



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 facilities, asking for an account, connection procedure, tree structure of files, command line syntax, TP1

Part II : 11h00 - 12h30

File types, permissions, manipulating files, displaying files, wildcard characters, disk space control, TP2

Part III : 14h00 - 17h00

Dowloading/transferring, compressing/uncompressing, utility commands, redirections, TP3

Part I

- Presentation of GenoToul bioinformatics 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

The team



Matthias Zytnicki (50% FTE) CR INRAE / Scientific animation / Development and data analysis
+33 (0)5 61 28 54 93
matthias.zytnicki@inrae.fr



Claire Hoede (100% FTE) IR INRAE / Technical animation / Development and data analysis
+33 (0)5 61 28 53 05
claire.hoede@inrae.fr



Christine Gaspin (30% FTE) DR INRAE / Development and data analysis
+33 (0)5 61 28 52 82
christine.gaspin@inrae.fr



Christophe Klopp (30% FTE) IR INRAE / Development and data analysis
+33 (0)5 61 28 50 36
christophe.klopp@inrae.fr



Didier Laborie (100% FTE) IE INRAE / System administrator
+33 (0)5 61 28 54 27
didier.laborie@inrae.fr



Jérôme Mariette (70% FTE) IE INRAE / Development and data analysis
+33 (0)5 61 28 57 25
jerome.mariette@inrae.fr



Céline Noirot (80% FTE) IE INRAE / Development and data analysis
+33 (0)5 61 28 57 24
celine.noiro@inrae.fr



Marie-Stéphane Trotard (80% FTE) IE INRAE / System administrator
+33 (0)5 61 28 52 76
marie-stephane.trotard@inrae.fr



Nathalie Vialaneix (25% FTE) DR INRAE / Statistical expertise and data analysis
+33 (0)5 61 28 55 73
nathalie.vialaneix@inrae.fr



Patrice Dehais (50% FTE) IE INRAE / System administrator
+33 (0)5 61 28 57 08
patrice.dehais@inrae.fr



Philippe Bordron (100% FTE) IE INRAE / Development and data analysis
+33 (0)5 61 28 50 72
philippe.bordron@inrae.fr



Julien Henry (100% FTE) IR INRAE / Data omics integration and statistical data analysis
julien.henry@inrae.fr



Julien Touchais (100% FTE) CDD ANR MetRibo/ Web Development
julien.touchais@inrae.fr



Hanae Chouali (100% FTE) CDD ANR Symway / Development and data analysis
hanae.chouali@inrae.fr

The users

1000+ registered users :

- 85% from INRAe or REGIONAL laboratories
- Biologists or computer scientists working in bio-informatics

Equipments

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

Disk spaces

`/usr/local/bioinfo/src`

Bioinformatics software

`/bank`

Databanks

`/home`

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

`/work`

Temporary computational disk space
(1000 GB user quota)

`/save`

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

<https://bioinfo.genotoul.fr>

Home - genotoul-bioinfo - Mozilla Firefox


File Edit View History Bookmarks Tools Help

LDAP Account Manager X Home - genotoul-bioinfo X

bioinfo.genotoul.fr 90% libreoffice upgrade ubuntu

Most Visited eGroupWare [Login] Bioinformatics : Home Ganglia: Genoclust Gri... Login - OTRS 5s Didier Laborie - Outloo... Slurm Workload Manager https://genocloud.toul...

Home About us Resources Training Software development Ask for FAQ

 Genotoul Bioinfo

Home

GENOTOUL BIOINFORMATICS HOME

The GenoToul bioinformatics facility is part of the [Genotoul GIS](#). 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 ([ng6](#)).

EQUIPMENT






- A computer farm : about 5000 cores (INTEL-2014, AMD-2012), 34 Tera Byte memory (3TB on a SMP machine), Infiniband interconnection (QDR), parallel file system (GPFS)
- Web servers and virtual machines hosting infrastructure
- More than 1Peta Byte disk space

SERVICES

- 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](#)

Use this [link](#) to create your user account. All questions about technical issues can be sent using one of the [Ask for](#) forms. The platform staff can help you to process your data or to develop specific databases or software packages. For any specific request please send a mail to anim.bioinfo@toulouse.inra.fr.

ACCESS SPECIFIC APPLICATIONS

 Galaxy
  Jvarkit
  NCBI BLAST
  NG6
  EMBOSS

NEWS

Newsletter #30
20 July 2018

Newsletter #29
12 March 2018

Newsletter #28
15 November 2017

Genotoul Bioinfo

<https://bioinfo.genotoul.fr/index.php/faq/>
or [mailto: support.bioinfo.genotoul@inrae.fr](mailto:support.bioinfo.genotoul@inrae.fr)

FAQ - genotoul-bioinfo - Mozilla Firefox


File Edit View History Bookmarks Tools Help

LDAP Account Manager P X FAQ - genotoul-bioinfo X +

bioinfo.genotoul.fr/index.php/faq/ 90% libreoffice upgrade ubuntu

Most Visited eGroupWare [Login] Bioinformatics : Home Ganglia:: Genoclust Gri... Login - OTRS 5s Didier Laborie - Outloo... Slurm Workload Manager https://genocloud.toul...

Home About us Resources Training Software development Ask for FAQ


 Genotoul Bioinfo


FAQ FREQUENTLY ASKED QUESTIONS


Welcome on the Genotoul FAQ, here under are listed the most frequently asked questions and some application examples. Theses questions are categorized in different sections:


FAQ


- User access
- Training and self-training
- Linux
- Default resources
- Job submission
- Databanks
- Software
- Errors and Tricks
- NGS data
- Cite us


 User Access


 Training and self-training


 Linux

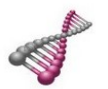
 Default Ressources


 Job Submission

 Databanks

 Software

 Errors and Tricks

 NGS data

 Cite us

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>

An account - genotoul-bioinfo - Mozilla Firefox


File Edit View History Bookmarks Tools Help

LDAP Account Manager P X An account - genotoul-bio X +

bioinfo.genotoul.fr/index.php/ask-for/create-an-account/ 90% libreoffice upgrade ubuntu

Most Visited eGroupWare [Login] Bioinformatics : Home Ganglia:: Genoclust Gri... Login - OTRS 5s Didier Laborie - Outloo... Slurm Workload Manager https://genocloud.toul...

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 [INRA charter](#) usage.

Please fill the supervisor's informations in the form with the director of this team.

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 *
<input type="text"/>	<input type="text"/>
EMail (academic only: NO gmail) *	Phone (format: 0561000000) *
<input type="text"/>	<input type="text"/>
Type *	
<input type="radio"/> Academic <input type="radio"/> Private	
Recruiter Institution *	
<input type="text"/>	
Laboratory (UMR...) *	Team or secondary group
<input type="text"/>	<input type="text"/>
Institution address *	City *
<input type="text"/>	<input type="text"/>

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)

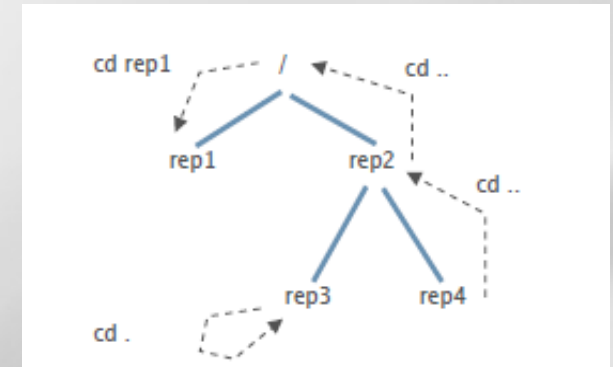
- Correct social behaviour is expected

The tree structure

Navigation

Tree structure

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



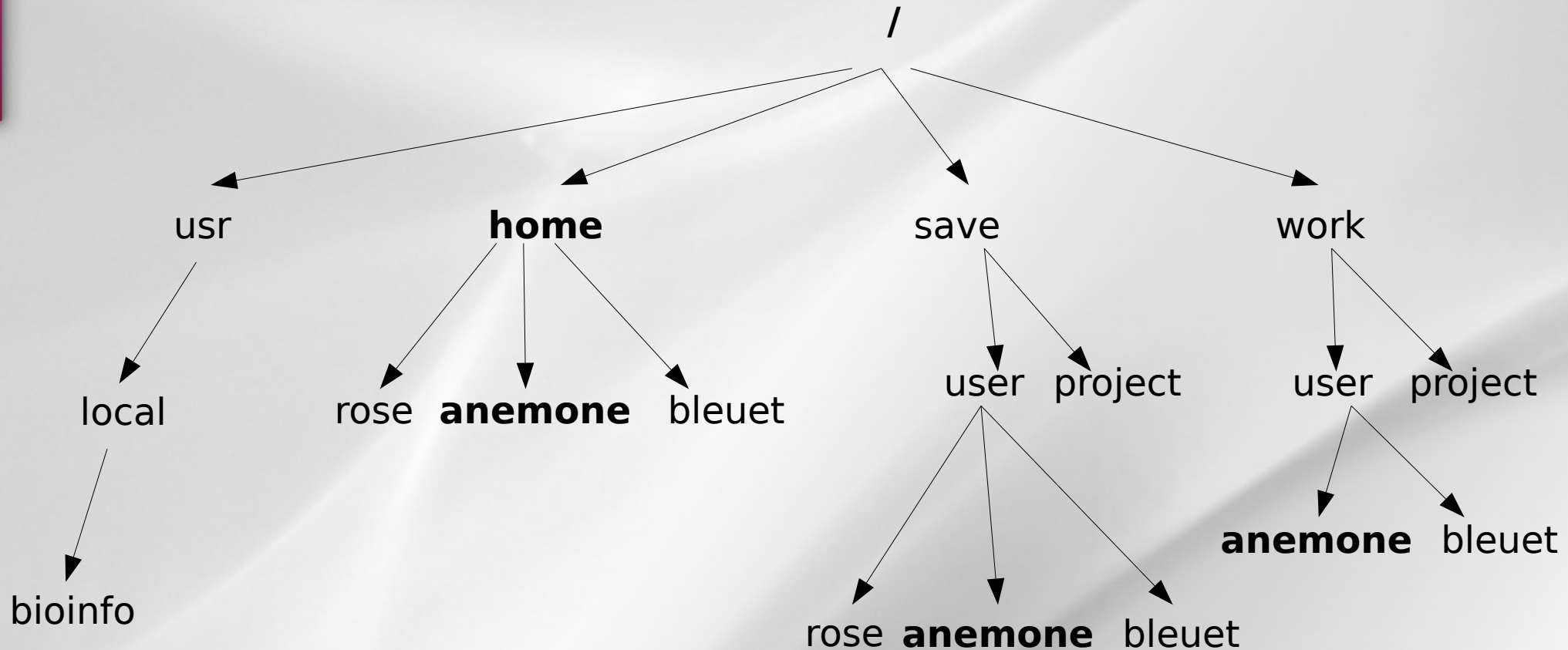
cd [nom_répertoire] : Change directory

Absolute path : /home/bleuet

Relative path : ../../usr

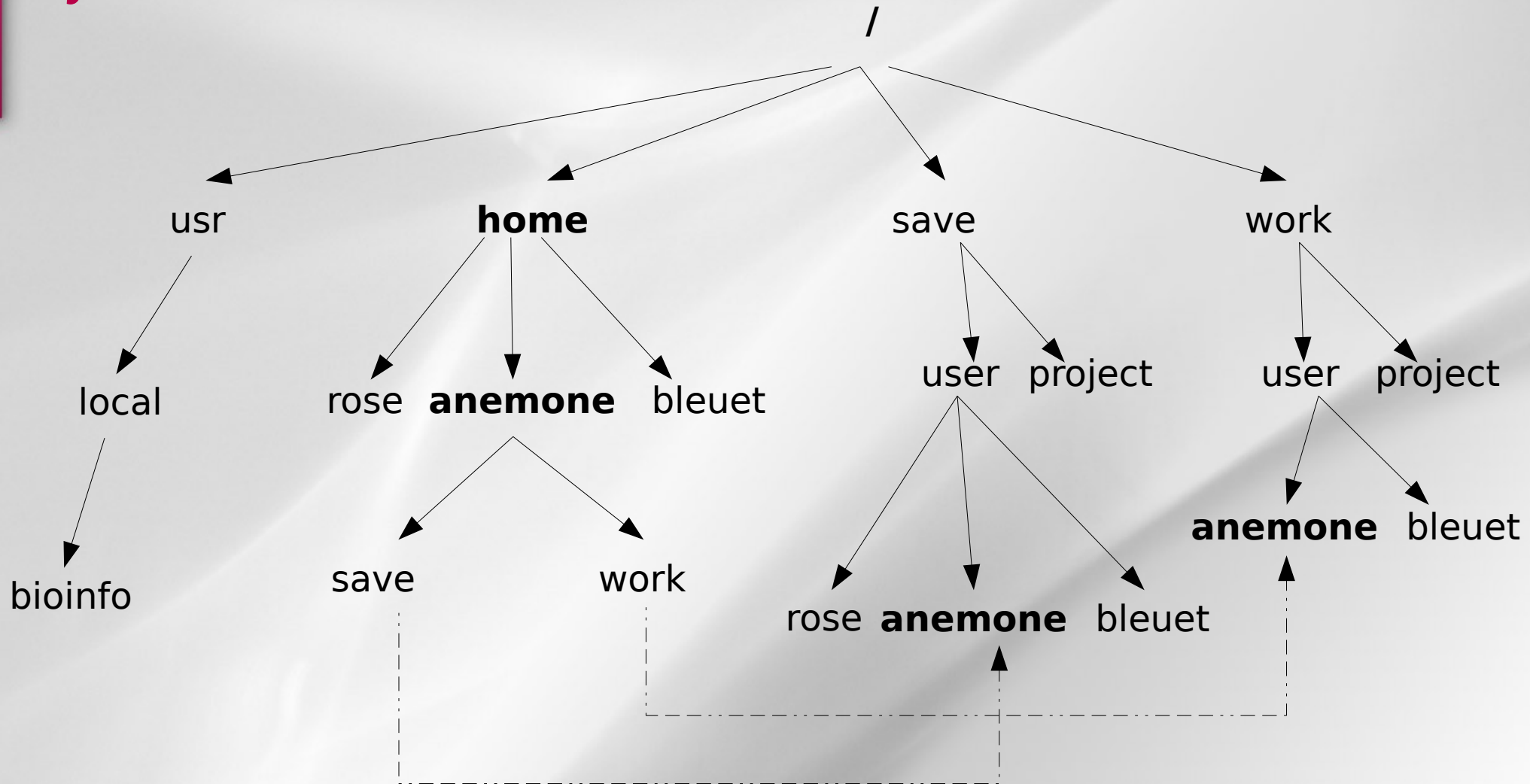
The tree structure

Notion of « Root »



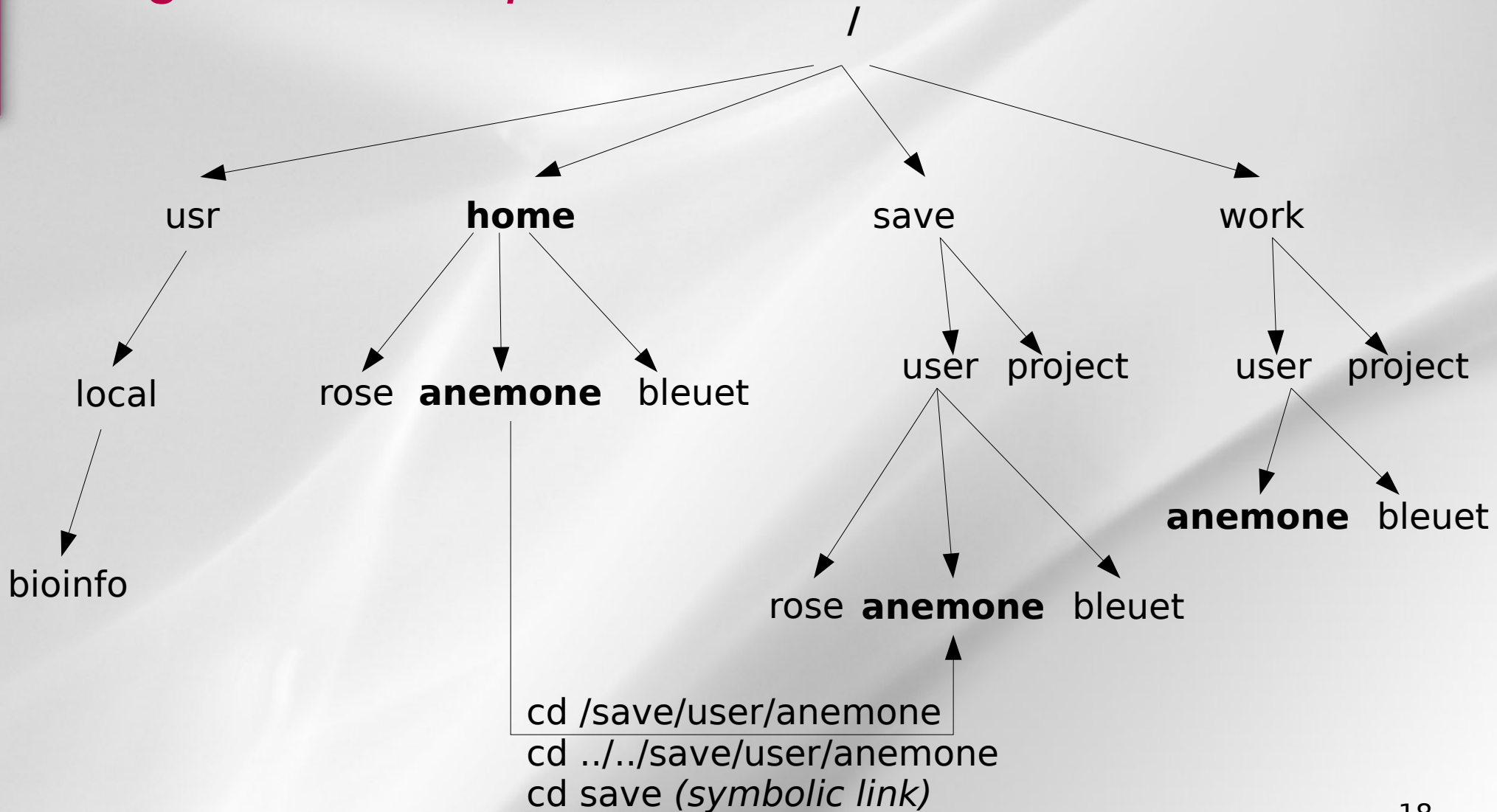
The tree structure

Symbolic link



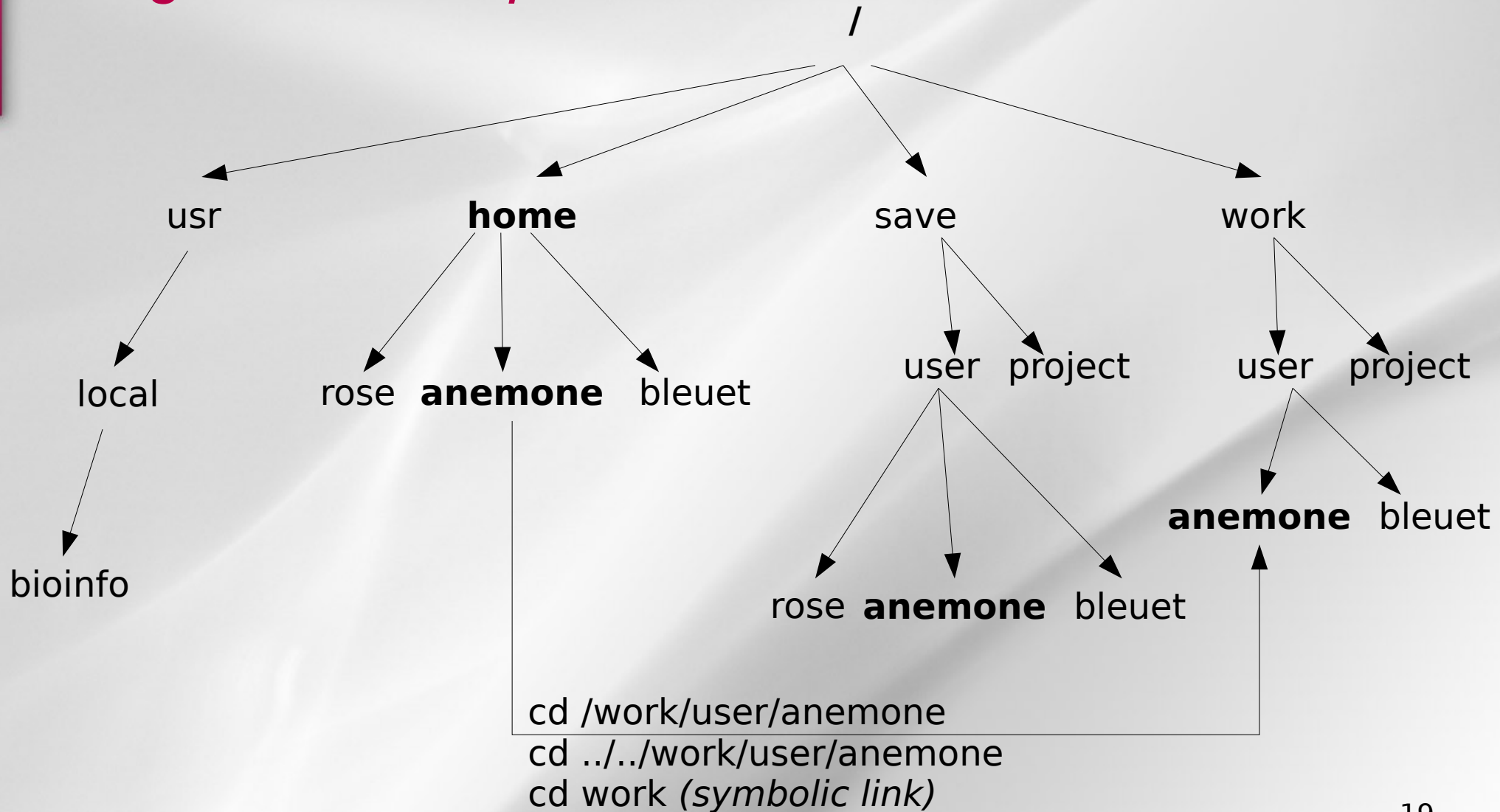
The tree structure

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 genologin ?

From Windows

- **MobaXterm**
- **Other tools : wsl / openssh**



MobaXterm Home Edition v23.2
(Portable edition)

From Linux / Mac

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

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


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.mia.inra.fr/linux-cluster/linux-initiation/tp1/>

Part II

Plan

- File types
- File permissions
- Manipulating files
- Displaying files
- wild card characters
- Disk space control
- TP2

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      13 Mar 18  2009 work -> /work/user/cnoirot

```

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

File types

Read, write, execute

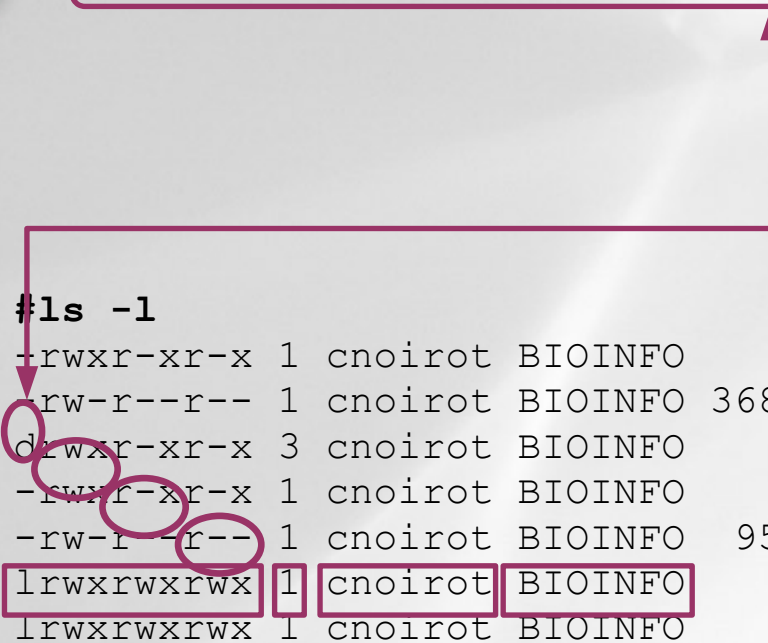
Type - User - Group - Others

#ls -l

```

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

```



Permissions

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
```

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
```

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]*
```

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
```

Looking into files

Display a file content

cat file_name : display the file content

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

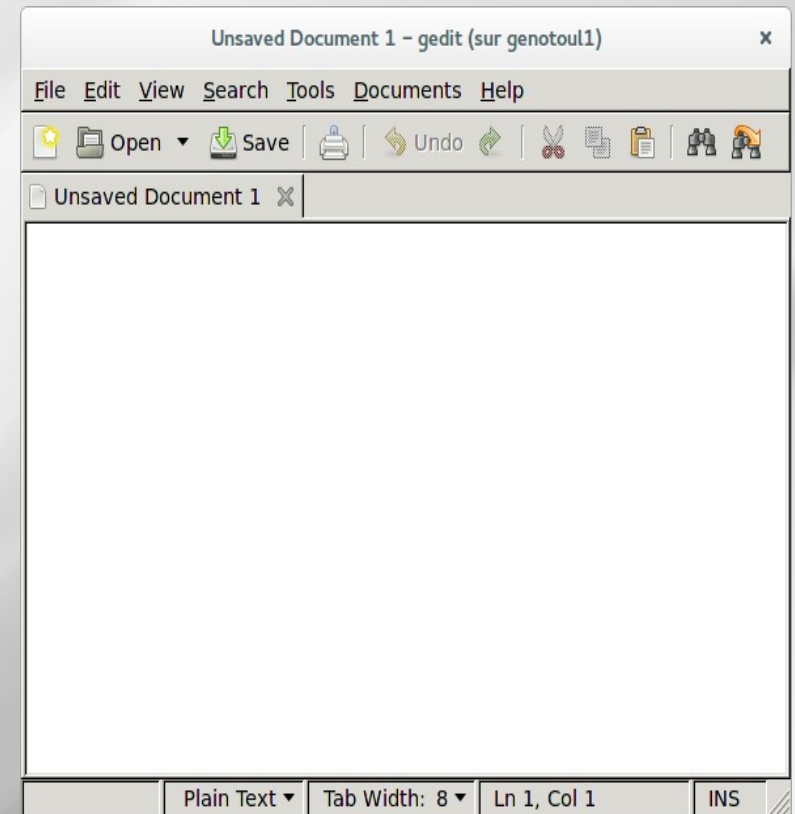
less file_name : display up and down (more than **more**)

```
#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



Disk space control

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

Disk space control

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

```
du -csh /home/formation/*  
483K   /home/formation/bin  
26K    /home/formation/comptes.txt  
242K   /home/formation/last.txt  
1.5K   /home/formation/public_html  
1.5K   /home/formation/save  
26K    /home/formation/tgjcl.cfg  
1.5K   /home/formation/work  
780K   total
```

Part II

TP

- Do the exercises

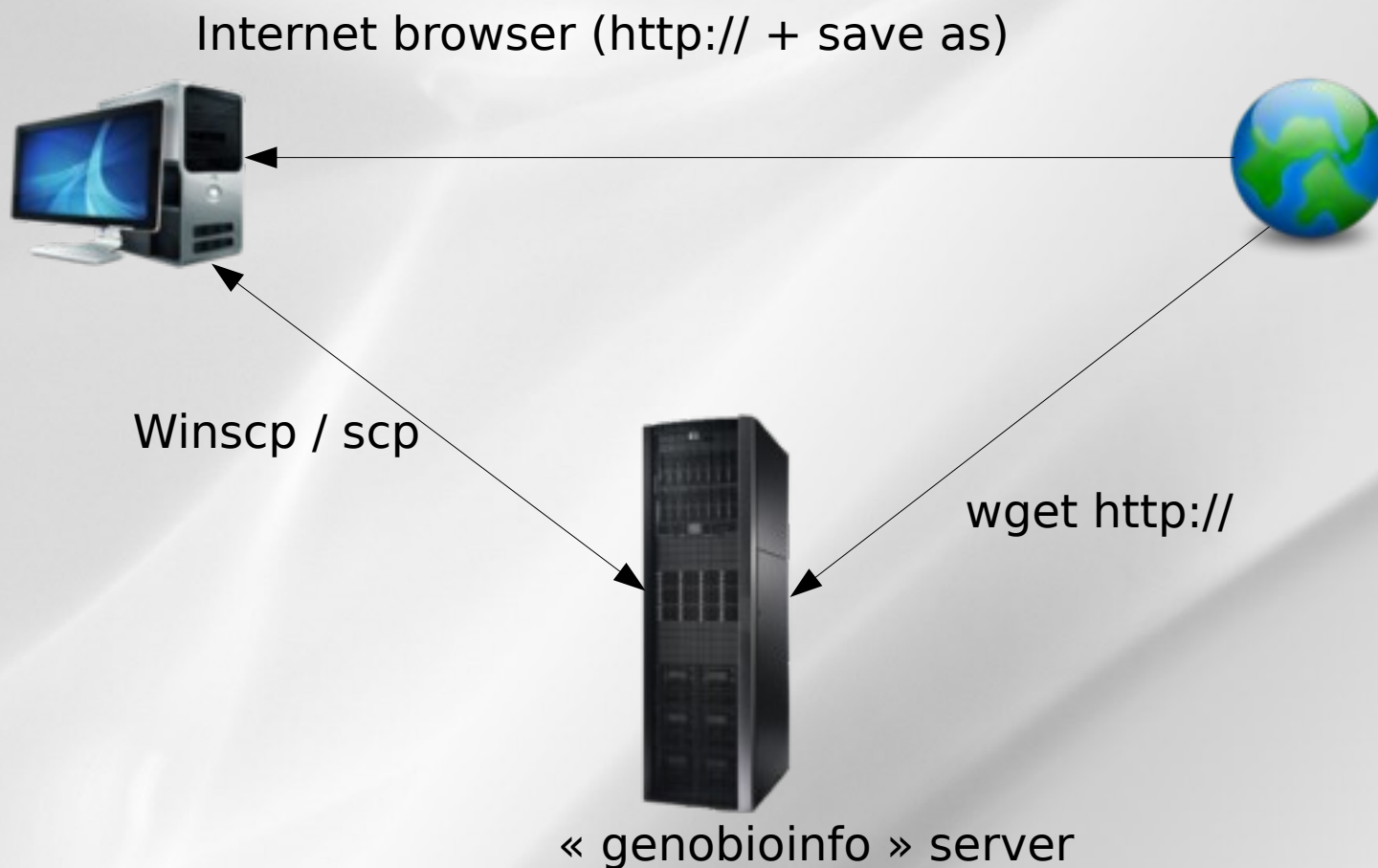
<https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/linux-initiation/tp2/>

Plan

- Downloading / transferring
- Compressing / uncompressing
- Utility commands
- Data extractions commands
- Redirections
- 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

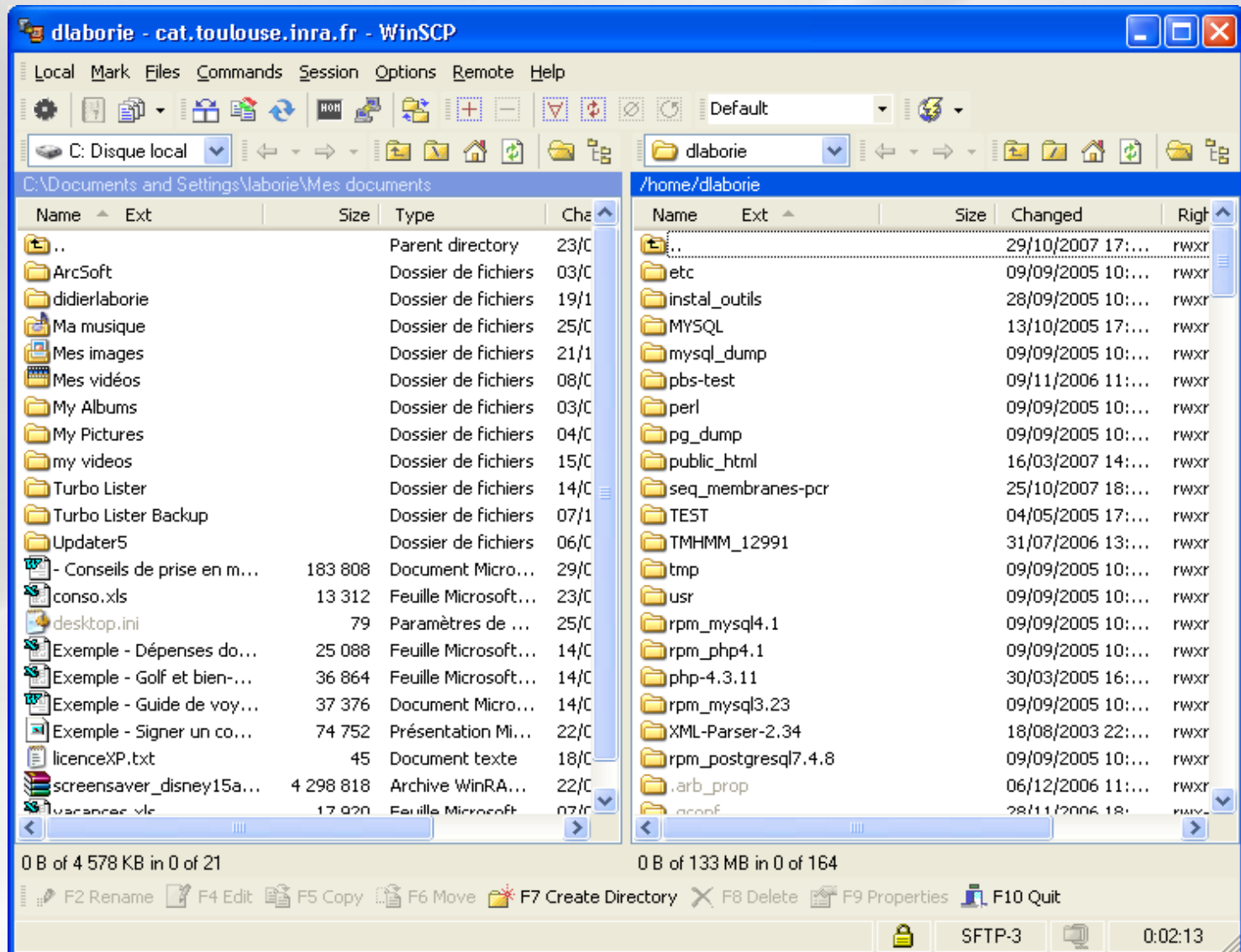
We recommend to use « scp » command (secure copy)

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

scp source_name **bleuet@genobioinfo:destination_name**
(copy from desktop to "genologin")

Downloading / transferring

WinSCP / FileZilla : copy via graphical interface



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

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

tkdiff [-options] file_name1 file_name2
compare two files (line per line)

```
tkdiff 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
```


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 languages (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
```

Part III

TP3

- Do the exercises

<https://genotoul-bioinfo.pages.mia.inra.fr/linux-cluster/linux-initiation/tp3/>

End of presentation

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

Thanks for your attention !

