Plateforme Bioinformatique Midi-Pyrénées

Training Day : Linux

Objectives

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

- Connect to «genologin» 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 - 10h45

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

Genotoul Bioinfo

Mission

Provide to public regional community :

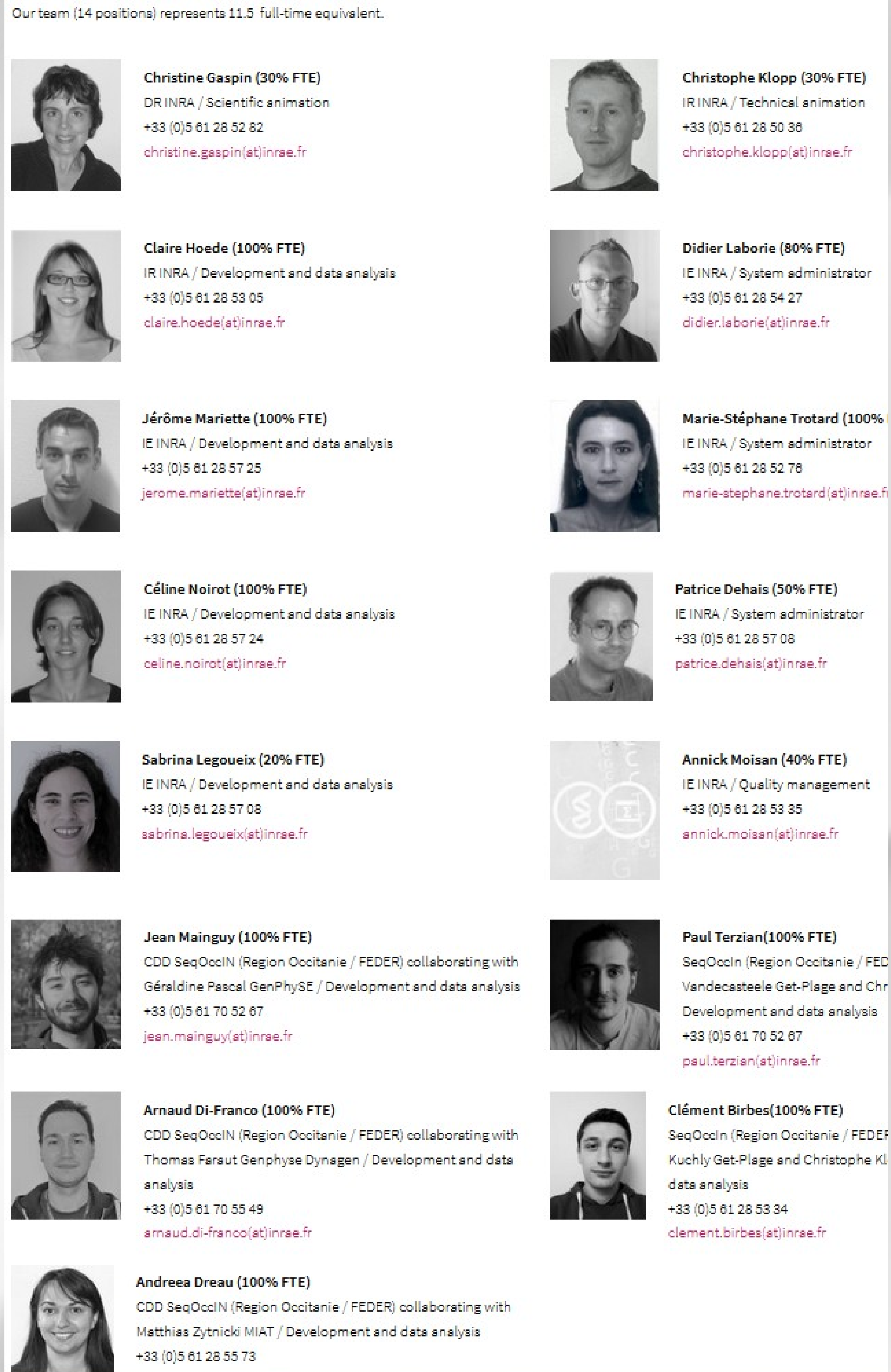
Equipment

- Storage disk space & computers farm
- Hosting facilities (virtual machine, web site)

Services

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

The team



Our team (14 positions) represents 11.5 full-time equivalent.

Christine Gaspin (30% FTE)
DR INRA / Scientific animation
+33 (0)5 61 28 52 82
christine.gaspin(at)inrae.fr

Christophe Klopp (30% FTE)
IR INRA / Technical animation
+33 (0)5 61 28 50 36
christophe.klopp(at)inrae.fr

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

Didier Laborie (80% FTE)
IE INRA / System administrator
+33 (0)5 61 28 54 27
didier.laborie(at)inrae.fr

Jérôme Mariette (100% FTE)
IE INRA / Development and data analysis
+33 (0)5 61 28 57 25
jerome.mariette(at)inrae.fr

Marie-Stéphane Trotard (100%
IE INRA / System administrator
+33 (0)5 61 28 52 76
marie-stephane.trotard(at)inrae.f

Céline Noirot (100% FTE)
IE INRA / Development and data analysis
+33 (0)5 61 28 57 24
celine.noirot(at)inrae.fr

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

Sabrina Legoueix (20% FTE)
IE INRA / Development and data analysis
+33 (0)5 61 28 57 08
sabrina.legoueix(at)inrae.fr

Annick Moisan (40% FTE)
IE INRA / Quality management
+33 (0)5 61 28 53 35
annick.moisan(at)inrae.fr

Jean Mainguy (100% FTE)
CDD SeqOccIN (Region Occitanie / FEDER) collaborating with Géraldine Pascal GenPhySE / Development and data analysis
+33 (0)5 61 70 52 67
jean.mainguy(at)inrae.fr

Paul Terzian(100% FTE)
SeqOccIn (Region Occitanie / FED
Vandecasteele Get-Plage and Chr
Development and data analysis
+33 (0)5 61 70 52 67
paul.terzian(at)inrae.fr

Arnaud Di-Franco (100% FTE)
CDD SeqOccIN (Region Occitanie / FEDER) collaborating with Thomas Faraut Genphyse Dynagen / Development and data analysis
+33 (0)5 61 70 55 49
arnaud.di-franco(at)inrae.fr

Clément Birbes(100% FTE)
SeqOccIn (Region Occitanie / FEDE
Kuchly Get-Plage and Christophe Kl
data analysis
+33 (0)5 61 28 53 34
clement.birbes(at)inrae.fr

Andreea Dreau (100% FTE)
CDD SeqOccIN (Region Occitanie / FEDER) collaborating with Matthias Zytnicki MIAT / Development and data analysis
+33 (0)5 61 28 55 73

Genotoul Bioinfo

The users

1000 authenticated users :

- Regional laboratories (+ some others)

(CNRS, INRA, ENSAT, INSERM, INSA, UPS...)

- Biologists and bio-informaticians

Genotoul Bioinfo

Equipments

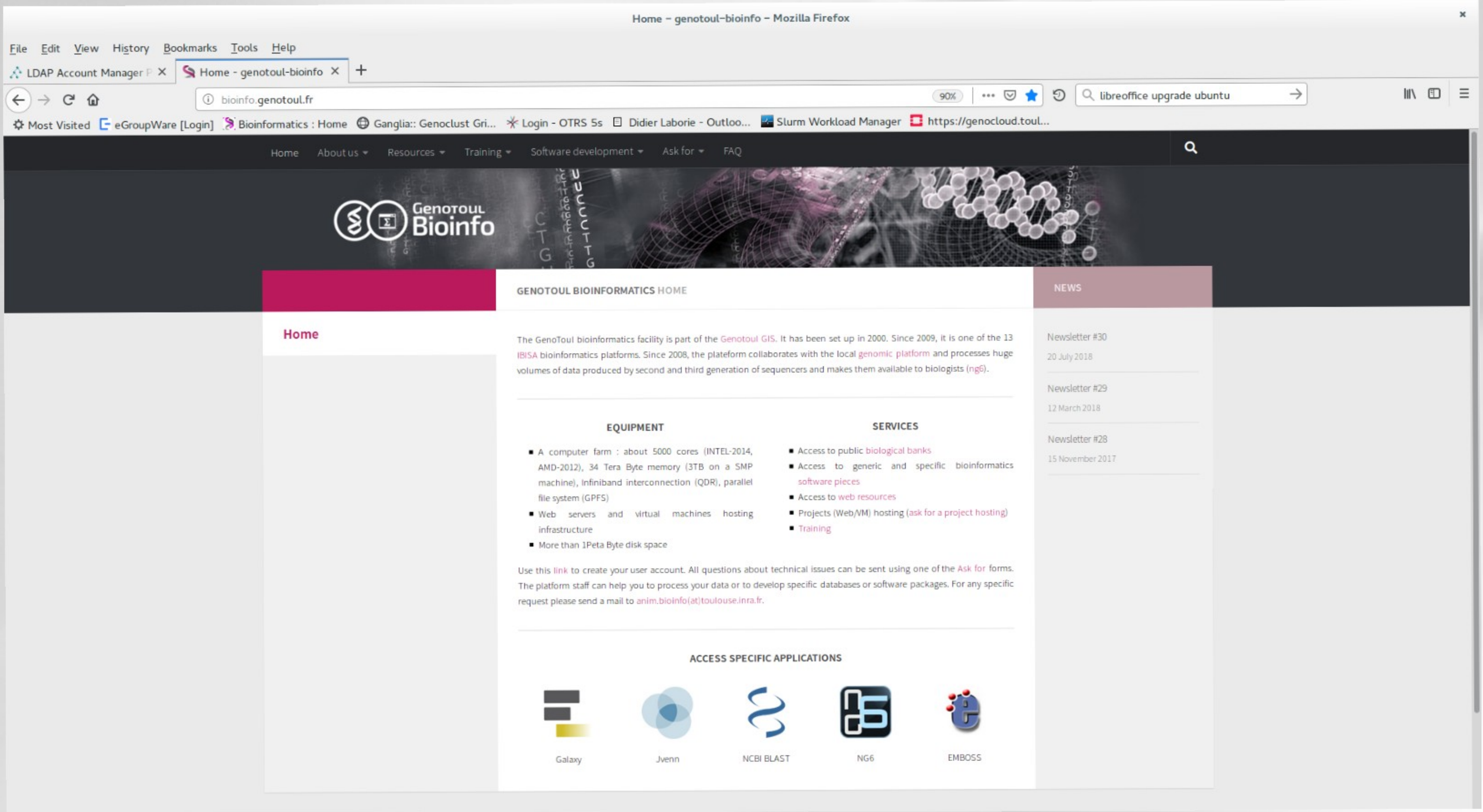
- **Several servers : physical or virtual machines**
  capacities for servers hosting and web services
- **Computational cluster**
  about 5000 cores
  4 PB disk space available for computing
- **Permanent Storage File System**
  2*1 PB disk space capacities (asynchronous replication)

Disk spaces

/usr/local/bioinfo/src	Bioinformatics Software
/bank	International genomics Databanks
/home	User configuration files (ONLY) (1 GB user quota)
/work	HPC computational disk space (TEMPORARY) (1 TB user quota)
/save	User disk space (with BACKUP) (250 GB user quota + replication)

Genotoul Bioinfo

http://bioinfo.genotoul.fr



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 plateform 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(at)toulouse.inra.fr.

ACCESS SPECIFIC APPLICATIONS

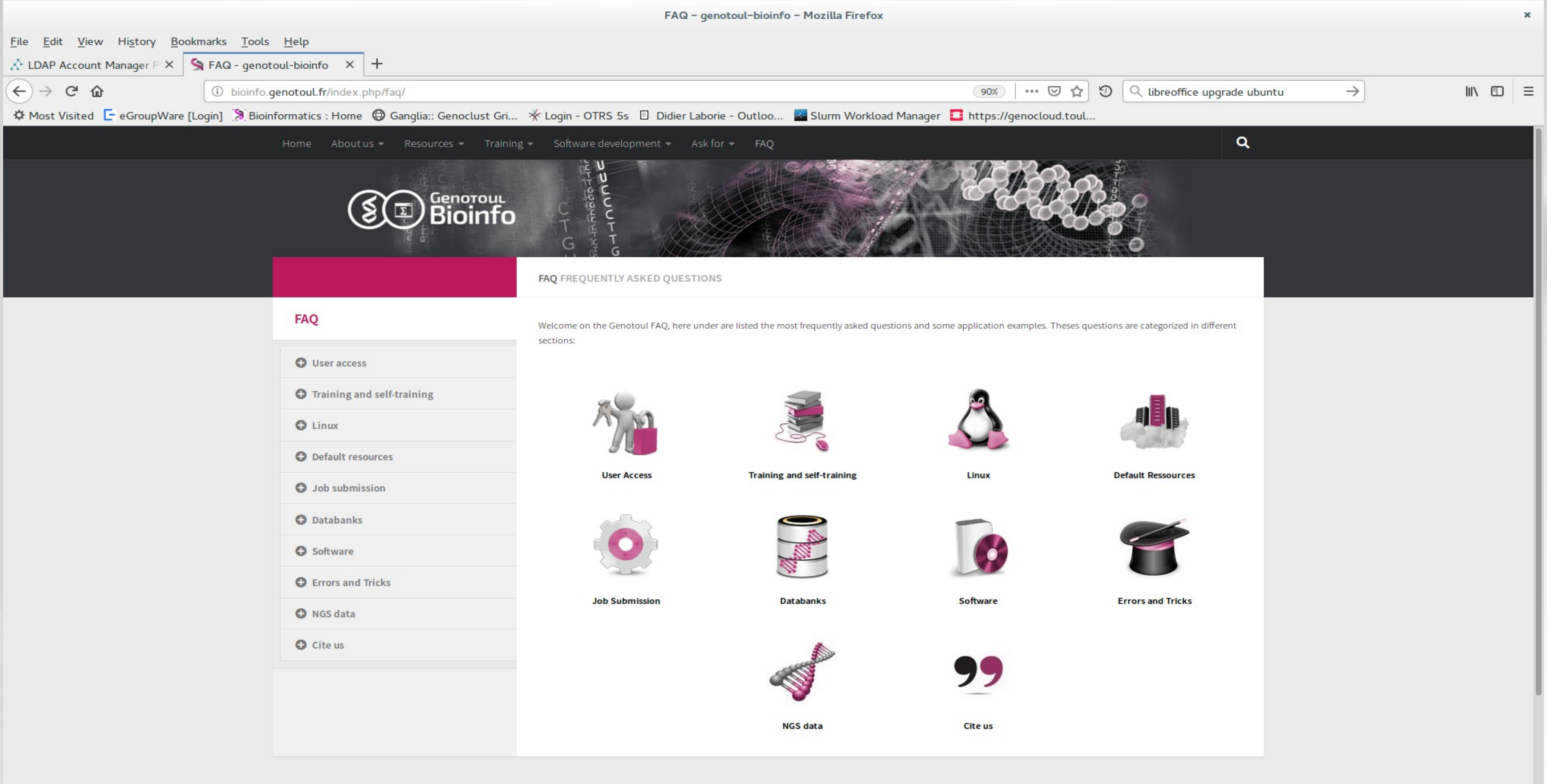
Newsletter #30
20 July 2018

Newsletter #29
12 March 2018

Newsletter #28
15 November 2017

Genotoul Bioinfo

Questions=> support.genopole@inrae.fr



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:

Linux

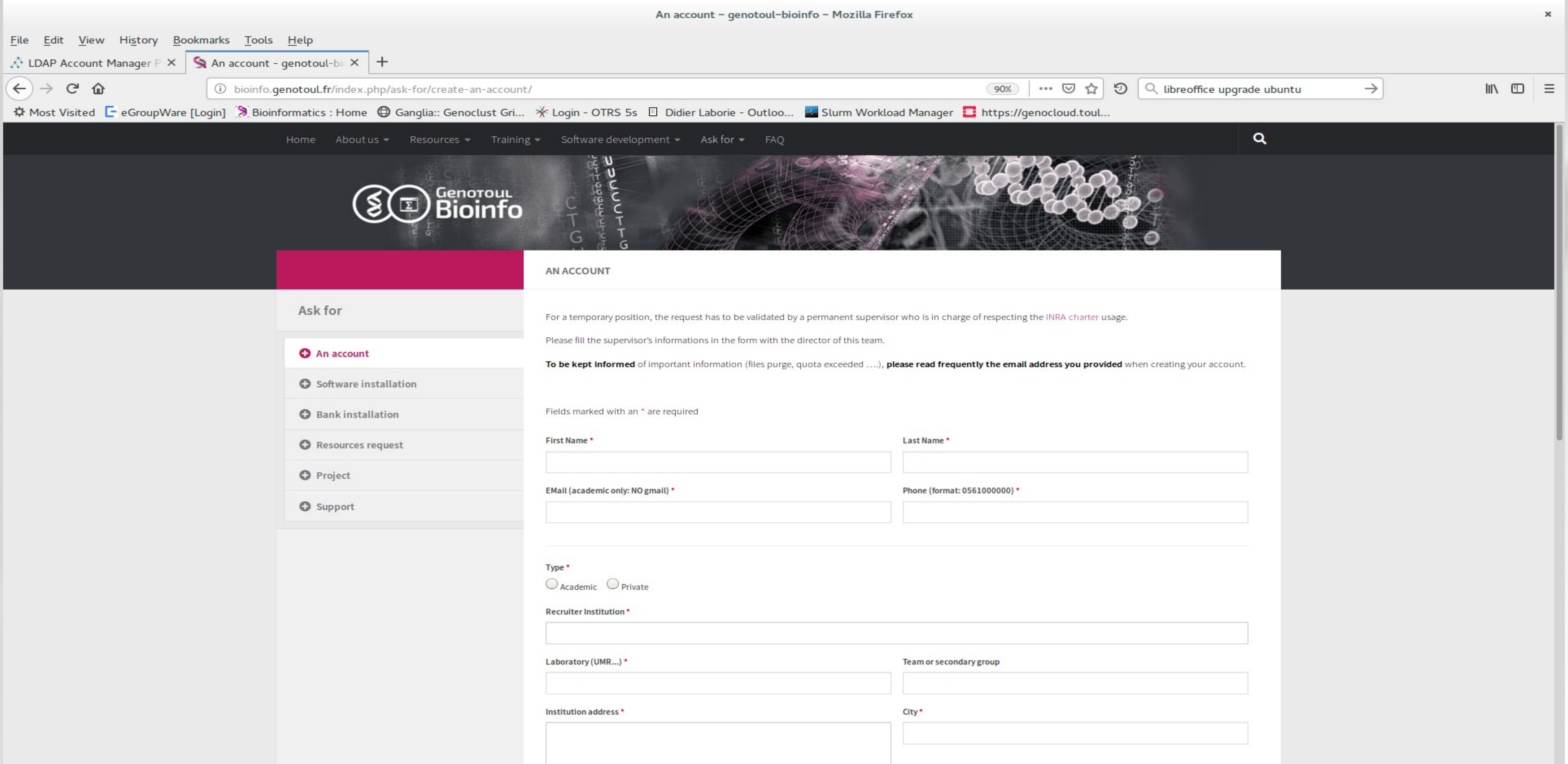
Introduction to Linux

GNU-Linux : Unix-like operating system

- Initial Developer = Linus Torvalds (Helsinki)
- Birth of kernel Linux on 1991
- GNU project = free and open source software
- Over three hundreds of active distributions (large community of developers)
- Some are commercial : Fedora (RedHat), openSUSE(SUSE), Ubuntu(Canonical), Mandriva

Linux

Asking for an account



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 *

EMail (academic only: NO gmail) *

Phone (format: 0561000000) *

Type *

○ Academic ○ Private

Recruiter Institution *

Laboratory (UMR...) *

Team or secondary group

Institution address *

City *

Linux

Linux account

Access to a work environment

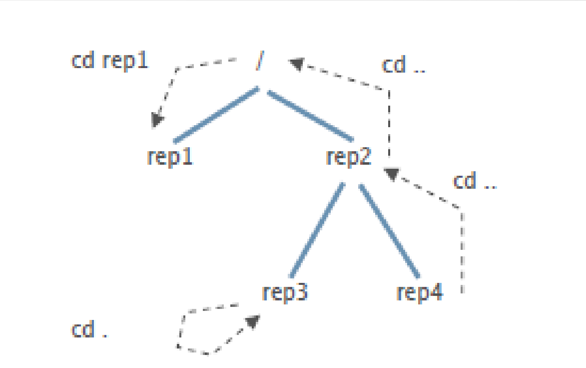
- Login + password
- Share resources (cpu, memory, disk)
- Usage of software installed
- Free access to computational cluster
- Own space disk (/save & /work directory)

The tree structure

Navigation

Tree structure

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



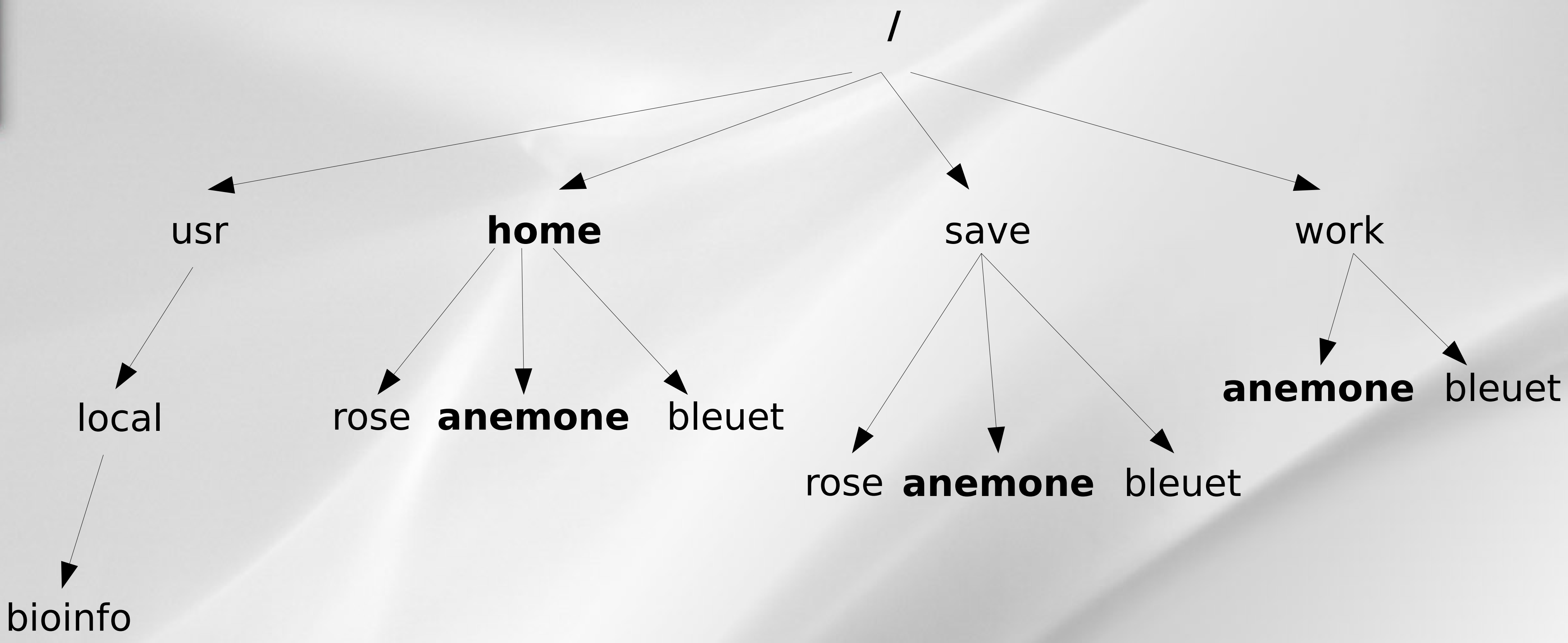
cd [nom_répertoire] : Change directory

Absolute path : `/home/bleuet`

Relative path : `../../usr`

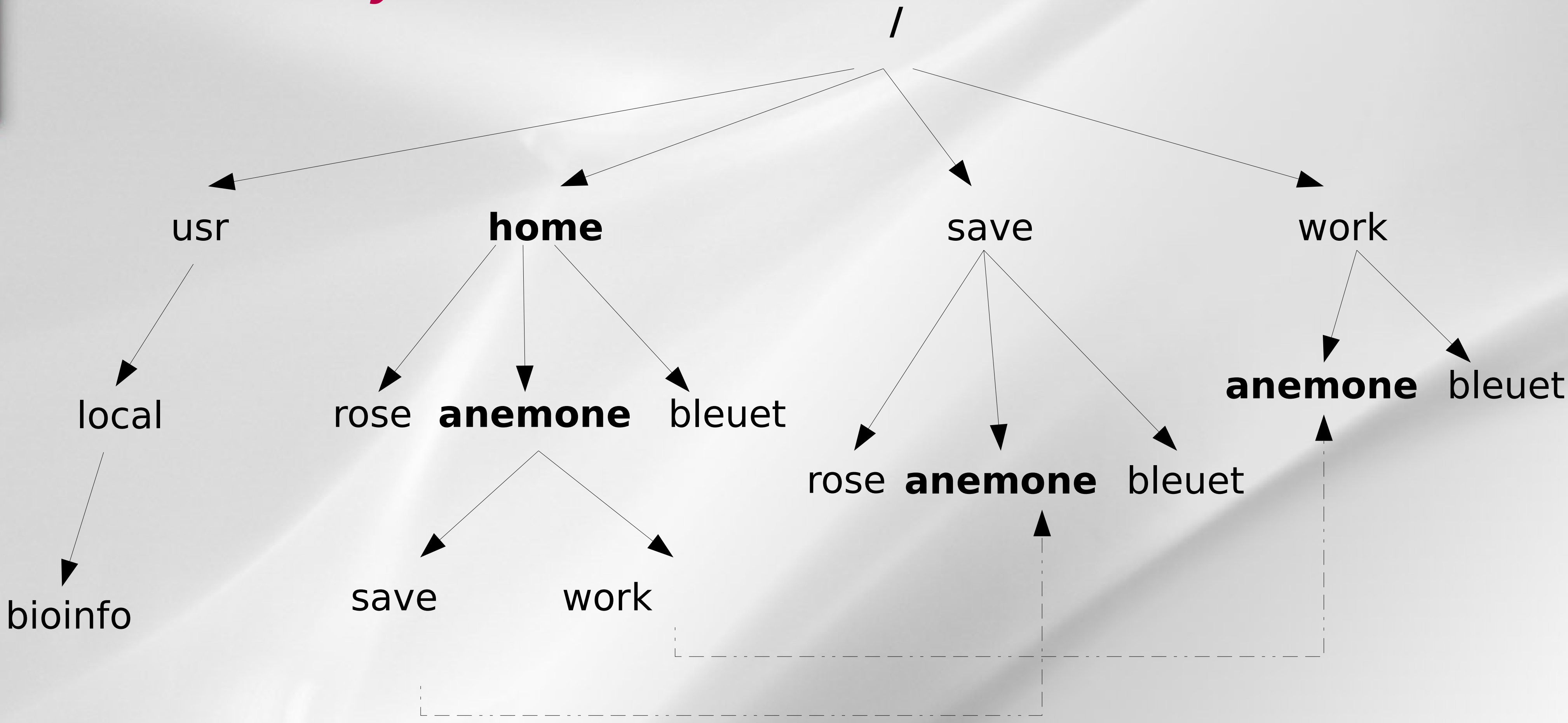
The tree structure

Notion of « Root »



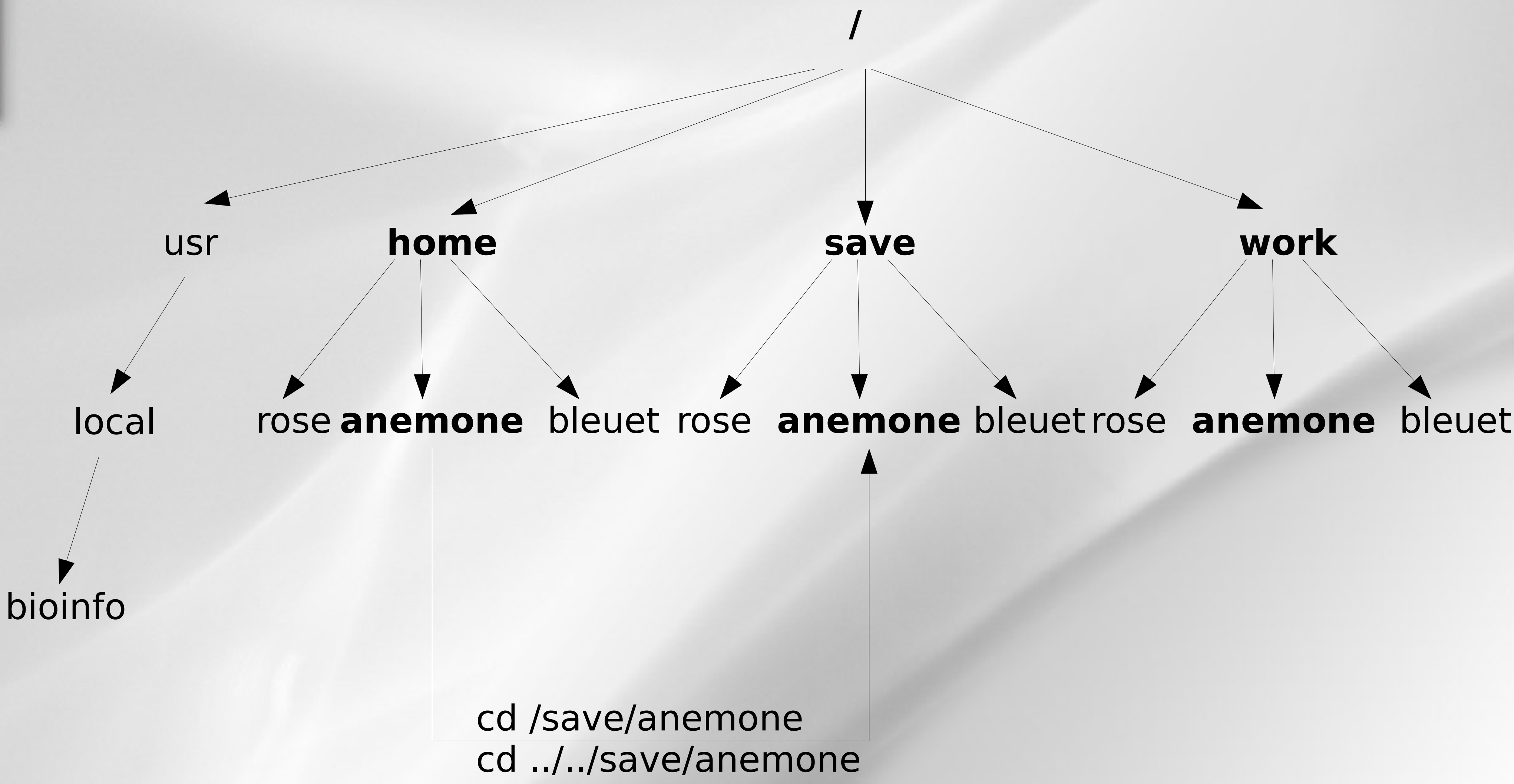
The tree structure

Notion of «symbolic links»



The tree structure

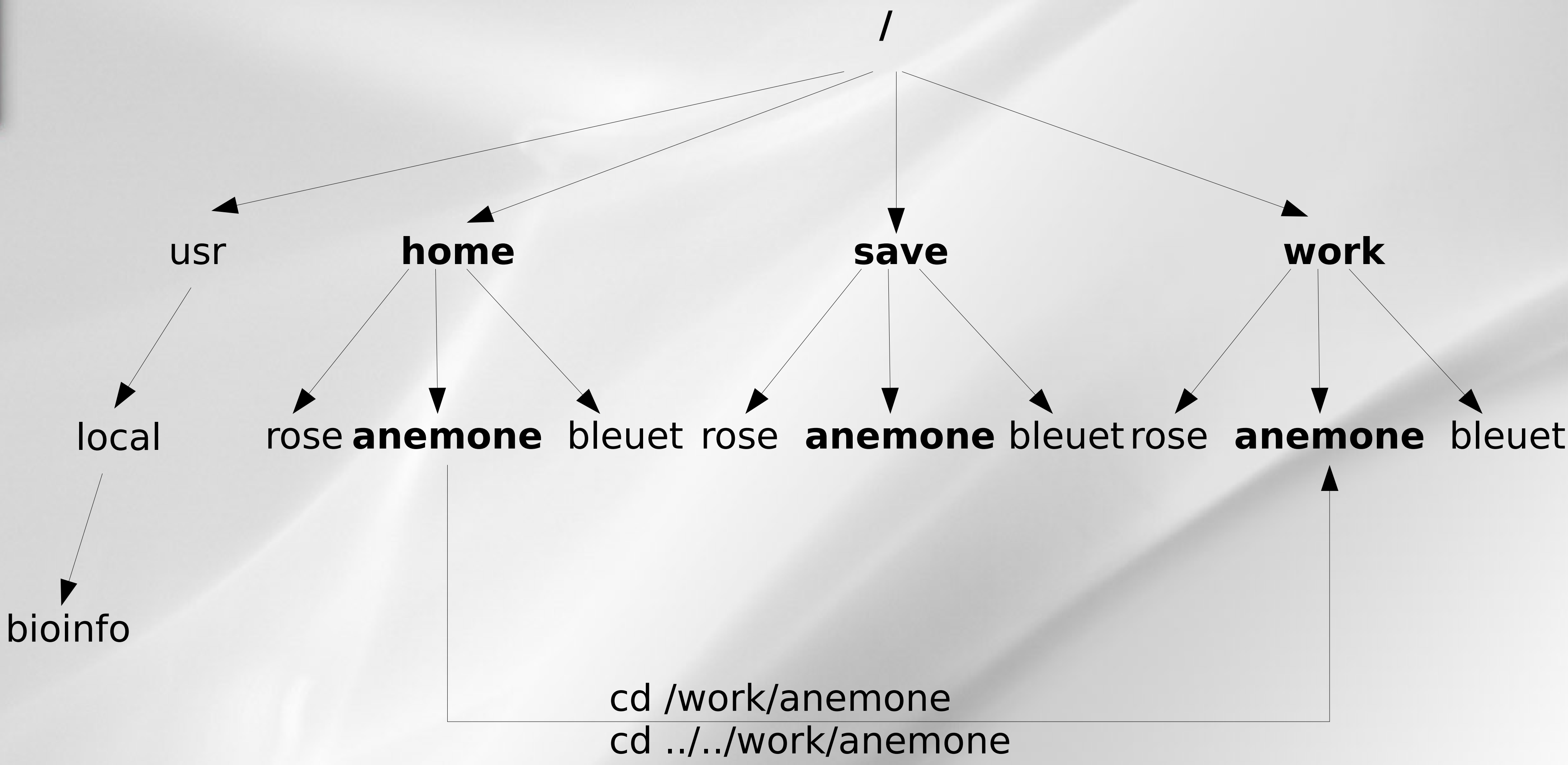
Navigation : examples



```
cd /save/anemone
cd ../../save/anemone
```

The tree structure

Navigation : examples



cd /work/anemone
cd ../../work/anemone

The command line : syntax

command_name [-option] [parameter]

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

```
ls -l /home

tree
```

The help on command

command_name -- help

man command_name

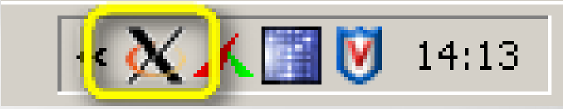
```
ls --help
blastall –help

man ls
man cd
```

Some basics commands

- **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



- **Putty + Xming** (Windows graphic)
- **MobaXterm**

From Linux / Mac

- **ssh username@genologin.toulouse.inra.fr** (command line)

Very important 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)

- **Command and path completion :**
  - Use the TAB key

- **Back to the previous used commands :**
  - Use the « up » and « down » keys

TP1

- Connect yourself to genologin server with your (training) login/password

anemone aster bleuet iris muguet
narcisse pensee rose tulipe violette…

- Do the exercices (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]

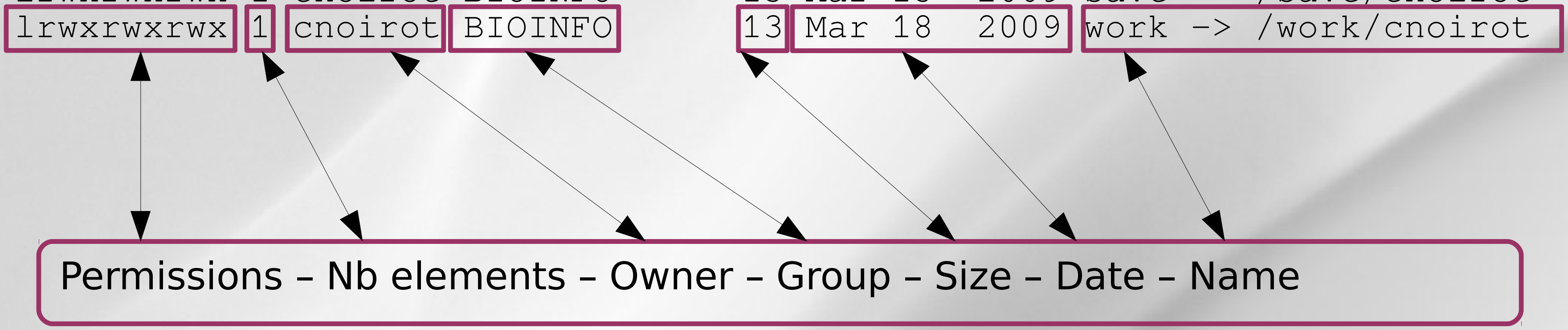
- **-a** : display hidden files/dir
- **-l** : use the long format
- **-t** : sort the content
- **-r** : reverse the sort order

```
ls -l /usr/local/bioinfo/src
drwxr-sr-x  3 laborie     bioadm     164 Mar 14  2014 VelvetOptimiser-2.2.5
drwxrwsr-x  6 dehais      bioadm     300 Feb 18  2015 VIENNA
drwxr-sr-x  3 mtrotard    bioadm     133 Sep 21 13:21 ViennaNGS
```

File types

“ls -l” command (long listing format)

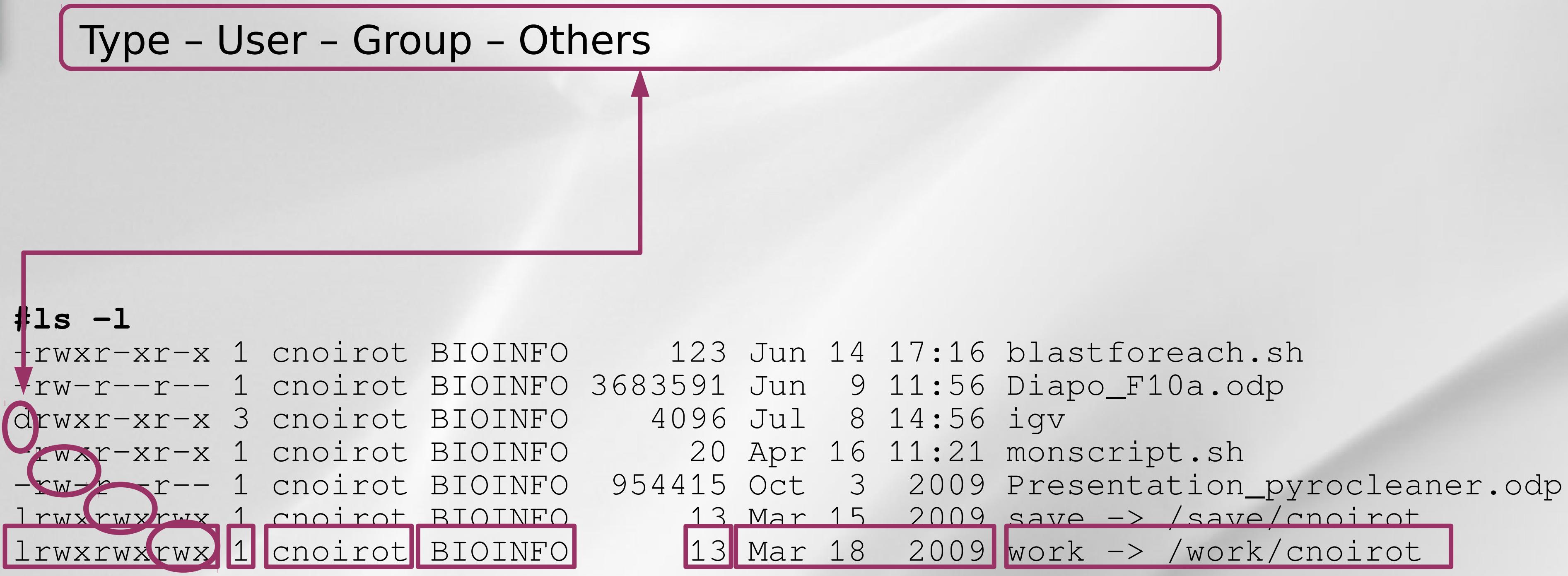
```
#ls -l
-rwxr-xr-x 1 cnoirot BIOINFO     123 Jun 14 17:16 blastforeach.sh
-rw-r--r-- 1 cnoirot BIOINFO 3683591 Jun  9 11:56 Diapo_F10a.odp
drwxr-xr-x 3 cnoirot BIOINFO    4096 Jul  8 14:56 igv
-rwxr-xr-x 1 cnoirot BIOINFO      20 Apr 16 11:21 monscript.sh
-rw-r--r-- 1 cnoirot BIOINFO  954415 Oct 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



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

File types

Read, write,execute



Type – User – Group – Others

```
#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/cnoirot
lrwxrwxrwx 1 cnoirot BIOINFO      13 Mar 18  2009 work -> /work/cnoirot
```

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
```

ln -s nom_fic_source nom_fic_destination
create a symbolic link

```
ln -s file_name link_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

Finding files/dir

find dirname [-option] [parameter]

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

    find . -type d

    find . -type f

    find / -size +1000k
```

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]*
```

Displaying 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.

ACGGTCAGACGTTTGGCCCGACCACCGGGATGAGGCTGACGCAGGTCAGAAATCTTTGTGACGAC
   AACCGTATCAATGCCGGTGTGG...
```

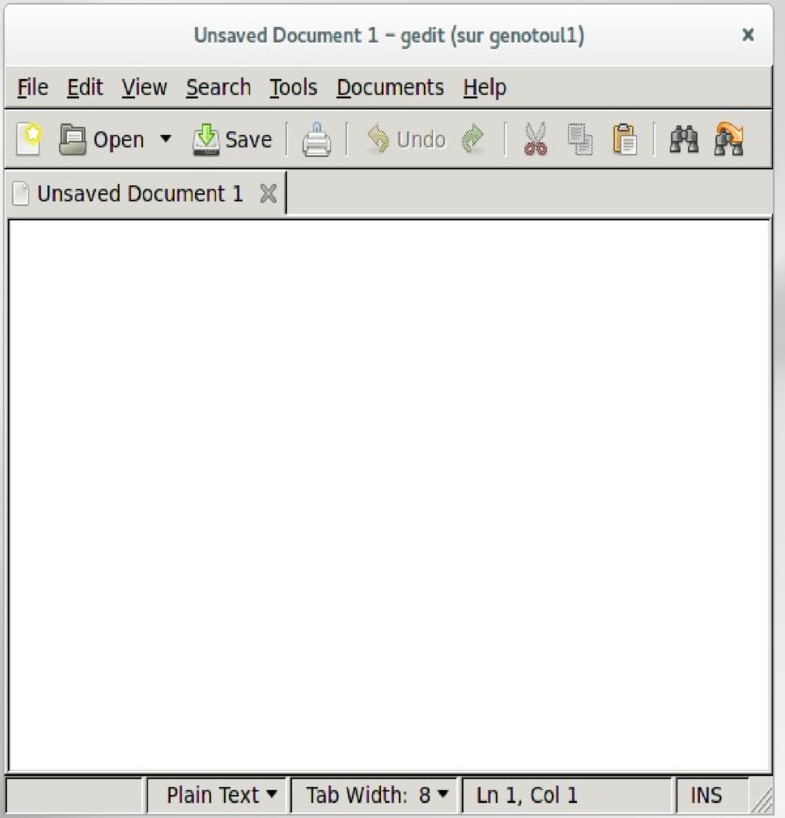
more file_name : display more and more

less file_name : display up and down

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/tgicl.cfg
1.5K    /home/formation/work
780K    total
```

Part II

TP

- Do the exercises

Part III

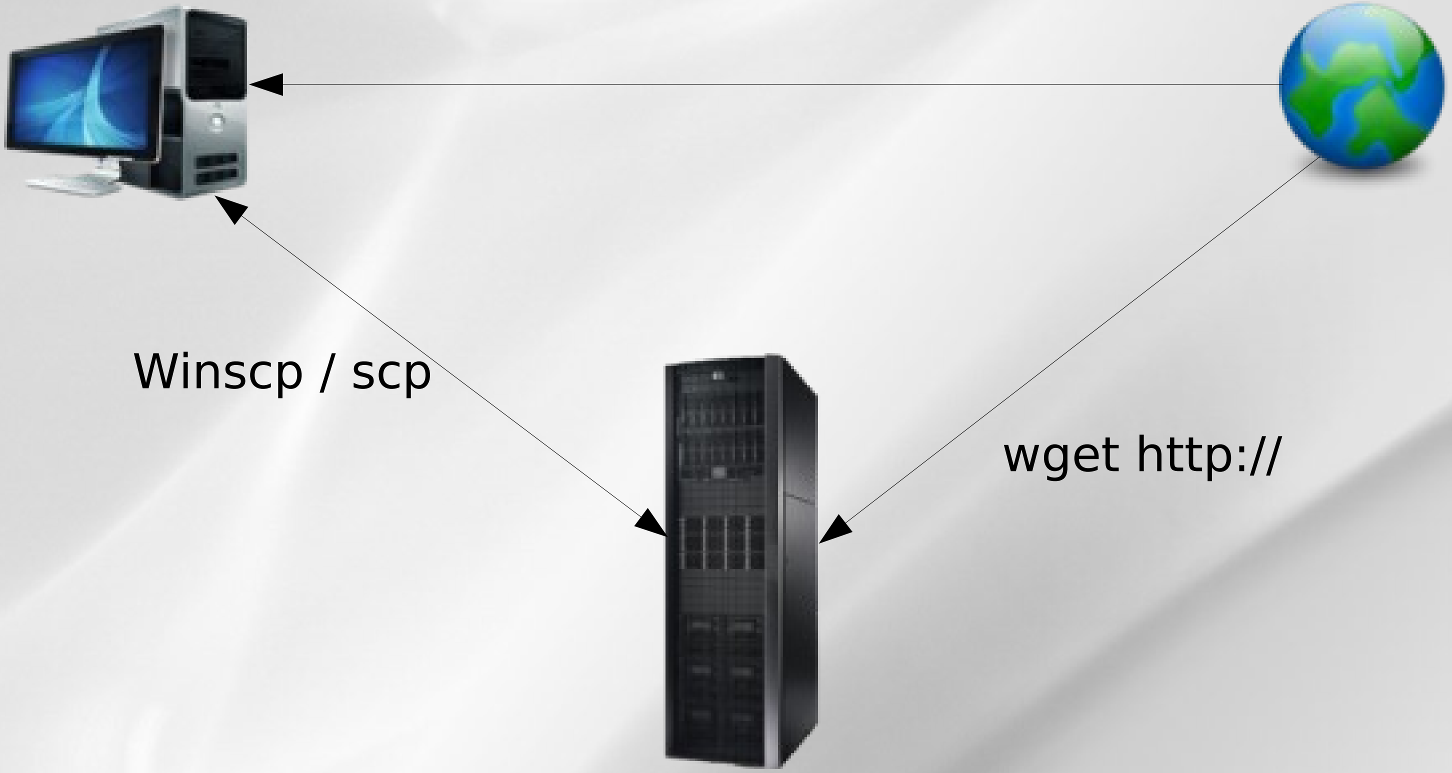
Plan

- Downloading / transferring
- Compressing / uncompressing

- Utility commands
- Data extractions commands
- Redirections
- My first script

Downloading / transferring

Several possible cases



Internet browser (http:// + save as)

Winscp / scp

wget http://

« genologin » server

Downloading / transferring

Directly from internet to genologin

File download from Internet to « genologin server »:

- Copy the URL of the file to download

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

Downloading / transferring

Transfer between genologin and desktop computer

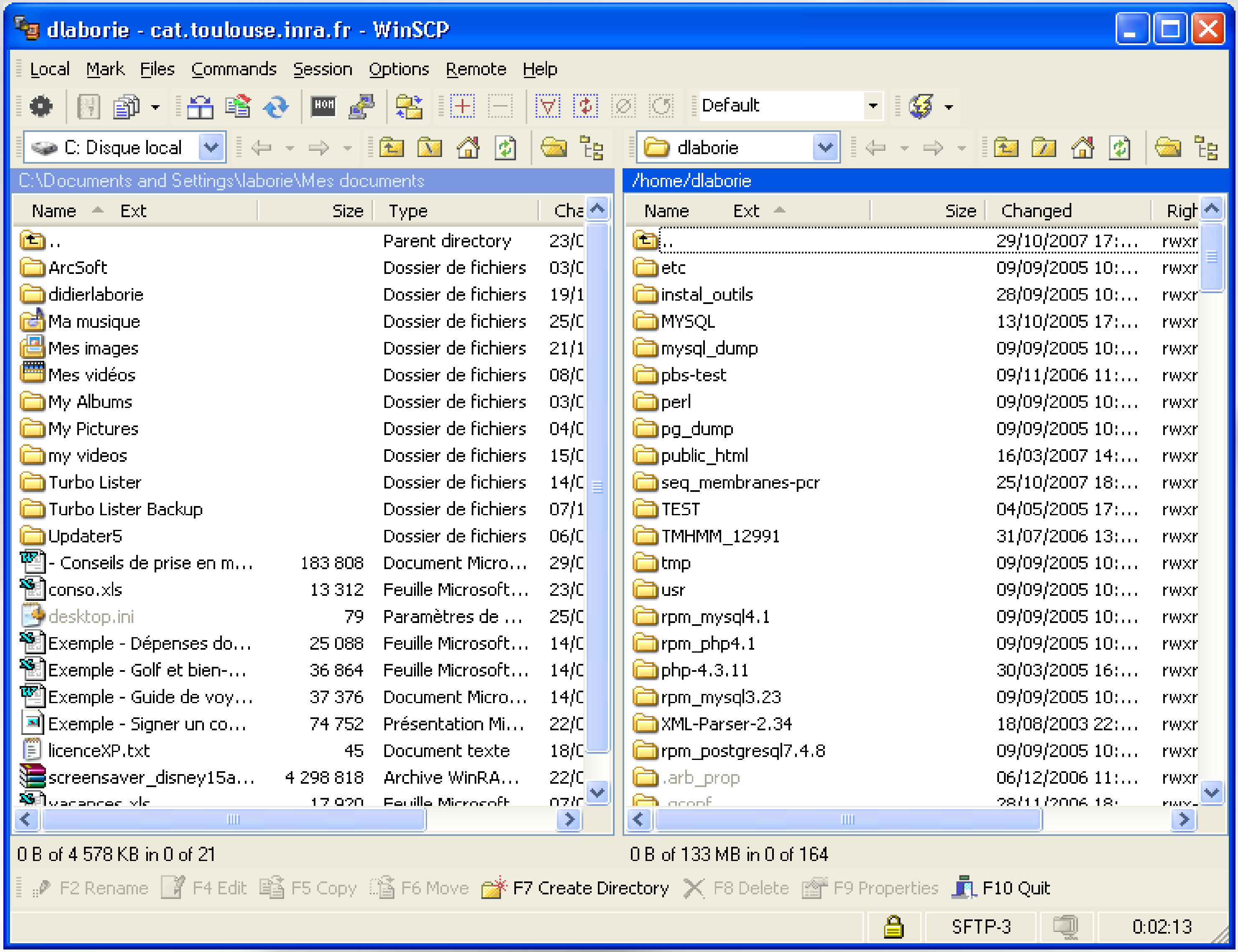
We recommend to use « scp » command (secure copy)

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

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

Downloading / transferring

WinSCP / FileZilla : copy via graphical interface



C:\Documents and Settings\laborie\Mes documents

Name	Ext	Size	Type	Cha
..			Parent directory	23/0
ArcSoft			Dossier de fichiers	03/0
didierlaborie			Dossier de fichiers	19/1
Ma musique			Dossier de fichiers	25/0
Mes images			Dossier de fichiers	21/1
Mes vidéos			Dossier de fichiers	08/0
My Albums			Dossier de fichiers	03/0
My Pictures			Dossier de fichiers	04/0
my videos			Dossier de fichiers	15/0
Turbo Lister			Dossier de fichiers	14/0
Turbo Lister Backup			Dossier de fichiers	07/1
Updater5			Dossier de fichiers	06/0
- Conseils de prise en m...		183 808	Document Micro...	29/0
conso.xls		13 312	Feuille Microsoft...	23/0
desktop.ini		79	Paramètres de ...	25/0
Exemple - Dépenses do...		25 088	Feuille Microsoft...	14/0
Exemple - Golf et bien-...		36 864	Feuille Microsoft...	14/0
Exemple - Guide de voy...		37 376	Document Micro...	14/0
Exemple - Signer un co...		74 752	Présentation Mi...	22/0
licenceXP.txt		45	Document texte	18/0
screensaver_disney15a...		4 298 818	Archive WinRA...	22/0
vacances.xls		17 920	Feuille Microsoft	07/0

0 B of 4 578 KB in 0 of 21

/home/dlaborie

Name	Ext	Size	Changed	Righ
..			29/10/2007 17:...	rwxr
etc			09/09/2005 10:...	rwxr
instal_outils			28/09/2005 10:...	rwxr
MYSQL			13/10/2005 17:...	rwxr
mysql_dump			09/09/2005 10:...	rwxr
pbs-test			09/11/2006 11:...	rwxr
perl			09/09/2005 10:...	rwxr
pg_dump			09/09/2005 10:...	rwxr
public_html			16/03/2007 14:...	rwxr
seq_membranes-pcr			25/10/2007 18:...	rwxr
TEST			04/05/2005 17:...	rwxr
TMHMM_12991			31/07/2006 13:...	rwxr
tmp			09/09/2005 10:...	rwxr
usr			09/09/2005 10:...	rwxr
rpm_mysql4.1			09/09/2005 10:...	rwxr
rpm_php4.1			09/09/2005 10:...	rwxr
php-4.3.11			30/03/2005 16:...	rwxr
rpm_mysql3.23			09/09/2005 10:...	rwxr
XML-Parser-2.34			18/08/2003 22:...	rwxr
rpm_postgresql7.4.8			09/09/2005 10:...	rwxr
.arb_prop			06/12/2006 11:...	rwxr
.gconf			28/11/2006 18:...	rwx-

0 B of 133 MB in 0 of 164

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
```

<u>Other formats</u> : 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
```

<u>Tips</u>: 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

```
nedit 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
```

Part III

TP3

- Do the exercises