



Training Day : Linux



Objectives

At the end of the day, you will be able to use Linux command line in order to :

- Connect to «genobioinfo» server
- Use available tools
- Transfer files between server and desktop
- Prepare data files
- Start processes with command line

Planning of the day

Part I : 09h00 - 10h30

Presentation of GenoToul bioinformatics core facilities, asking for an account, connection procedure, tree structure of files, command line syntax

Part II : 11h00 - 12h30

File types, permissions, manipulating files, displaying files, wildcard characters, disk space control

Part III : 14h00 - 17h00

Downloading/transferring, compressing/uncompressing, utility commands, redirections

Part I

- Presentation of GenoToul bioinformatics core facilities (mission, the team, the users, equipments, disk spaces, website)
- Introduction to Linux,
- Creating an account,
- Tree structure of files,
- Basic Linux commands,
- Connection procedure

Mission

Provide to public regional community :

Equipment

- Storage disk space
- High Performance Computer
- Hosting facilities (virtual machine, web site)

Services

- Access to public biologic banks
- Access to bioinformatics software
- Support, Training sessions

Our team (17 positions) represents 11.85 full-time equivalent.



Matthias Zytnicki (50% FTE)

DR INRAE / Scientific animation / Development and data analysis

+33 (0)5 61 28 54 93

matthias.zytnicki(at)inrae.fr



Claire Hoede (100% FTE)

IR INRAE / Technical animation / Development and analysis

+33 (0)5 61 28 53 05

claire.hoede(at)inrae.fr



Philippe Bordron (100% FTE)

IE INRAE / Development and data analysis

+33 (0)5 61 28 50 72

philippe.bordron(at)inrae.fr



Philippe Ruiz (100% FTE)

IE INRAE / Development and data analysis

philippe.ruiz(at)inrae.fr



Christine Gaspin (30% FTE)

DR INRAE / Development and data analysis

+33 (0)5 61 28 52 82

christine.gaspin(at)inrae.fr



Christophe Klopp (30% FTE)

IR INRAE / Development and data analysis

+33 (0)5 61 28 50 36

christophe.klopp(at)inrae.fr



Vincent Dominguez (60% FTE)

work-study student / Web development

vincent.dominguez(at)inrae.fr



Florent Blaise (40% FTE)

IE INRAE / System administrator

+33 (0)5 61 28 54 21

florent.blaise(at)inrae.fr



Didier Laborie (100% FTE)

IE INRAE / System administrator

+33 (0)5 61 28 54 27

didier.laborie(at)inrae.fr



Marie-Stéphane Trotard (80% FTE)

IE INRAE / System administrator

+33 (0)5 61 28 52 76

marie-stephane.trotard(at)inrae.fr



Martin Racoupeau (100% FTE)

CDD PEPR BRéIF / Development and data analysis

martin.racoupeau(at)inrae.fr



Fabien Graziani (100% FTE)

IE INRAE / Development and data analysis

fabien.graziani(at)inrae.fr



Nathalie Vialaneix (25% FTE)

DR INRAE / Statistical expertise and data analysis

+33 (0)5 61 28 55 73

nathalie.vialaneix(at)inrae.fr



Patrice Dehais (50% FTE)

IE INRAE / System administrator

+33 (0)5 61 28 57 08

patrice.dehais(at)inrae.fr



Théo Moser (100% FTE)

CDD Inspect and HOSP_ATB projects / Development and data analysis

theo.moser(at)inrae.fr



Mohamed Reda Derrassi (20% FTE)

CDD Mudis4LS / DevOps

mohamed-reda.derrassi(at)inrae.fr



Nicolas Enjalbert-Courrech (100% FTE)

IR INRAE / Statistical expertise and data analysis

nicolas.enjalbertcourrech(at)inrae.fr

The users

1200+ registered users :

- 50% from INRAe (national)
- 50% from other academic laboratories (regional)
- Biologists or computer scientists working in bio-informatics

Equipments

- **Computing cluster (HPC)**
about 5000 cores, 80TB RAM
4,5 PB disk space available for computing
- **Permanent Storage File System**
2*1,6 PB disk space capacities (asynchronous replication)
- **Virtual machines (VM)**
capacities for servers hosting and web services

Disk spaces

/usr/local/bioinfo/src

Bioinformatics software

/bank

Genomic reference databanks

/home/username

User **configuration files ONLY**
(10 GB user quota)

/work/user/username

Temporary computational disk space
(1 TB user quota)

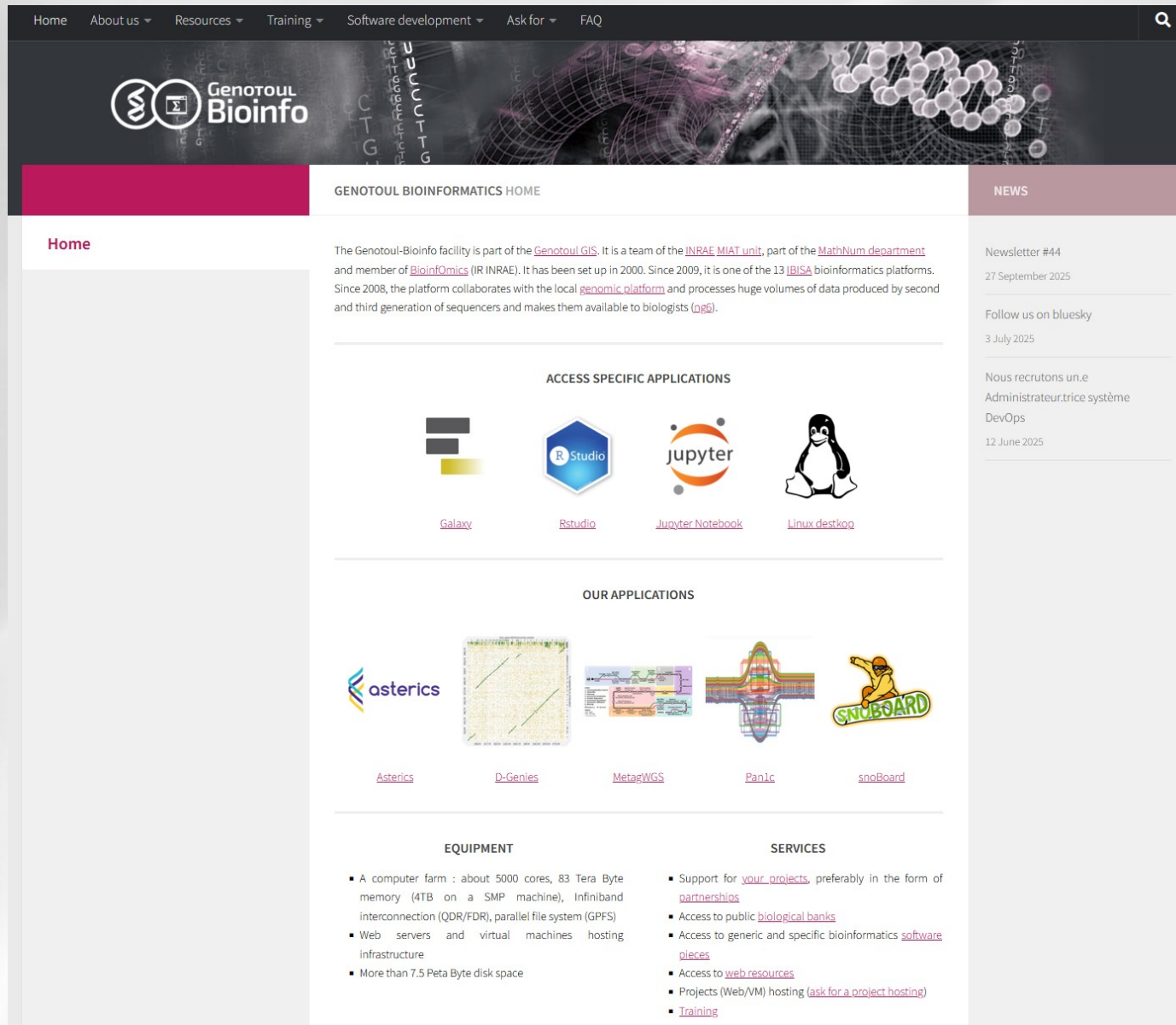
/save/user/username

Permanent disk space (with **BACKUP**)
(250 GB user quota + replication)

/save/project/XXX
/work/project/XXX

Billed project disk space
(**work** and / or **save**)

<https://bioinfo.genotoul.fr>



The screenshot shows the Genotoul Bioinformatics Home website. The navigation menu includes Home, About us, Resources, Training, Software development, Ask for, and FAQ. The main content area is titled 'GENOTOUL BIOINFORMATICS HOME' and features a 'Home' section with a paragraph describing the facility. Below this are sections for 'ACCESS SPECIFIC APPLICATIONS' (Galaxy, Rstudio, Jupyter Notebook, Linux desktop) and 'OUR APPLICATIONS' (Asterics, D-Genies, MetagWGS, Panic, snoBoard). At the bottom, there are sections for 'EQUIPMENT' and 'SERVICES' with bulleted lists of details.

Home

The Genotoul-Bioinfo facility is part of the [Genotoul GIS](#). It is a team of the [INRAE MIAT unit](#), part of the [MathNum department](#) and member of [BioinfoMics](#) (IR INRAE). It has been set up in 2000. Since 2009, it is one of the 13 [IBISA](#) bioinformatics platforms. Since 2008, the platform collaborates with the local [genomic platform](#) and processes huge volumes of data produced by second and third generation of sequencers and makes them available to biologists ([ngg](#)).

ACCESS SPECIFIC APPLICATIONS

- [Galaxy](#)
- [Rstudio](#)
- [Jupyter Notebook](#)
- [Linux desktop](#)

OUR APPLICATIONS

- [Asterics](#)
- [D-Genies](#)
- [MetagWGS](#)
- [Panic](#)
- [snoBoard](#)

EQUIPMENT

- A computer farm : about 5000 cores, 83 Tera Byte memory (4TB on a SMP machine), Infiniband interconnection (QDR/FDR), parallel file system (GPFS)
- Web servers and virtual machines hosting infrastructure
- More than 7.5 Peta Byte disk space

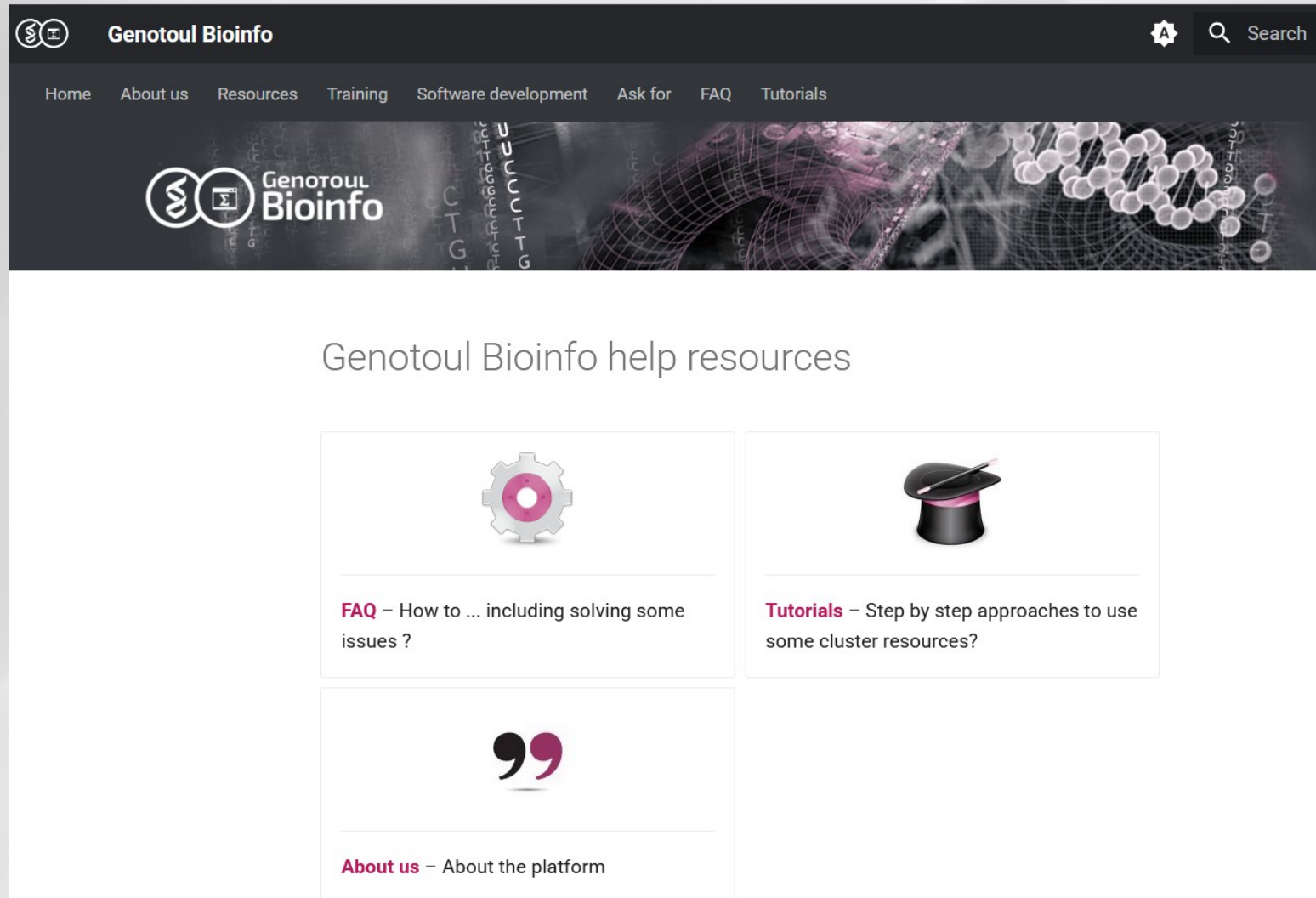
SERVICES

- Support for [your projects](#), preferably in the form of [partnerships](#)
- Access to public [biological banks](#)
- Access to generic and specific bioinformatics [software pieces](#)
- Access to [web resources](#)
- Projects (Web/VM) hosting ([ask for a project hosting](#))
- [Training](#)

NEWS

- Newsletter #44
27 September 2025
- Follow us on bluesky
3 July 2025
- Nous recrutons un.e
Administrateur.trice système
DevOps
12 June 2025

<https://bioinfo.genotoul.fr/index.php/faq/>



The screenshot shows the Genotoul Bioinfo website interface. At the top, there is a navigation bar with the site name and a search function. Below the navigation bar is a banner image with the Genotoul Bioinfo logo and a molecular structure. The main content area is titled "Genotoul Bioinfo help resources" and contains three cards: "FAQ - How to ... including solving some issues?", "Tutorials - Step by step approaches to use some cluster resources?", and "About us - About the platform".

Genotoul Bioinfo

Home About us Resources Training Software development Ask for FAQ Tutorials

Genotoul Bioinfo

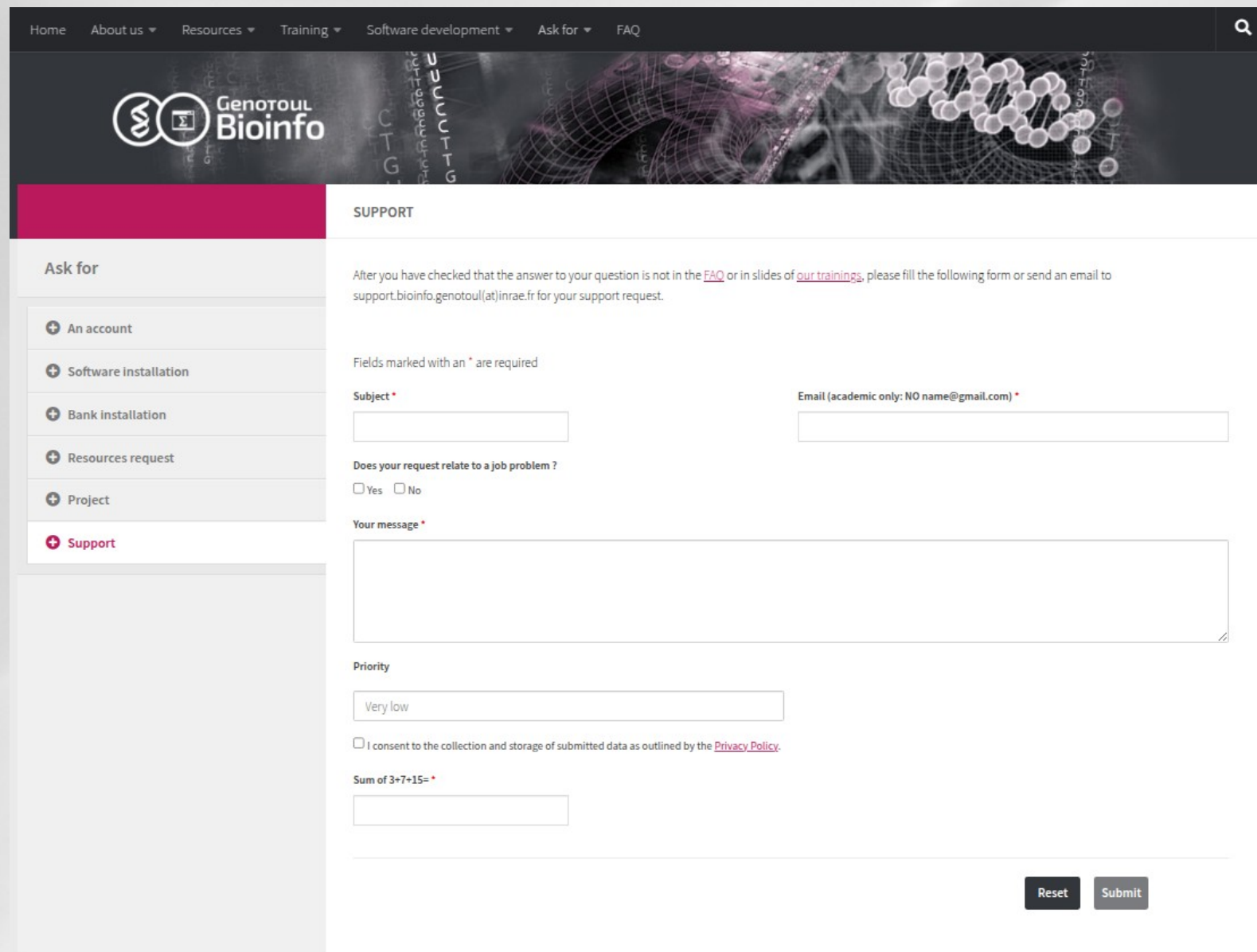
Genotoul Bioinfo help resources

FAQ – How to ... including solving some issues ?

Tutorials – Step by step approaches to use some cluster resources?

About us – About the platform

mailto: support.bioinfo.genotoul@inrae.fr



The screenshot shows the 'SUPPORT' page of the Genotoul Bioinfo website. The page has a dark header with navigation links: Home, About us, Resources, Training, Software development, Ask for, and FAQ. A search icon is in the top right. Below the header is a banner image with the Genotoul Bioinfo logo and a background of DNA and data visualizations.

The main content area is titled 'SUPPORT' and contains the following text: 'After you have checked that the answer to your question is not in the [FAQ](#) or in slides of [our trainings](#), please fill the following form or send an email to support.bioinfo.genotoul(at)inrae.fr for your support request.'

Below this text is a form with the following fields and options:

- Subject ***: A text input field.
- Email (academic only: NO name@gmail.com) ***: A text input field.
- Does your request relate to a job problem ?**: Radio buttons for 'Yes' and 'No'.
- Your message ***: A large text area for the support request.
- Priority**: A dropdown menu currently set to 'Very low'.
- I consent to the collection and storage of submitted data as outlined by the [Privacy Policy](#).
- Sum of 3+7+15= ***: A text input field for a CAPTCHA.

At the bottom right of the form are two buttons: 'Reset' and 'Submit'.

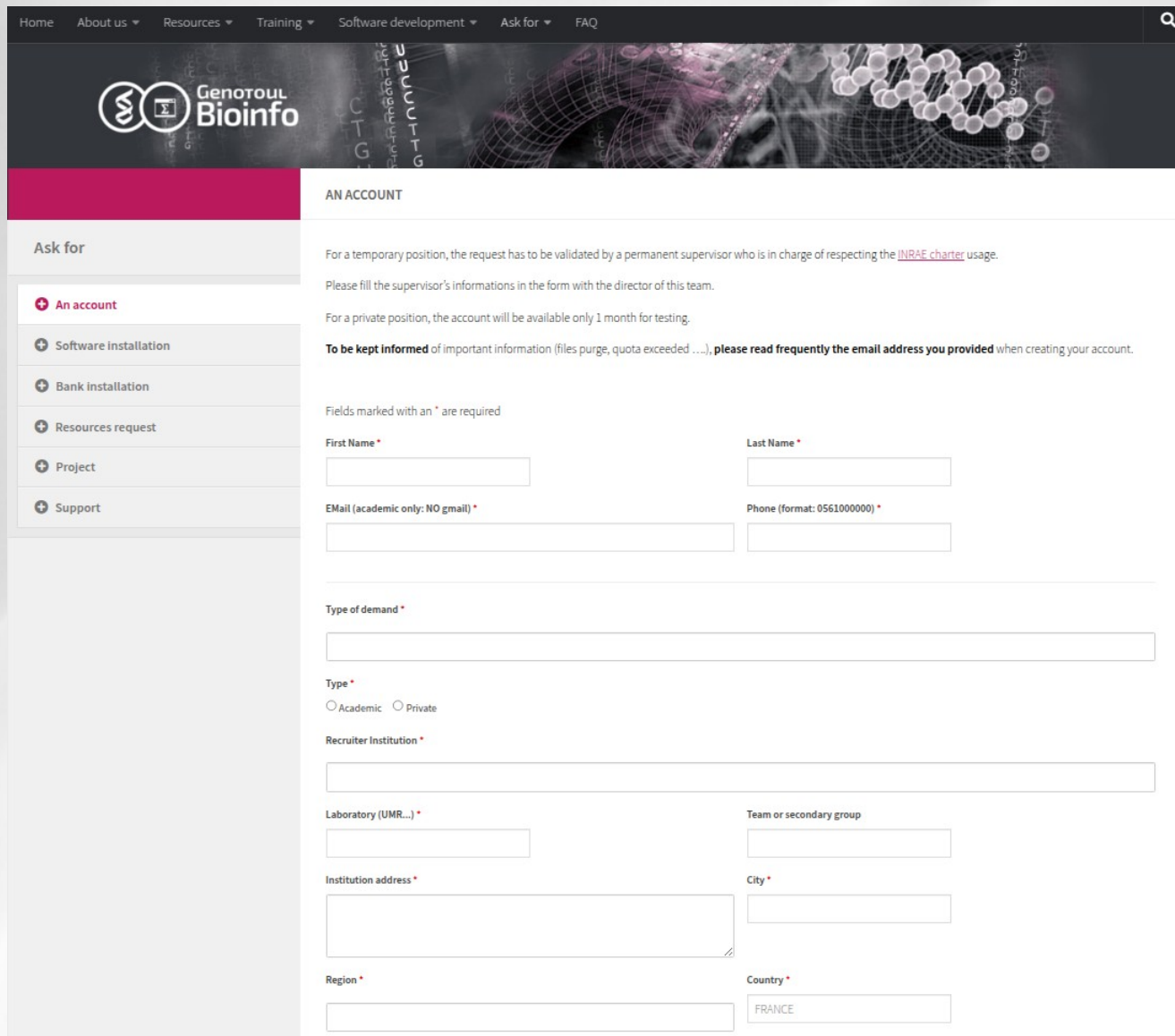
Introduction to Linux

GNU-Linux : Unix-like operating system

- Initial Developer = Linus Torvalds (Helsinki)
- Birth of kernel Linux on 1991
- GNU project = free and open source software
- Hundreds of active distributions : Ubuntu, Fedora, Debian...
- Large community of developers
- Some are commercial : RedHat, SUSE...

Asking for an account

<https://bioinfo.genotoul.fr/index.php/ask-for/create-an-account>



Home About us Resources Training Software development Ask for FAQ

Genotoul Bioinfo

AN ACCOUNT

For a temporary position, the request has to be validated by a permanent supervisor who is in charge of respecting the [INRAE charter](#) usage.
Please fill the supervisor's informations in the form with the director of this team.

For a private position, the account will be available only 1 month for testing.

To be kept informed of important information (files purge, quota exceeded ...), **please read frequently the email address you provided** when creating your account.

Fields marked with an * are required

First Name *

Last Name *

E-Mail (academic only: NO gmail) *

Phone (format: 0561000000) *

Type of demand *

Type * Academic Private

Recruiter Institution *

Laboratory (UMR...) *

Team or secondary group

Institution address *

City *

Region *

Country *

Linux account

Access to a “work” environment

- Login + password (login nodes)
- Share resources (cpu, memory, disk)
- Usage of software installed
- Free access to computational cluster
- Own space disk (**/save** & **/work** directory)

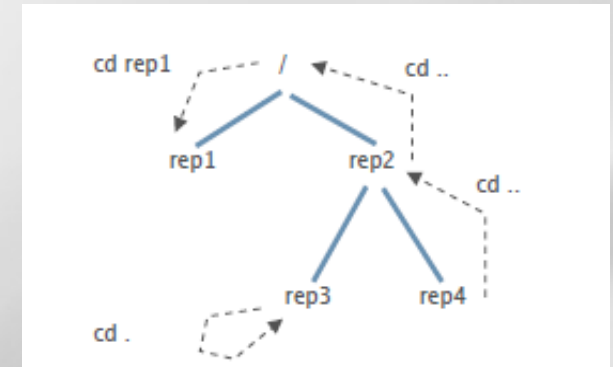
- Correct social behaviour is expected

The tree structure

Navigation

Tree structure

- « / » root directory
- « ~ » home directory (user)
- « . » current directory
- « .. » parent directory



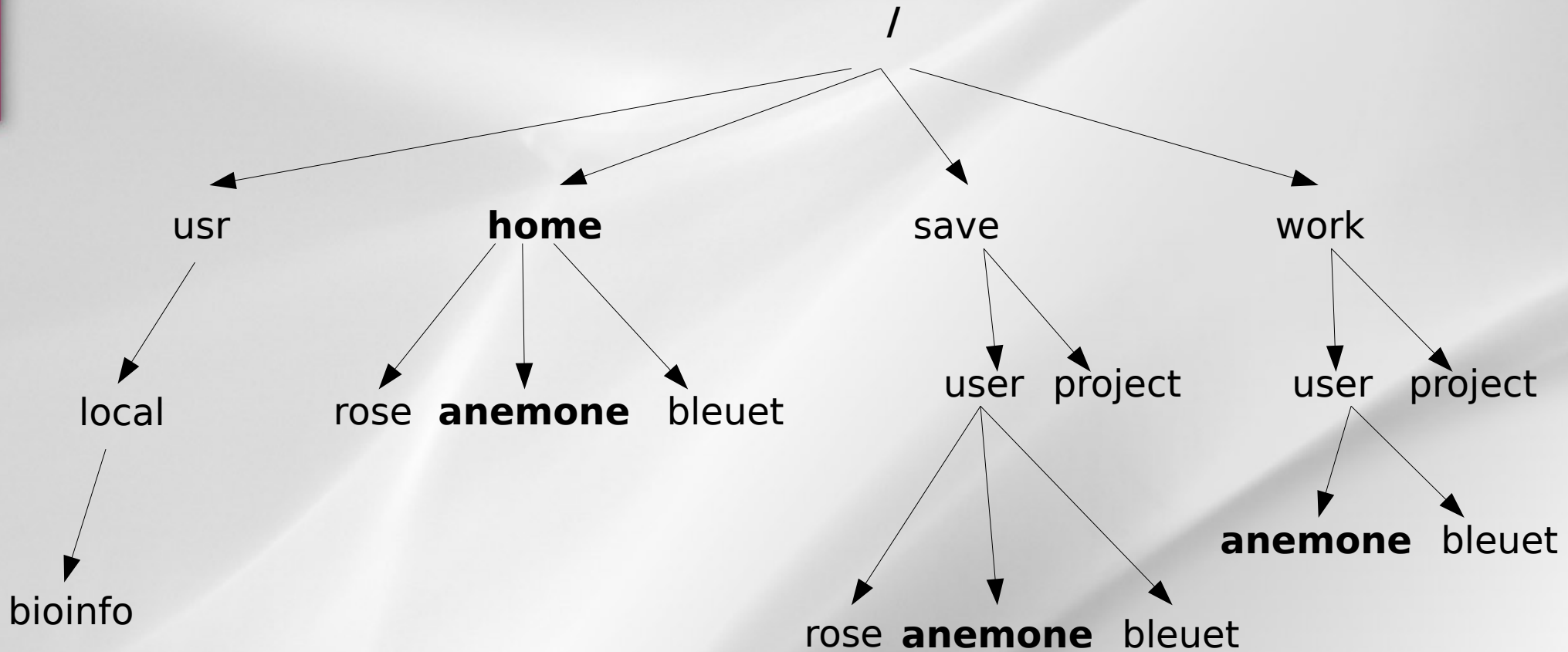
cd [nom_répertoire] : Change directory

Absolute path : /home/bleuet

Relative path : ../../usr

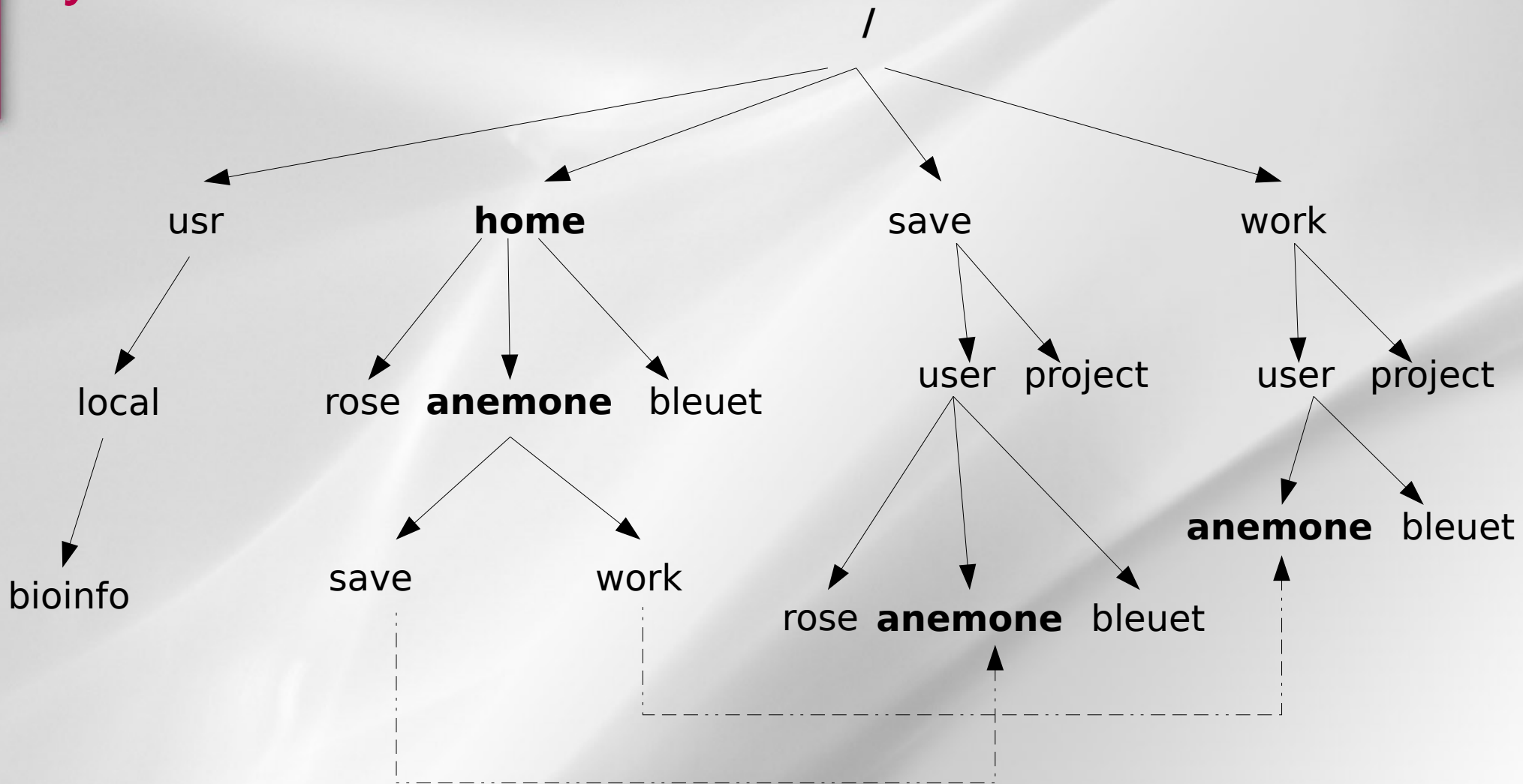
The tree structure

Notion of « Root »



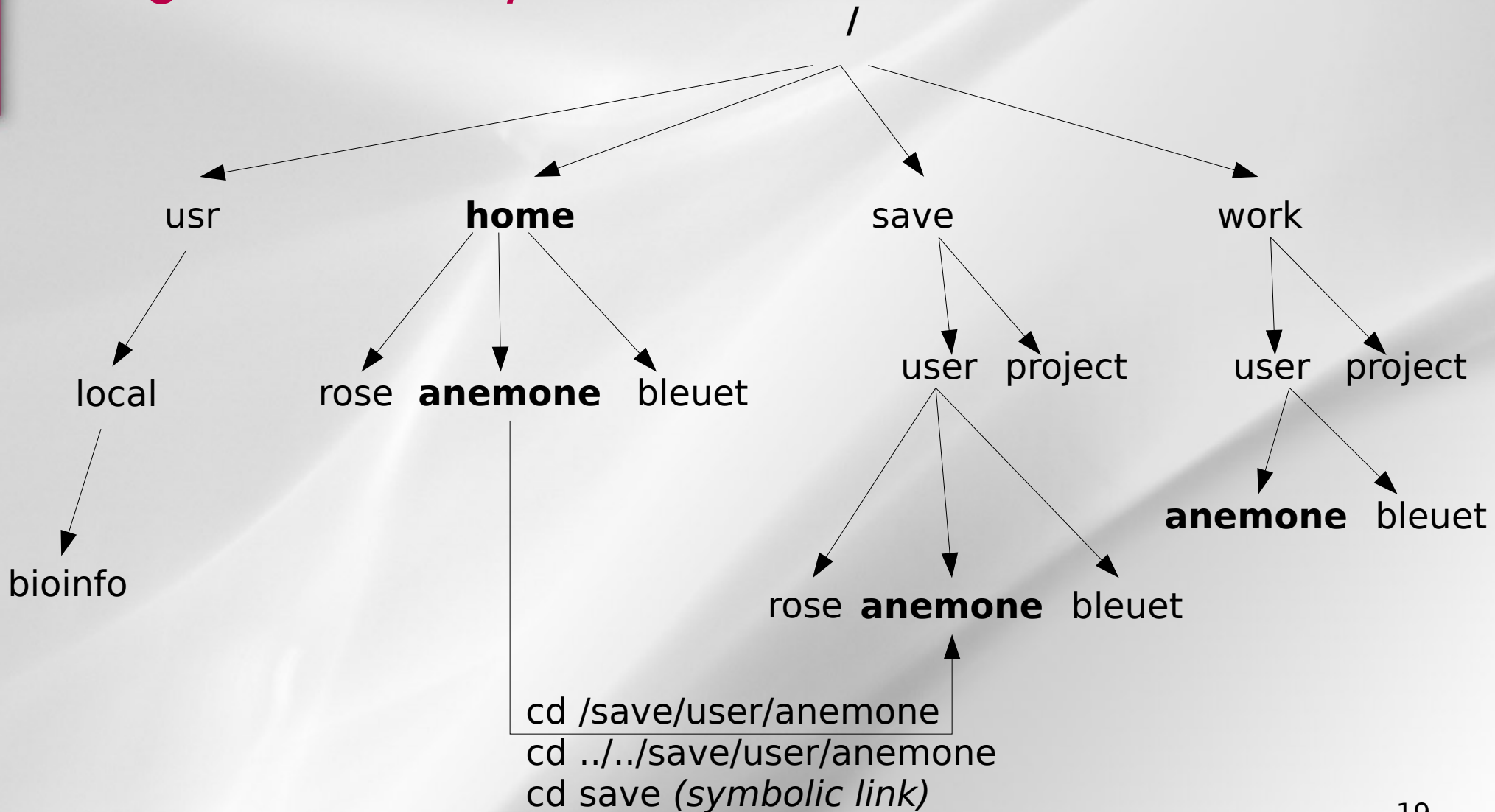
The tree structure

Symbolic link



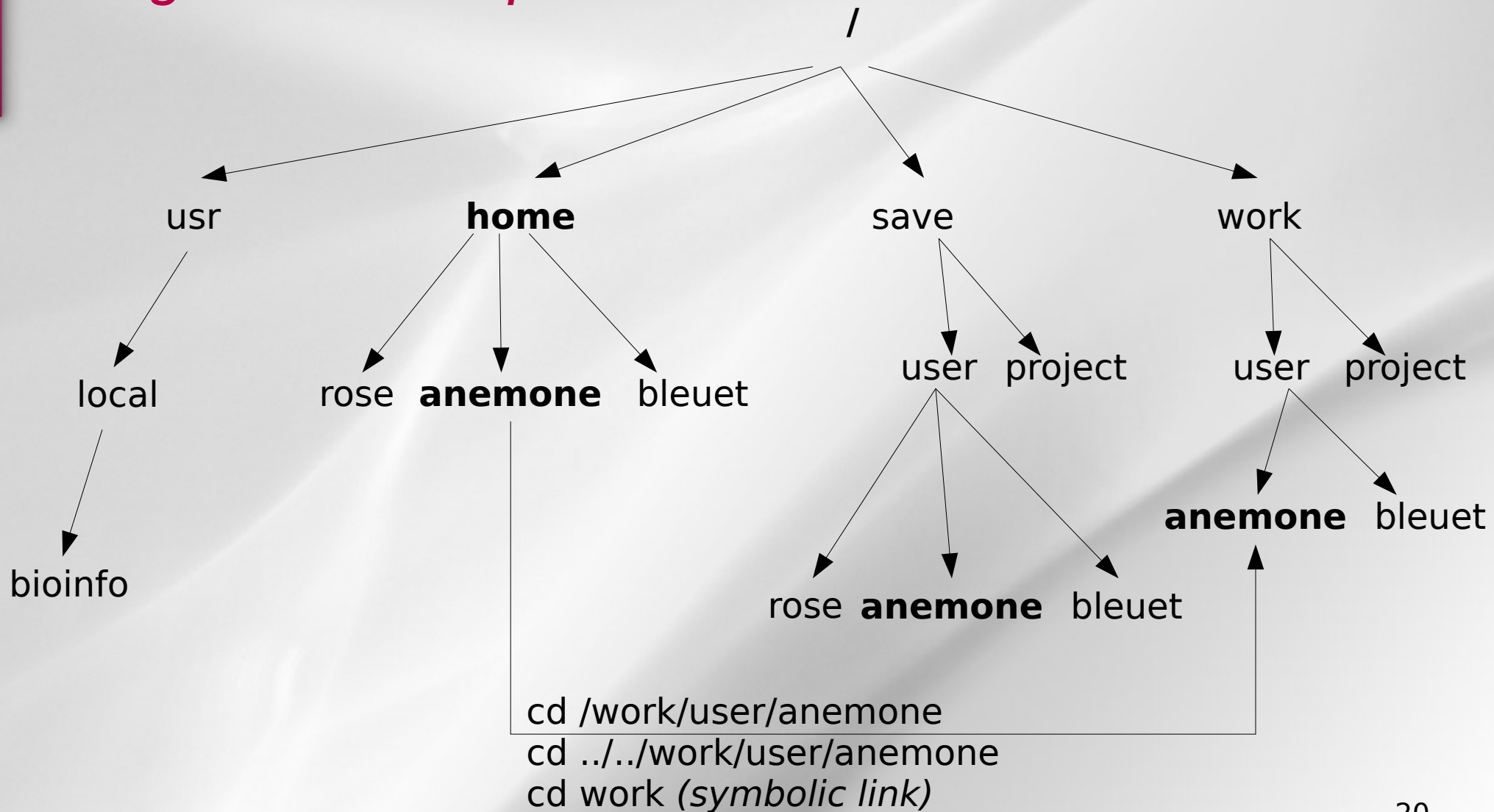
The tree structure

Navigation: exemple



The tree structure

Navigation: exemple



The command line : syntax

command_name [-option] [parameter]

- Command_name : what you want to do ?
- Option : how to do it ?
- Parameter : on which ?

```
ls
ls -l /home
ls -a
ls -l -a
ls -a -l
ls -la
ls -al
```

} same thing

The help on command

command_name [--help / - help / -h]

man command_name

```
ls --help  
blastn --help  
  
man ls  
man cd
```

Some basics commands

- Examples

cd : change directory

pwd : print working directory

ls [nom_répertoire]: list directory contents

tree : list contents in a tree like format

who : show who is logged on the server

passwd : update user's authentication token

history : display the commands history

How to connect to genobioinfo ?

From Windows

- **MobaXterm**



MobaXterm Home Edition v23.2
(Portable edition)

- **Other tools : wsl / openssh**

From Linux / Mac

- **Open command line tool (terminal)**

```
ssh -X username@genobioinfo.toulouse.inrae.fr
```



How to connect to genobioinfo ?

<https://geno-ood.toulouse.inrae.fr/>

The screenshot displays a web browser window with the URL `geno-ood-1.toulouse.inrae.fr/pun/sys/dashboard/batch_connect/sessions`. The page header includes navigation links: `GENOBIOINFO HPC Portal`, `Files`, `Jobs`, `Shells`, `Interactive Apps`, and `My Interactive Sessions`. A dropdown menu for `Shells` is open, showing `>_genobioinfo Shell Access`. An inset window titled `Shell - GENOBIOINFO HF` shows the terminal interface with the URL `geno-ood-1.toulouse.inrae.fr/pun/sys/shell/ssh/genobioinfo.toulouse.inrae.fr`. The terminal output shows the host `genobioinfo.toulouse.inrae.fr` and the prompt `laborie@genobioinfo.toulouse.inrae.fr's password:` with a red cursor.

Tricks & Tips

- **Copy / Paste with the mouse**
 - Select a text (it is automatically copied)
 - Click on the mouse wheel (the text is pasted where the cursor is located)
- **Stop a process :**
 - Use the CTRL <C> keys
- **Command and path completion :**
 - Use the TAB key (tab tab)
- **Back to the previous used commands :**
 - Use the « up » and « down » keys

- Connect yourself to genobioinfo server with your (training) login/password
- Do the exercices (TP1)
<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/linux-initiation/tp1/>

Plan

- File types
- Manipulating files
- TP 2.1
- File permissions
- wild card characters
- TP 2.2

File types

The « ls » command

List the content of a directory

ls [-options] [dir_name]

- a** : display hidden files/dir
- l** : use the long format
- t** : sort the content
- r** : reverse the sort order
- h** : human readable

```
ls -lh
-rw-r--r-- 1 laborie BIOINFO 69 22 oct. 2015 SAMPLE.nhr
-rw-r--r-- 1 laborie BIOINFO 144 22 oct. 2015 SAMPLE.nin
-rw-r--r-- 1 laborie BIOINFO 106K 22 oct. 2015 SAMPLE.nsq
```

File types

"ls -l" command (long listing format)

```
#ls -l
```

```
-rwxr-xr-x 1 cnoirot BIOINFO      123 Jun 14 17:16 blastforeach.sh
-rw-r--r-- 1 cnoirot BIOINFO 3683591 Jun  9 11:56 Diapo_F10a.odp
drwxr-xr-x 3 cnoirot BIOINFO      4096 Jul  8 14:56 igv
-rwxr-xr-x 1 cnoirot BIOINFO        20 Apr 16 11:21 monscript.sh
-rw-r--r-- 1 cnoirot BIOINFO 954415 Oct  3  2009 Presentation_pyrocleaner.odp
lrwxrwxrwx 1 cnoirot BIOINFO      13 Mar 15  2009 save -> /save/user/cnoirot
lrwxrwxrwx 1 cnoirot BIOINFO       3 Mar 18  2009 work -> /work/user/cnoirot
```

Permissions - Nb elements - Owner - Group - Size - Date - Name

Manipulating files

File/Dir. Creating and removing

mkdir / rmdir [dir_name] : create/remove an empty directory

```
mkdir dir_name
```

touch / rm [file_name] : create/remove a file

```
touch file_name
```

Manipulating files

Copying files/dir

cp src_filename dest_filename

=> copy source file to destination file

```
cp file1 file2
```

cp -r src_dirname dest_dirname

=> copy source dir. to destination dir.

```
cp -r dir1 dir2
```

Manipulating files

Moving / renaming a file

mv source destination

→ **Move** :

```
mv file_name existing_dir_name
```

→ **Rename** :

```
mv old_file_name new_file_name
```

→ **Move and rename** :

```
mv old_file_name existing_dir_name/new_file_name
```

Manipulating files

Wild cards characters

? replace any (one) character

```
ls bov?.seq
```

* replace 0, 1 ou any character

```
ls *.seq
```

```
rm bacterie*
```

[] replace any character between a selection

```
ls [123]*
```

```
ls f[a-c]*
```

Symbolic link

A kind of shortcut

ln -s nom_fic_source nom_fic_destination
create a symbolic link

```
ln -s file_name link_name
```

```
ls -l link_name
```

```
link_name -> file_name
```

TP

- Do the exercises (TP 2.1)
<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/linux-initiation/tp2.1/>

File types

Read, write, execute

Type - User - Group - Others

```
#ls -l
```

```
-rwxr-xr-x 1 cnoiot BIOINFO 123 Jun 14 17:16 blastforeach.sh
-rw-r--r-- 1 cnoiot BIOINFO 3683591 Jun 9 11:56 Diapo_F10a.odp
drwxr-xr-x 3 cnoiot BIOINFO 4096 Jul 8 14:56 igv
-rwxr-xr-x 1 cnoiot BIOINFO 20 Apr 16 11:21 monscript.sh
-rw-r--r-- 1 cnoiot BIOINFO 954415 Oct 3 2009 Presentation_pyrocleaner.odp
lrwxrwxrwx 1 cnoiot BIOINFO 13 Mar 15 2009 save -> /save/user/cnoiot
lrwxrwxrwx 1 cnoiot BIOINFO 13 Mar 18 2009 work -> /work/user/cnoiot
```

File permission modification

chmod [options] filename

modifies the permissions of a file

→ **u** :user, **g** : group, **o** : other, **a** : all

→ **r** : read, **w** : write, **x** : execute

```
chmod g+w file_name
```

Manipulating files

Finding files/dir

find dirname [-option] [parameter]

```
find /home/formation -name "*.seq"
```

```
find . -type d
```

```
find . -type f
```

```
find / -size +1000k
```

Kind of files

file command

file file_name

file find_result.txt:	ASCII text
file create_account	symbolic link
file extract_users_stats.py	Python script, ASCII text executable

TP

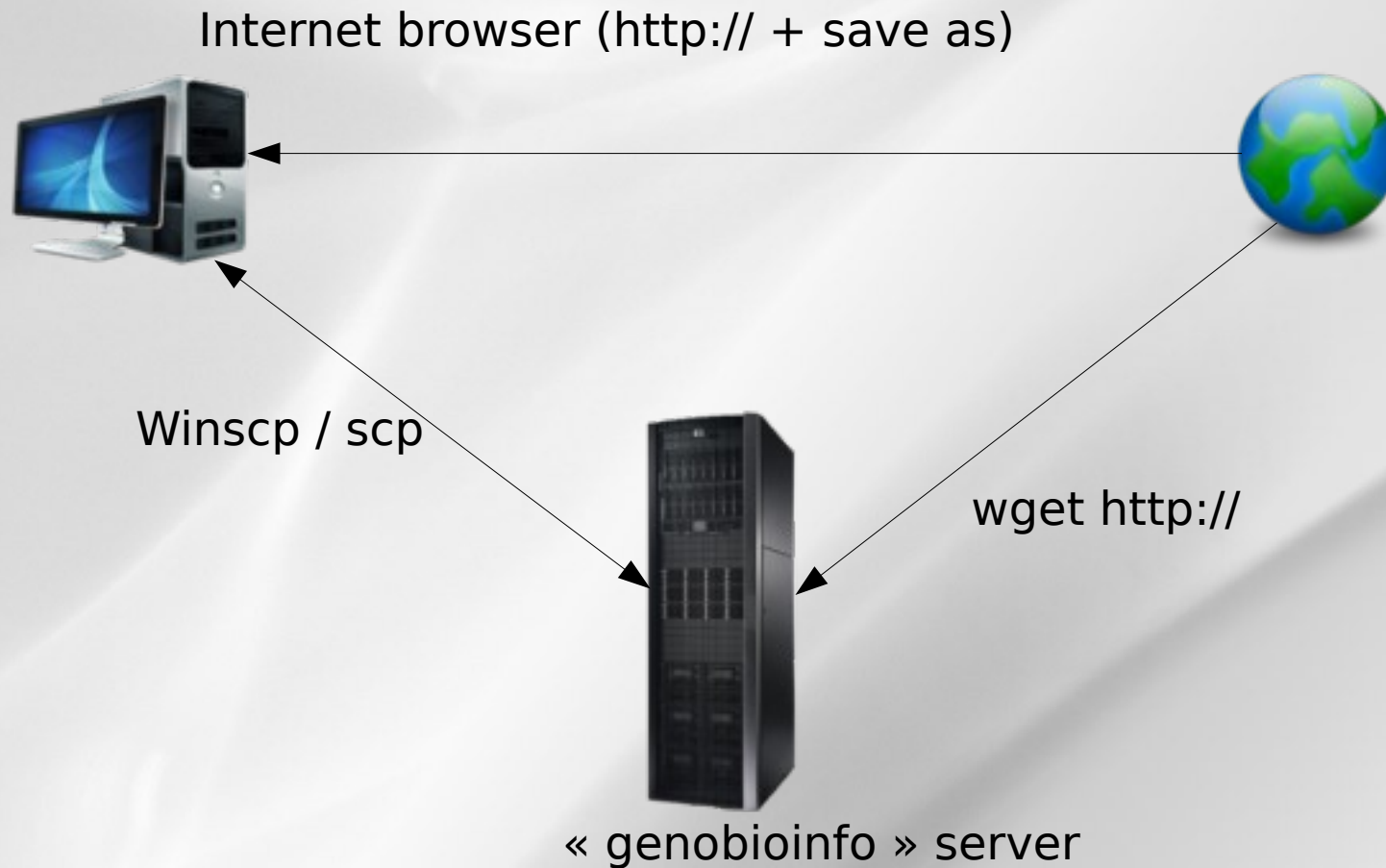
- Do the exercises (TP 2.2)
<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/linux-initiation/tp2.2/>

Plan

- Downloading / transferring
- Disk space control
- Compressing / uncompressing
- TP 3.1
- Displaying files
- Utility commands
- Data extraction
- Redirections
- TP 3.2
- My first script

Downloading / transferring

Several possible cases



Downloading / transferring

Directly from internet to genobioinfo

File download from Internet to « genobioinfo server »:

- Copy the URL of the file to download

```
wget http://url.a.telecharger/nom_fichier
```

Downloading / transferring

Transfer between genobioinfo and desktop computer

We recommend to use « scp » command (secure copy)

scp [user@host1:]file1 [user@host2:]file2
copy file from the network

scp source_name **bleuet@genobioinfo.toulouse.inrae.fr:/dest_name**
(copy from desktop to "genobioinfo server")

Downloading / transferring

FileZilla : copy via graphical interface

The screenshot shows the FileZilla interface with the following details:

- Window Title:** sftp://laborie@genobioinfo.toulouse.inrae.fr - FileZilla
- Host:** sftp://genobioinf
- Identifiant:** laborie
- Mot de passe:** [masked]
- Port:** 22
- Connexion rapide:** [dropdown menu]
- Status Log:**
 - Erreur : Impossible d'établir une connexion au serveur.
 - Statut : Connexion à genobioinfo.toulouse.inrae.fr...
 - Statut : Connected to genobioinfo.toulouse.glb-pra.inra.fr
 - Statut : Récupération du contenu du dossier...
 - Statut : Listing directory /home/laborie
 - Statut : Contenu du dossier "/home/laborie" affiché avec succès
- Local Site:** /home/dlaborie/
 - home
 - dlaborie
 - homer
 - laborie
 - lib
 - lib64
 - lost+found
 - media
 - mnt
- Remote Site:** /home/laborie
 - /? /
 - /? home
 - laborie
- Local File List:**

Nom de fichier	Taille de fic	Type de fichier	Dernière modif
..			
.GlobalProt...		Dossier	25/02/2025 08:...
.cache		Dossier	25/02/2025 10:...
.config		Dossier	25/02/2025 10:...
.cups		Dossier	26/03/2024 10:...
.docker		Dossier	13/10/2023 11:...
.fltk		Dossier	30/06/2023 09:...
.gnome		Dossier	17/02/2022 08:...
.gnupg		Dossier	17/02/2022 11:...
.gphoto		Dossier	05/11/2024 08:...

15 fichiers et 32 dossiers. Taille totale : 312,5 Ko
- Remote File List:**

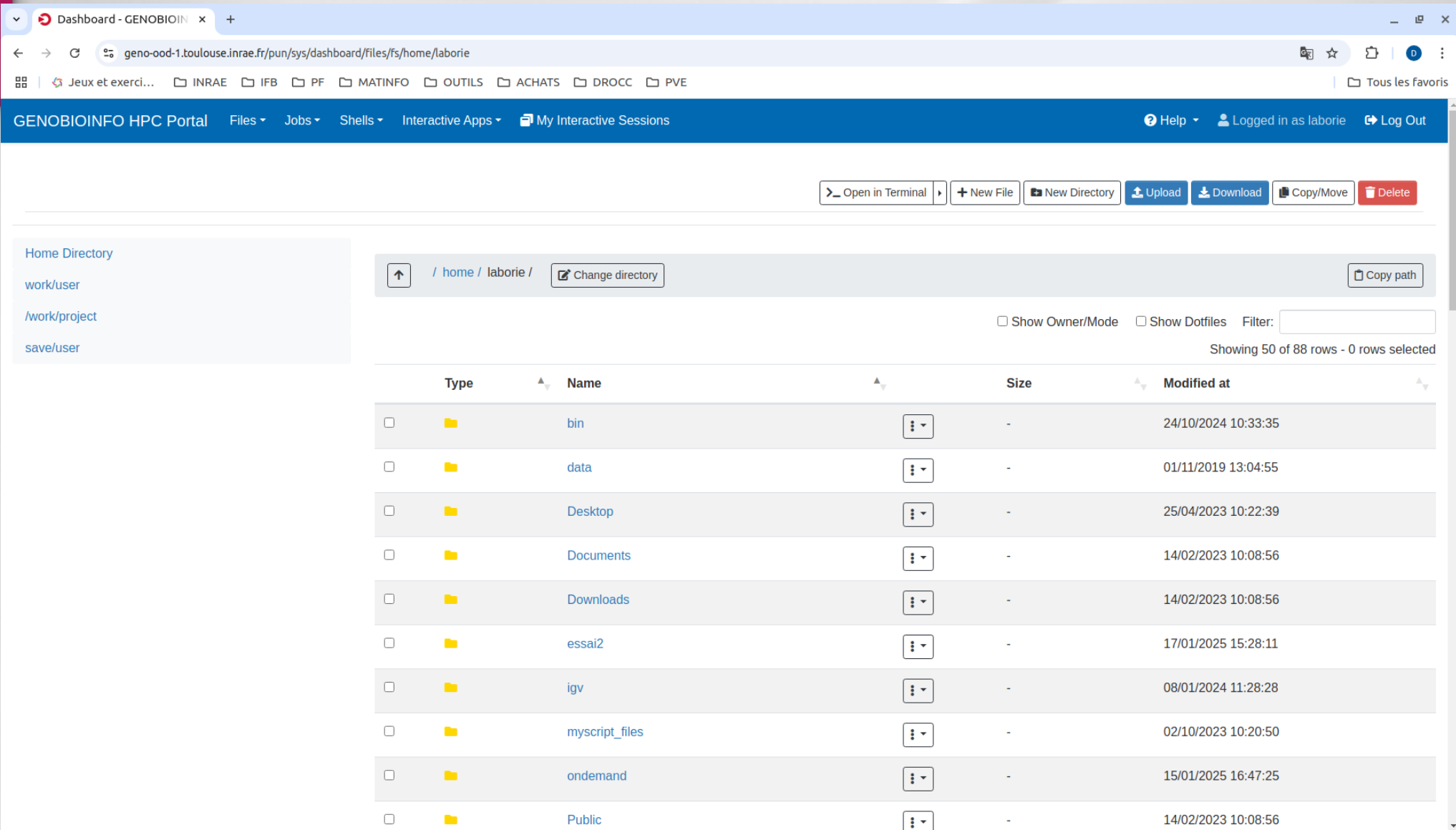
Nom de fic	Taille de fi	Type de ficl	Dernière mod	Droits d'ac	Propriétaire
..					
.aws		Dossier	06/06/2024 ...	drwxr-xr-x	laborie B...
.cache		Dossier	19/02/2025 ...	drwxr-xr-x	laborie B...
.cargo		Dossier	30/06/2023 ...	drwxr-xr-x	laborie B...
.conda		Dossier	17/02/2025 ...	drwxrws...	laborie B...
.config		Dossier	02/12/2024 ...	drwx---	laborie B...
.dbus		Dossier	14/02/2023 ...	drwx---	laborie B...
.ema...		Dossier	22/04/2024 ...	drwx---	laborie B...
.igv		Dossier	08/01/2024 ...	drwxr-xr-x	laborie B...
.ipyn...		Dossier	15/02/2023 ...	drwxr-xr-x	laborie B...

46 fichiers et 42 dossiers. Taille totale : 1,2 Go
- Transfer Queue:**

Serveur / Fichier local	Directio	Fichier distant	Taille	Priorité	Statut
- Bottom Panel:**
 - Fichiers en file d'attente
 - Transferts échoués
 - Transferts réussis

Downloading / transferring

OpenOnDemand : Upload / Download



Dashboard - GENOBIOIN x +

geno-ood-1.toulouse.inrae.fr/pun/sys/dashboard/files/fs/home/laborie

Jeux et exerci... INRAE IFB PF MATINFO OUTILS ACHATS DROCC PVE

Tous les favoris

GENOBIOINFO HPC Portal Files Jobs Shells Interactive Apps My Interactive Sessions

Help Logged in as laborie Log Out

Open in Terminal New File New Directory Upload Download Copy/Move Delete

Home Directory

- work/user
- /work/project
- save/user

/ home / laborie / Change directory Copy path

Show Owner/Mode Show Dotfiles Filter:

Showing 50 of 88 rows - 0 rows selected

Type	Name	Size	Modified at
Folder	bin	-	24/10/2024 10:33:35
Folder	data	-	01/11/2019 13:04:55
Folder	Desktop	-	25/04/2023 10:22:39
Folder	Documents	-	14/02/2023 10:08:56
Folder	Downloads	-	14/02/2023 10:08:56
Folder	essai2	-	17/01/2025 15:28:11
Folder	igv	-	08/01/2024 11:28:28
Folder	myscript_files	-	02/10/2023 10:20:50
Folder	ondemand	-	15/01/2025 16:47:25
Folder	Public	-	14/02/2023 10:08:56

Disk spaces available

df [-option] [partition_name] :

Show the differences disk spaces

```
df -h
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda5	204G	8.7G	185G	5%	/
tmpfs	63G	16K	63G	1%	/dev/shm
/dev/sda1	124M	35M	84M	30%	/boot
/dev/sda3	9.9G	559M	8.8G	6%	/var
isi-ceri:/ifs/save	60T	47T	14T	78%	/save
isi-ceri:/ifs/home	100G	47G	54G	47%	/home

du [-option] [dir_name] :
Show the disk usage

```
du -csh --apparent-size /save/user/<username>
```

<https://vm-genoword.toulouse.inrae.fr/FAQ/faq/default-resources/#how-to-get-my-disk-quota-and-usage>

Compressing / uncompressing

Several formats

gzip : compress a file to **.gz**

```
gzip file_to_compress  
=>gz file creation
```

gunzip : uncompress a file **.gz**

```
gunzip file_to_uncompress.gz
```

Other formats : bz2, zip, rar, Z, 7z

Archiving

Tar command

tar -cvf : archive a file tree

```
tar -cvf formation.tar /home/formation  
=> .tar file creation
```

tar -xvf : deploy a file tree

```
tar -xvf formation.tar /tmp
```

Tips: combination of tar + gzip (.tgz)

tar -cvzf : archive + compression

tar -xvzf : uncompress-ion + deploy

TP3

- Do the exercises (TP 3.1)
<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/linux-initiation/tp3.1/>

Looking into files

Display a file content

cat file_name : display the file content

```
cat /bank/ncbi/genbank/genbankRelease/current/fasta/gbphg1.fasta
```

```
>AB000833.1 Bacteriophage Mu DNA for ORF1, sheath protein gpL, ORF2,  
ORF3, complete cds.
```

```
ACGGTCAGACGTTTGGCCCGACCAACCGGGATGAGGCTGACGCAGGTCAGAAATCTT  
TGTGACGACAACCGTATCAATGCCGGTGTGG...
```

cat file_name1 file_name2 ... : display many files content

Pagers

Navigate into the file

more file_name : display more and more

less file_name : display up and down (more than **more**)

```
less my_file.txt
```

Text editors

Modify a file content

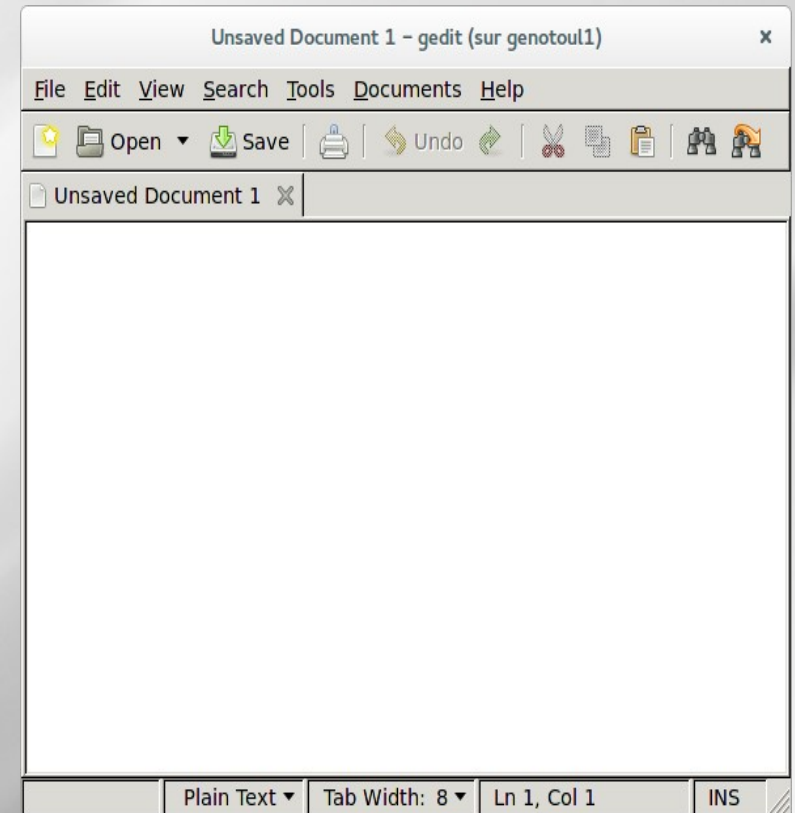
vi : standard but difficult

nano : easy to use

gedit : graphic mode, intuitive

nedit : idem as gedit

emacs : advanced features



Utility commands

sort [-options] file_name : sort a file

```
sort -n -k 1  
    (num. sort, first col.)
```

wc [-options] file_name : words count

```
wc -c file_name
```

```
wc -w file_name
```

```
wc -l file_name
```

Data extraction

Filters (1)

cat [-options] file (s) name : merge files

```
cat nom_fic1 nom_fic2
```

head [-number] file_name : read the beginning of a file

```
head -100 file_name (first 100 lines)
```

tail [-f] [+/-number] file_name : read the end of a file

```
tail -n 100 file_name (last 100 lines)
```

```
tail -n +6 file_name (from the 6th line)
```

Data extraction

Filters (2)

cut [-options] file_name :
cuts the fields (vertically)

```
cut -c 1 (gets the first char.)
```

```
cut -f 2,3 (gets the #2 and #3 fields)
```

split [-options] file_name :
cuts the fields (horizontally)

```
split -l 500 file_name.txt (default size 500 lines)
```

Data extraction

File Comparison

meld file_name1 file_name2
compare two files (line per line)

```
meld fic_1 fic_2
```

Data extraction

Tex research

grep [-options] 'motif' file_name[s]

- Text research tool in the file contents
- Wild card characters may be used

```
grep SEQRES fichier_pdb (simple research)
grep -i (case insensitive)
grep -c (counts the line amount)
grep -v (all the lines except)
```

Redirections

Standard input / output

Most commands use the standard input / output :

Standard input = the keyboard

Standard output = the console

Input / Output (I/O) may be redirected by using the following operators : "<", ">", "|", ">>"

Redirections

Redirection

command > output_file_name

redirects the standard output to a new file

```
grep -i Human uniprot.fasta > fic_result
```

command1 | command2

redirects the standard output to another software

```
grep -i Human uniprot.fasta | wc -l
```

command >> output_file_name

redirects the standard output to an existing file and appends it

```
grep -i bovin uniprot.fasta >> fic_result
```

TP3

- Do the exercises (TP 3.2)
<https://genotoul-bioinfo.pages-forge.inrae.fr/linux-cluster/linux-initiation/tp3.2/>

My first script

- A script = a succession of commands
- Put commands into a text file

```
edit prog &
```

- Give the execution right

```
chmod +x prog
```

- Execute the script

```
./prog
```

My first script

- Automation and plan
- Win of time (re-utilization)
- Templates : easy to find on the web
- Portable (running on all Unix-like systems)

Be careful to the syntax between different shell langages (csh,bash...)

My first script

- Run a « blast » for all the fasta files of the directory :

```
#!/bin/bash

## COMMENT : THIS IS THE INPUT VARIABLE
REPERTOIRE=$1

## COMMENT : LIST ALL OF FILES
LISTE=$(ls $REPERTOIRE)
echo $LISTE

## COMMENT : REPETITION
for FILE in $LISTE
do
    blastall -p blastn -i $REPERTOIRE/$FILE -d swissprot -o
$FILE.out
    echo "Blastall sur le fichier : $FILE: ok"
done
exit
```

End of presentation

<https://sondages.inrae.fr/index.php/84236?lang=fr>

Thanks for your attention !

