and (almost) everything that will help them intergrate with the bio-computational community of EMBL Heidelberg.

The information in this guide overlaps with the material contained in the other documents useful for newcomers to EMBL. We also recommend that you check out the following resources:

- Bio-IT EMBL website for general information about Bio-IT activities.
- EMBL website for general information about EMBL.
- EMBL newcomer's booklet including information about first day best practices, housing, banking, medical service, chidlcare and family care, transportation, free time activities, additional informaton about Germany.
- EMBL Staff Association website, including SA events calendar, forums and working groups and initiatives for the EMBL-ites' well being.
- Heidelberg city guide, featuring city tour and events announcements.
- Szilard library guide, including the catalogue and access information.

If you can't find the information that you were looking for, please go ahead and directly interact with the EMBL-ites via the EMBL chat.

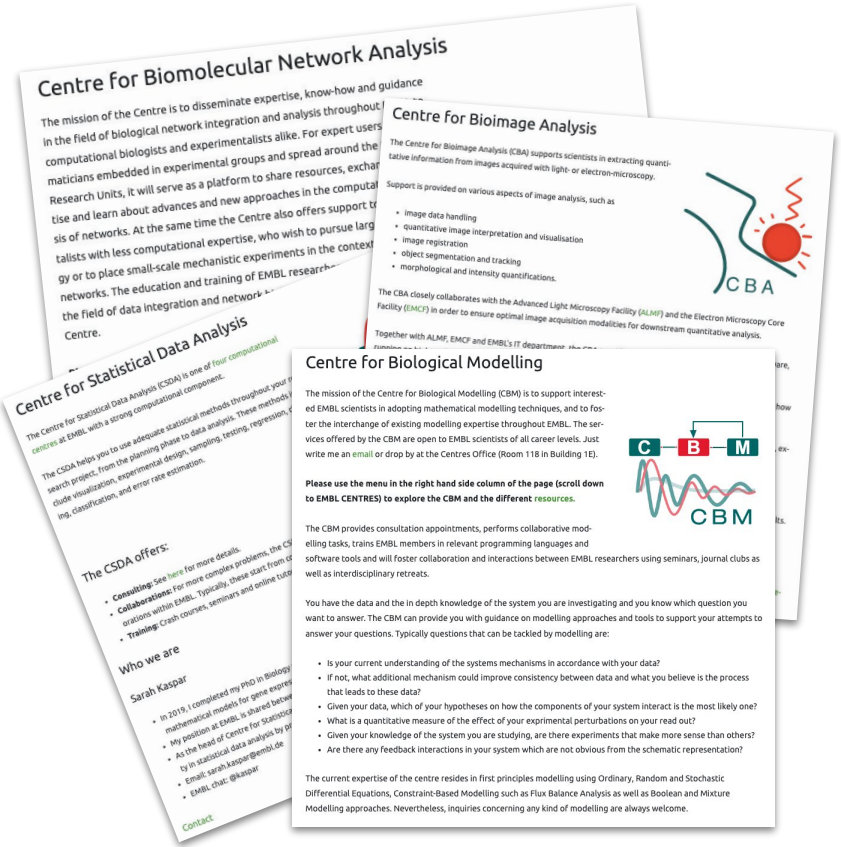
Acknowledgments

This guide borrows heavily (with permission) from the newcomers guides previously developed in the Zaugg and Huber Labs at EMBL. We thank the authors of those documents for their efforts and help.

Several sections also link out to the excellent Netherlands eScience Center Guide.

EMBL 50 YEARS | 1974–2024

EMBL centers



Bio-IT

Bio-IT

The Bio-IT community initiative provides support to computational biology research through training, courses, consulting, networking opportunities, and computational resources and tools.

Contacts: Renato Alves and Lisanna Paladin (bio-it@embl.de)
Stop by our Drop-In Sessions Tuesdays 10am-12pm

Centre for Bioimage Analysis (CBA)

The CBA supports you in extracting quantitative information from images acquired with light- or electron-microscopy.

Contacts: Christian Tischer, Sebastian Gonzalez, and Arif Khan (image-analysis-support@embl.de)
Send a mail to schedule a consultancy session.

Centre for Biological Modelling (CBM)

The CBM aims to support you to adopt mathematical modeling techniques into your everyday research.

Contact: Eva Geissen (eva.geissen@embl.de)
Write me an email to request an appointment or come over to Room 118.

Centre for Statistical Data Analysis (CSDA)

The CSDA helps you to use adequate statistical methods throughout your research project, from the planning phase to data analysis.

Contact: Sarah Kaspar (sarah.kaspar@embl.de)

Centre for Biomolecular Network Analysis (CBNA)

The CBNA disseminates expertise, know-how and guidance in network integration and analysis throughout EMBL.

Contact: Santhust (santhust@embl.de), Room 118

[*]:
```
while True:
    print("Thank you!")
```

```
Thank you!
Thank you!
Thank you!
```