Plateforme Bioinformatique Midi-Pyrénées

geno toul bioinfo

Training Day : Linux

Objectives

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

- Connect to « genotoul » 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

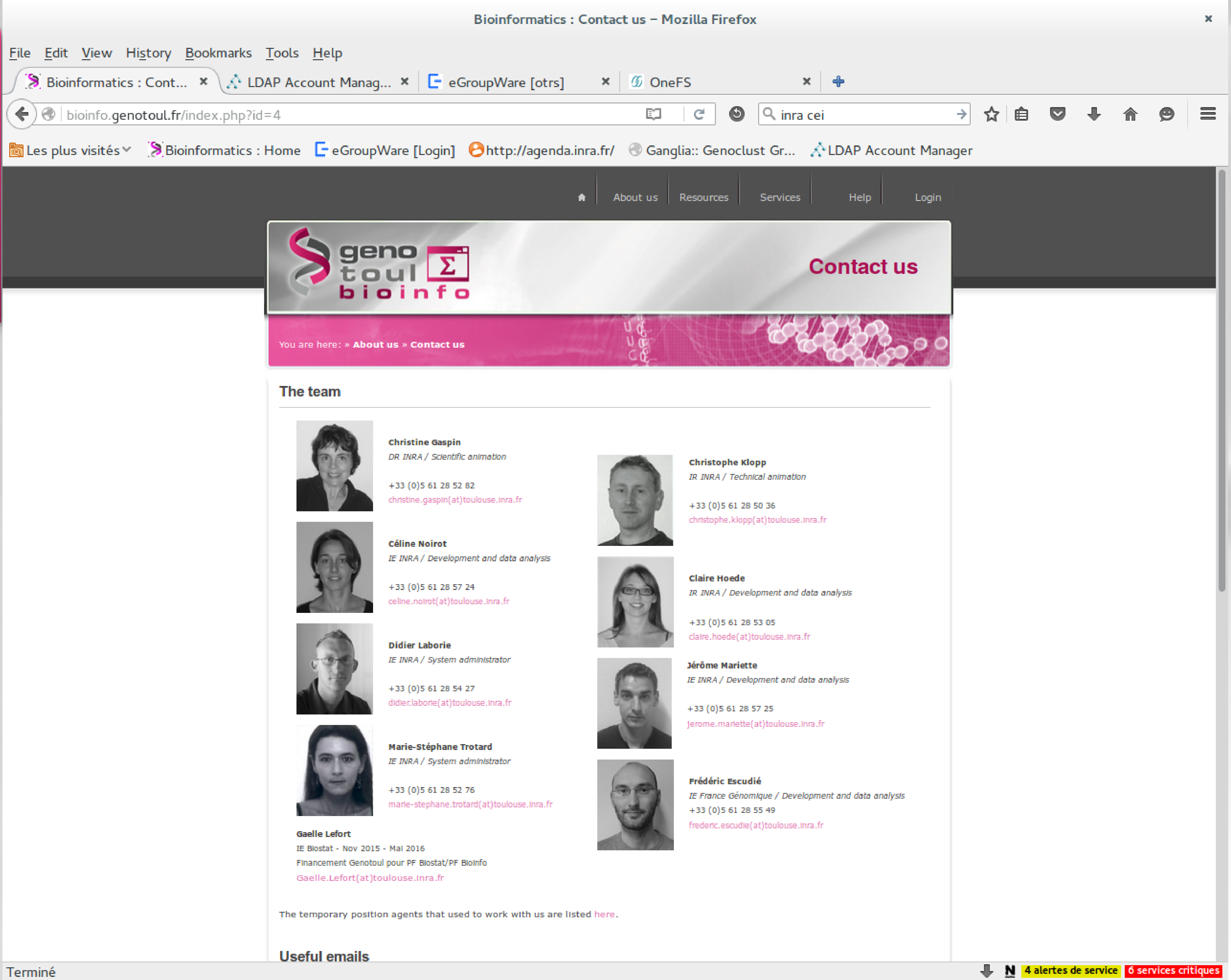
Provide to public regional community :

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



Contact us

You are here: » **About us** » **Contact us**

The team

Christine Gaspin
DR INRA / Scientific animation

+33 (0)5 61 28 52 82
christine.gaspin(at)toulouse.inra.fr

Céline Noirot
IE INRA / Development and data analysis

+33 (0)5 61 28 57 24
celine.noirot(at)toulouse.inra.fr

Didier Laborie
IE INRA / System administrator

+33 (0)5 61 28 54 27
didier.laborie(at)toulouse.inra.fr

Marie-Stéphane Trotard
IE INRA / System administrator

+33 (0)5 61 28 52 76
marie-stephane.trotard(at)toulouse.inra.fr

Gaelle Lefort
IE Biostat - Nov 2015 - Mai 2016
Financement Genotoul pour PF Biostat/PF BioInfo
Gaelle.Lefort(at)toulouse.inra.fr

Christophe Klopp
IR INRA / Technical animation

+33 (0)5 61 28 50 36
christophe.klopp(at)toulouse.inra.fr

Claire Hoede
IR INRA / Development and data analysis

+33 (0)5 61 28 53 05
claire.hoede(at)toulouse.inra.fr

Jérôme Mariette
IE INRA / Development and data analysis

+33 (0)5 61 28 57 25
jerome.mariette(at)toulouse.inra.fr

Frédéric Escudié
IE France Génomique / Development and data analysis
+33 (0)5 61 28 55 49
frederic.escudie(at)toulouse.inra.fr

The temporary position agents that used to work with us are listed here.

Useful emails

The users

800 **authenticated users :**

- Regional laboratories (+ some others)

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

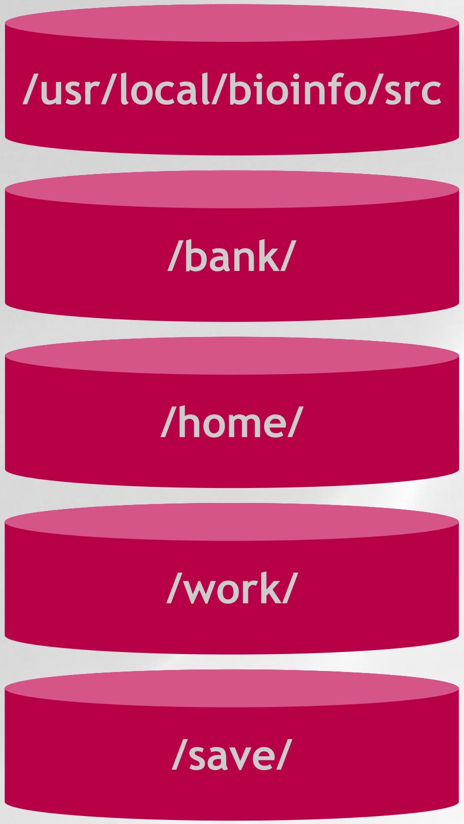
- Biologists and bio-informaticians

Genotoul Bioinfo

Equipments

- **Several servers : physical or virtual machines**
  capacities for servers hosting and web services
- **Computational cluster :**
  about 4000 cores, 34 TB memory
  2*200 TB disk space available for computing
- **Permanent Storage File System :**
  2*500 TB disk space capacities (asynchronous replication)

Disk spaces

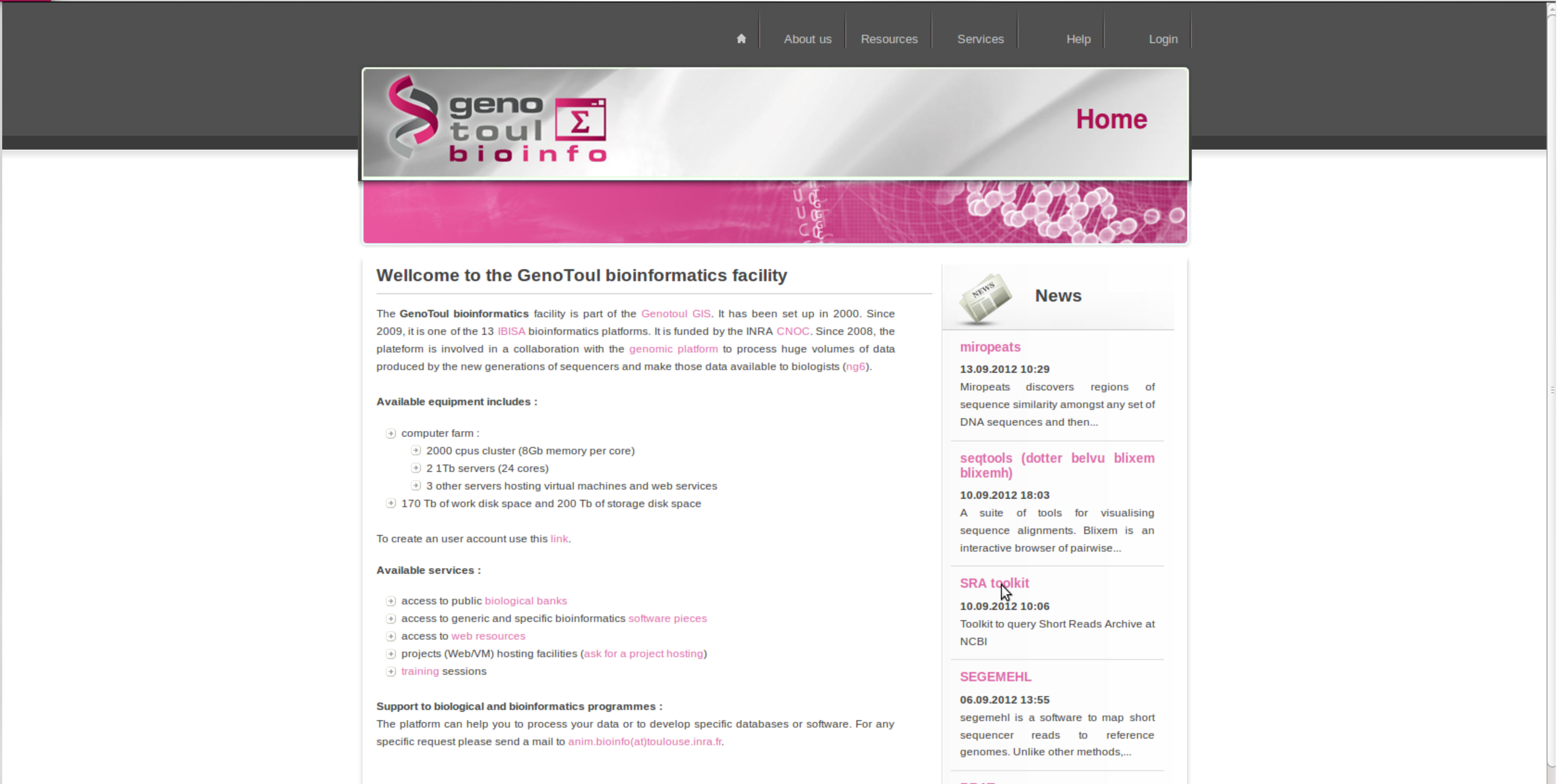


/usr/local/bioinfo/src	Bioinformatics Software
/bank/	International genomics Databanks
/home/	User configuration files (ONLY) (100 MB user quota)
/work/	HPC computational disk space (TEMPORARY) (1 TB user quota)
/save/	User disk space (with BACKUP) (250 GB user quota)

Plateforme Bioinformatique Midi-Pyrénées

Genotoul Bioinfo

http://bioinfo.genotoul.fr



Wellcome to the GenoToul bioinformatics facility

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. It is funded by the INRA CNOC. Since 2008, the plateform is involved in a collaboration with the genomic platform to process huge volumes of data produced by the new generations of sequencers and make those data available to biologists (ng6).

Available equipment includes :

- computer farm :
  - 2000 cpus cluster (8Gb memory per core)
  - 2 1Tb servers (24 cores)
  - 3 other servers hosting virtual machines and web services
- 170 Tb of work disk space and 200 Tb of storage disk space

To create an user account use this link.

Available services :

- access to public biological banks
- access to generic and specific bioinformatics software pieces
- access to web resources
- projects (Web/VM) hosting facilities (ask for a project hosting)
- training sessions

Support to biological and bioinformatics programmes :
The platform can help you to process your data or to develop specific databases or software. For any specific request please send a mail to anim.bioinfo(at)toulouse.inra.fr.

News

miropeats

13.09.2012 10:29

Miropeats discovers regions of sequence similarity amongst any set of DNA sequences and then...

seqtools (dotter belvu blixem blixemh)

10.09.2012 18:03

A suite of tools for visualising sequence alignments. Blixem is an interactive browser of pairwise...

SRA toolkit

10.09.2012 10:06

Toolkit to query Short Reads Archive at NCBI

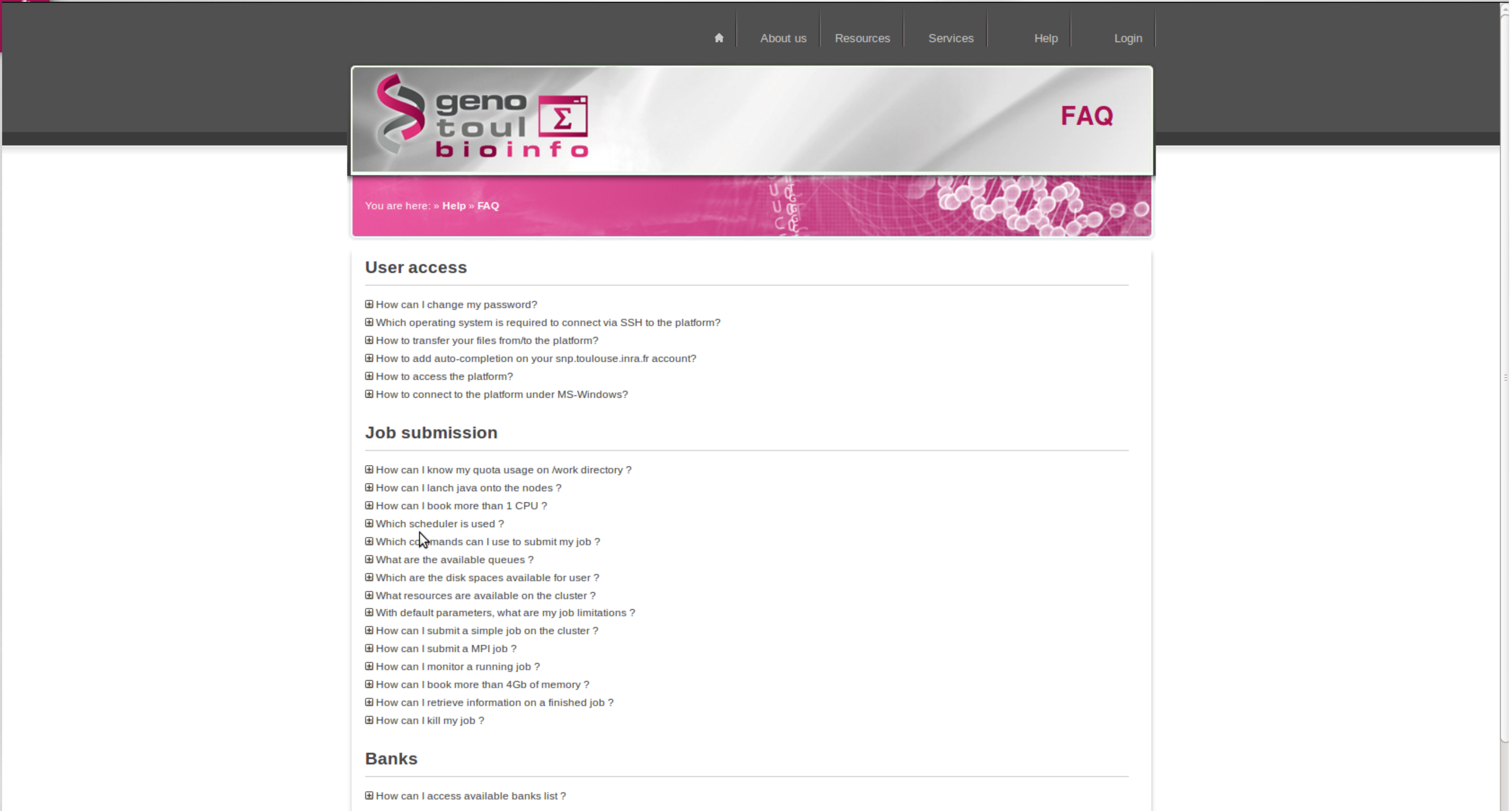
SEGEMEHL

06.09.2012 13:55

segemehl is a software to map short sequencer reads to reference genomes. Unlike other methods,...

Genotoul Bioinfo

Questions=> support.genopole@toulouse.inra.fr



FAQ

You are here: » **Help** » **FAQ**

User access

- ⊞ How can I change my password?
- ⊞ Which operating system is required to connect via SSH to the platform?
- ⊞ How to transfer your files from/to the platform?
- ⊞ How to add auto-completion on your snp.toulouse.inra.fr account?
- ⊞ How to access the platform?
- ⊞ How to connect to the platform under MS-Windows?

Job submission

- ⊞ How can I know my quota usage on /work directory ?
- ⊞ How can I lanch java onto the nodes ?
- ⊞ How can I book more than 1 CPU ?
- ⊞ Which scheduler is used ?
- ⊞ Which commands can I use to submit my job ?
- ⊞ What are the available queues ?
- ⊞ Which are the disk spaces available for user ?
- ⊞ What resources are available on the cluster ?
- ⊞ With default parameters, what are my job limitations ?
- ⊞ How can I submit a simple job on the cluster ?
- ⊞ How can I submit a MPI job ?
- ⊞ How can I monitor a running job ?
- ⊞ How can I book more than 4Gb of memory ?
- ⊞ How can I retrieve information on a finished job ?
- ⊞ How can I kill my job ?

Banks

- ⊞ How can I access available banks list ?

Linux

Introduction to Linux

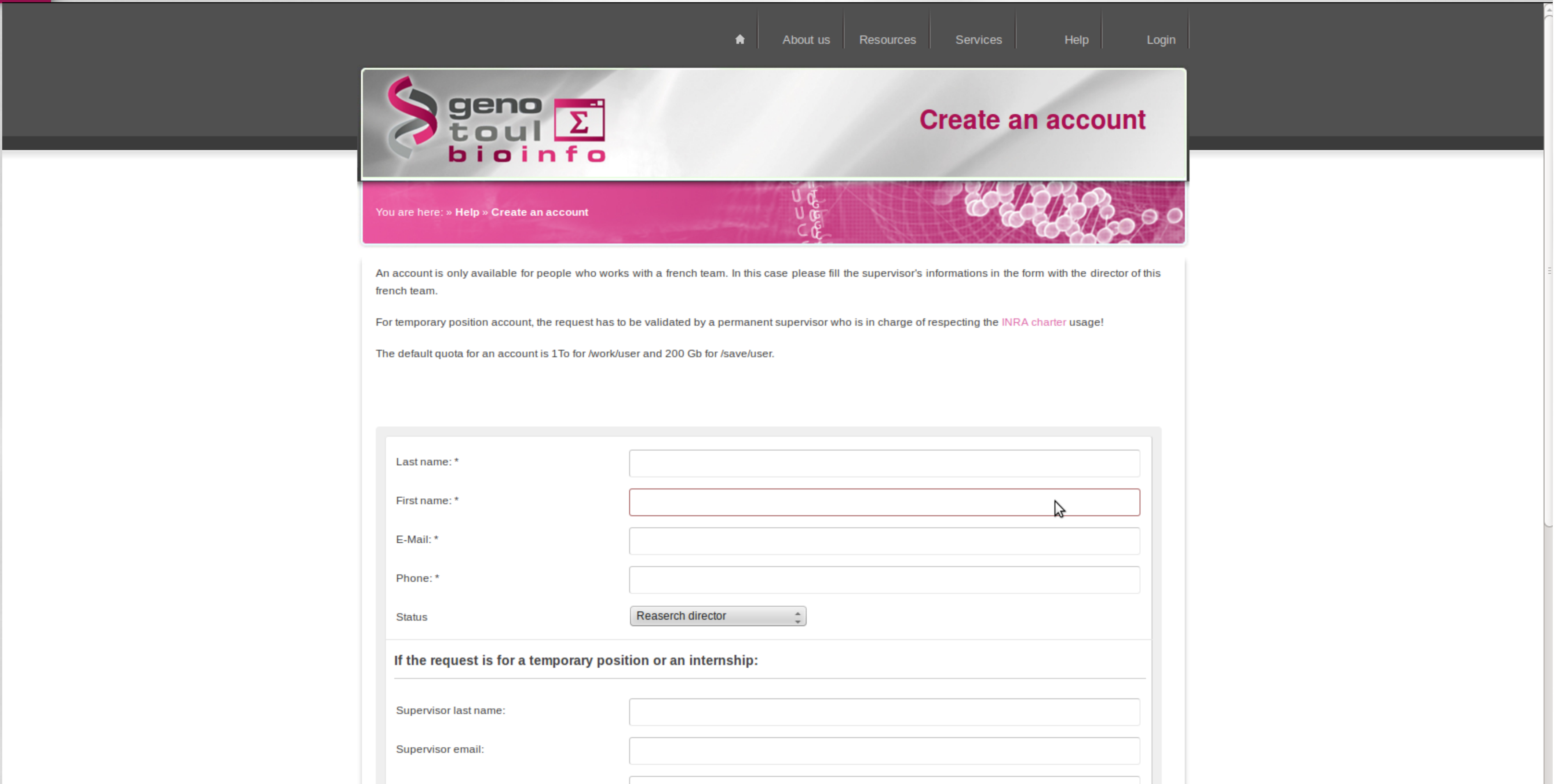
GNU-Linux : Unix-like operating system

- 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

Plateforme Bioinformatique Midi-Pyrénées

Linux

Asking for an account



You are here: » Help » Create an account

An account is only available for people who works with a french team. In this case please fill the supervisor's informations in the form with the director of this french team.

For temporary position account, the request has to be validated by a permanent supervisor who is in charge of respecting the INRA charter usage!

The default quota for an account is 1To for /work/user and 200 Gb for /save/user.

Last name: *	
First name: *	
E-Mail: *	
Phone: *	
Status	Reaserch director

If the request is for a temporary position or an internship:

Supervisor last name:	
Supervisor email:	

Linux

Linux account

Access to a work environnement

- 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

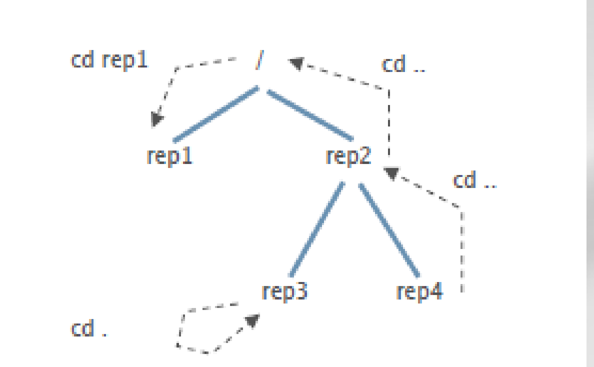
Tree structure

« / » root directory

« ~ » home directory (user)

« . » current directory

« .. » parent directory



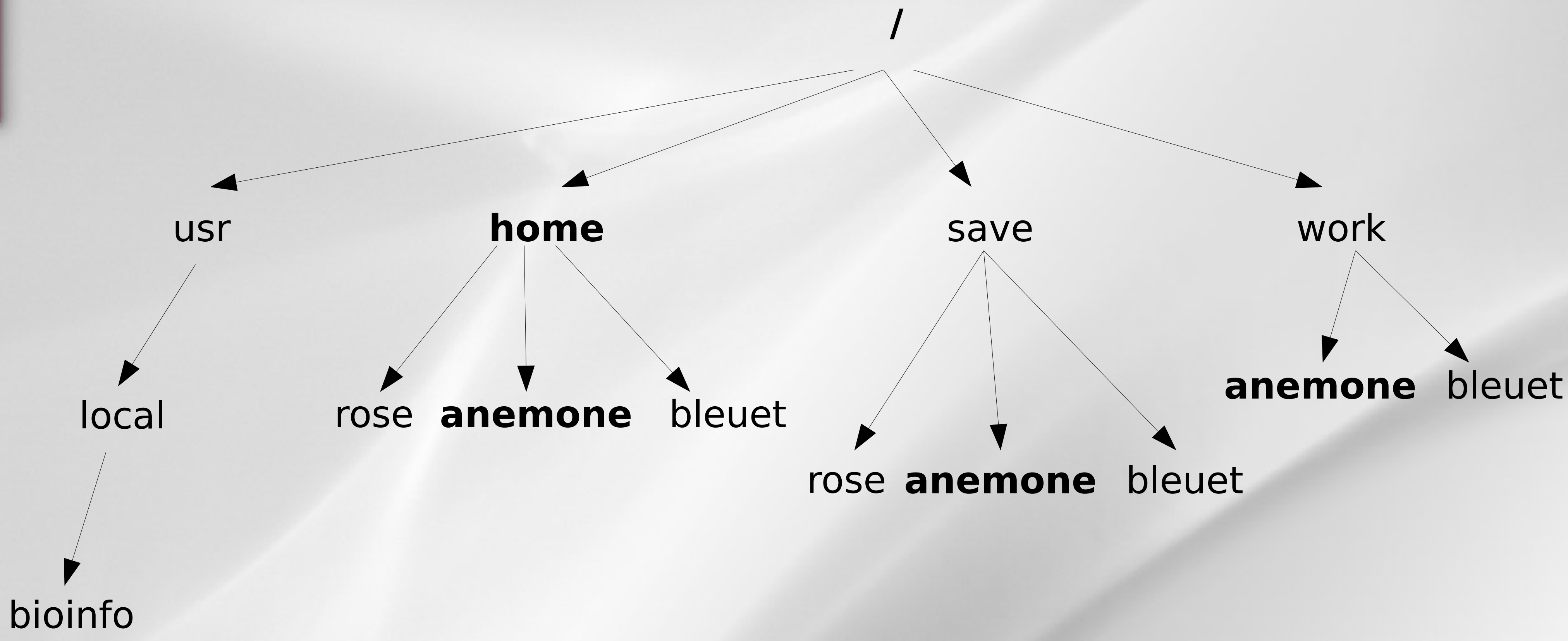
cd [nom_répertoire] : Change directory

Absolute path : `/home/bleuet`

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

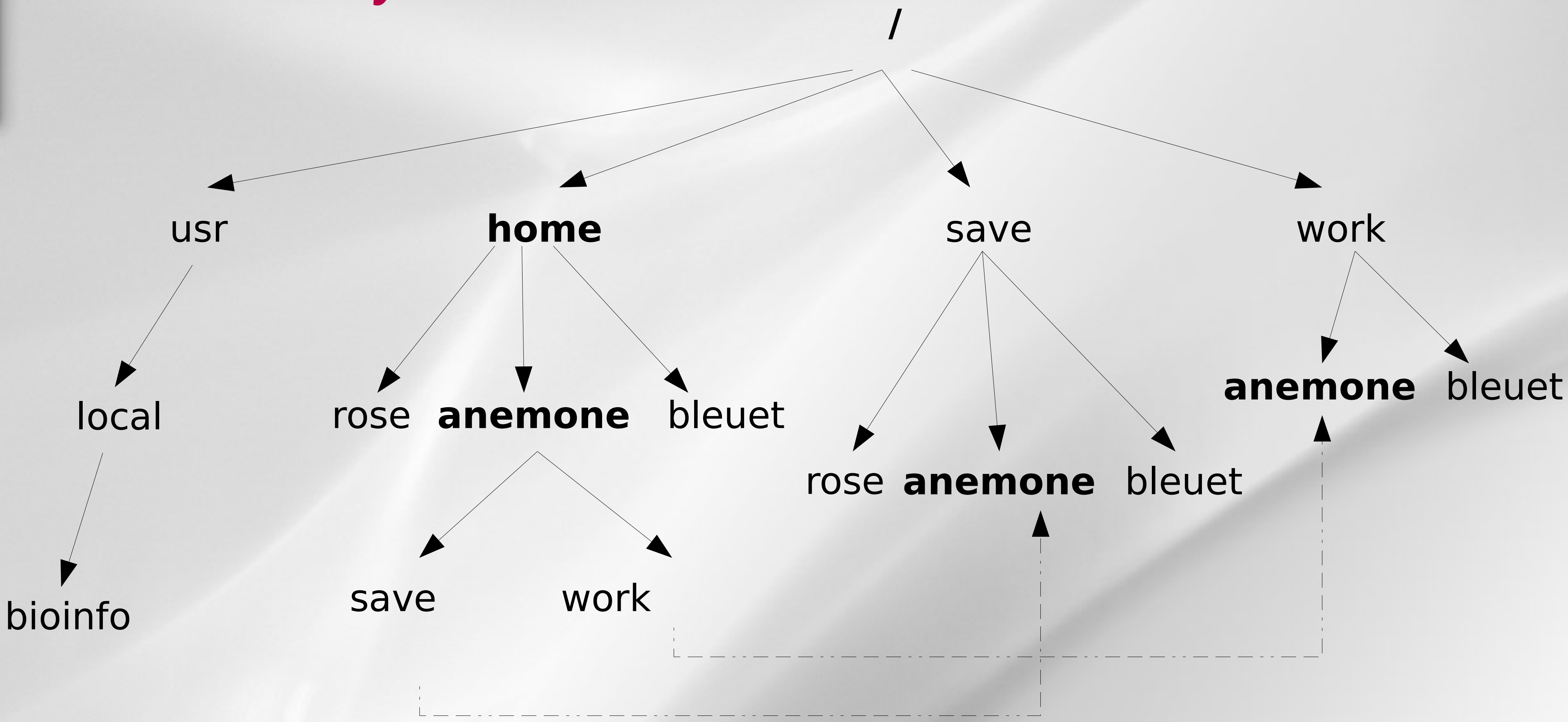
The tree structure

Notion of « Root »



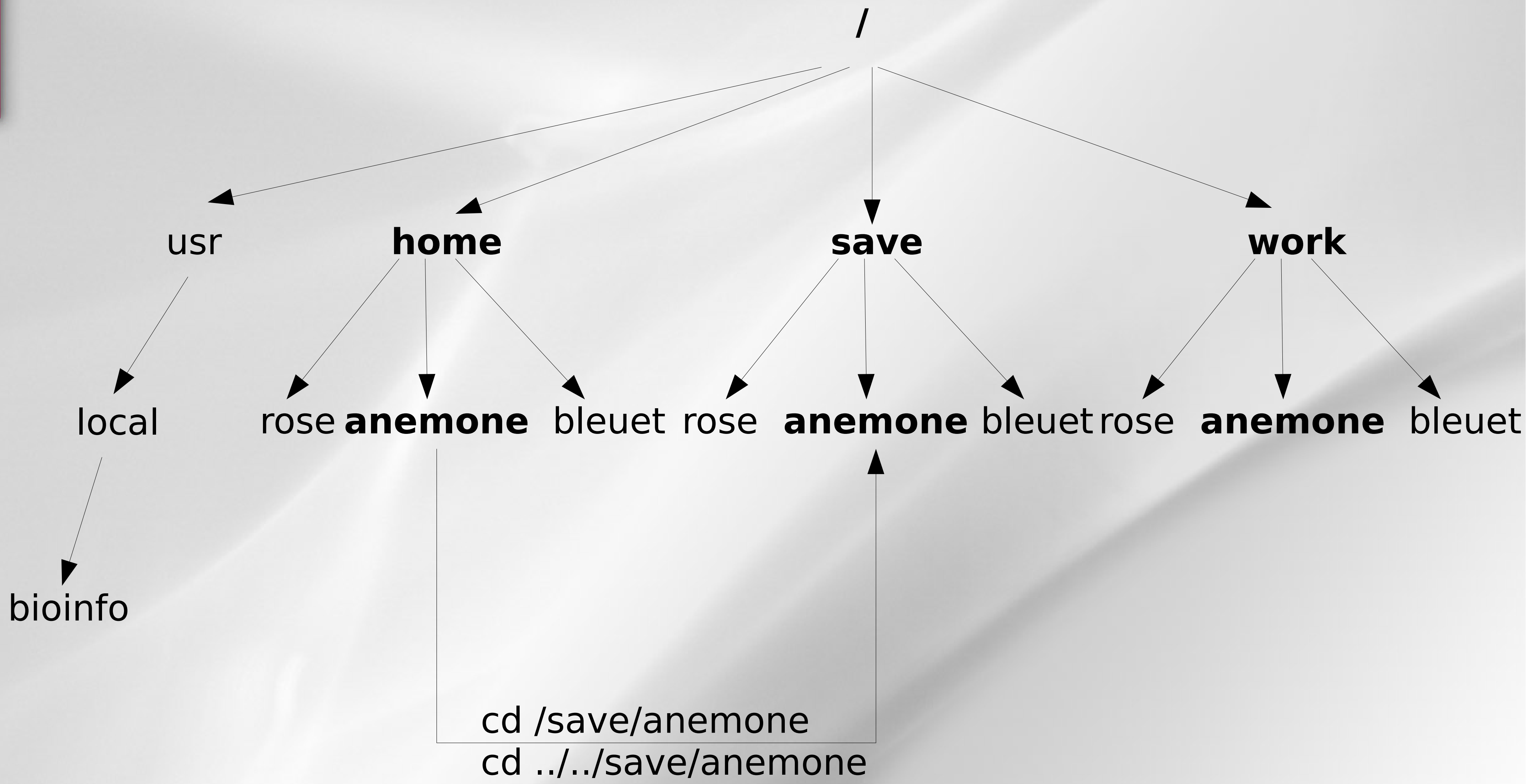
The tree structure

Notion of «symbolic links»



The tree structure

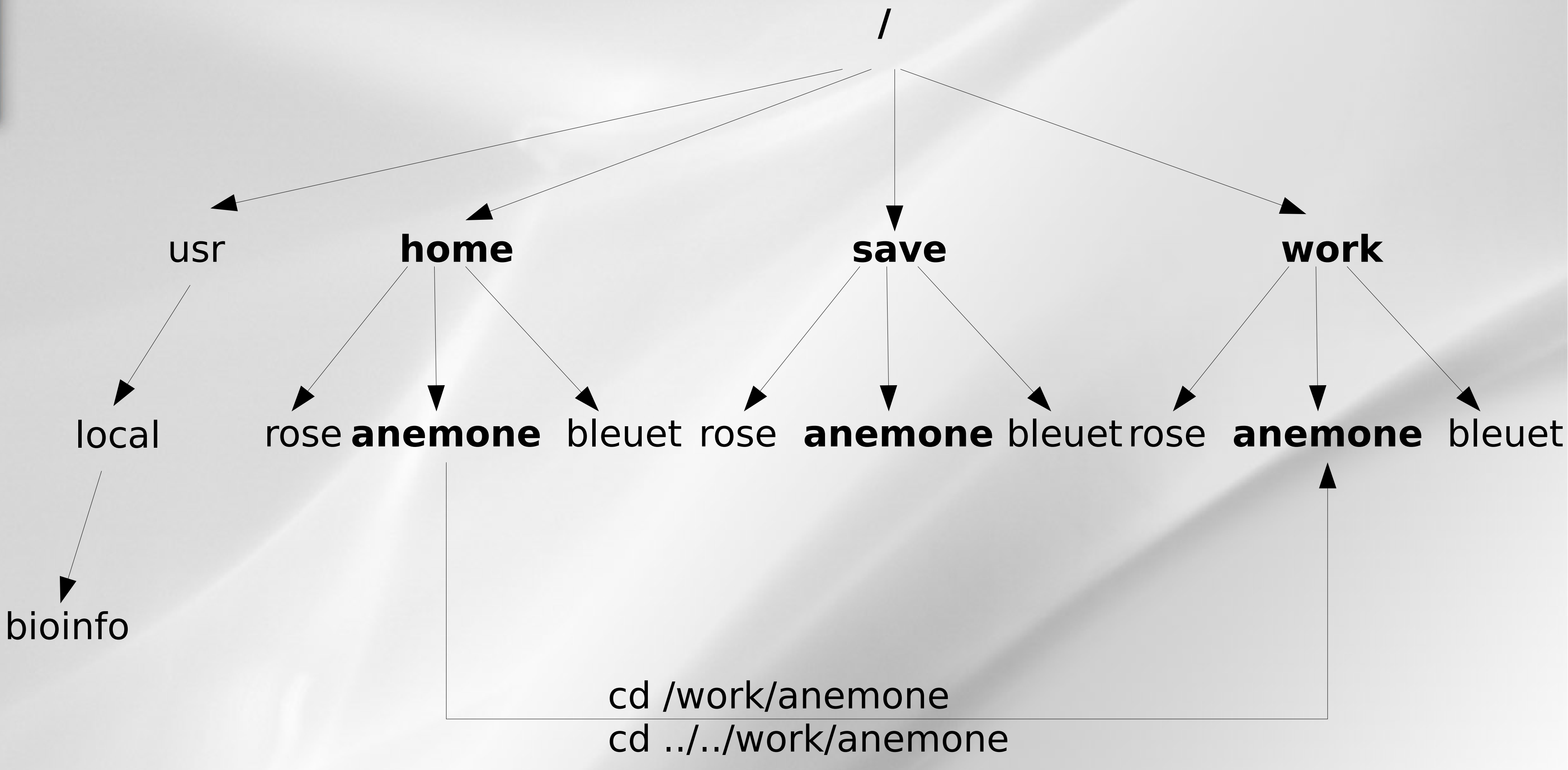
Navigation : examples



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

The tree structure

Navigation : examples



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

The commands : 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 commands

command_name -- help

man command_name

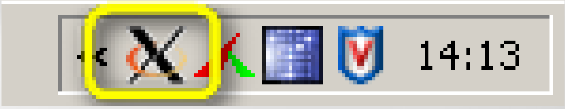
```
#ls --help
#blastall -help

#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 genotoul ?

From Windows



- **Xming** (Windows graphic)
- **Putty** (Connection)

From Linux / Mac

- **ssh username@genotoul.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 genotoul 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

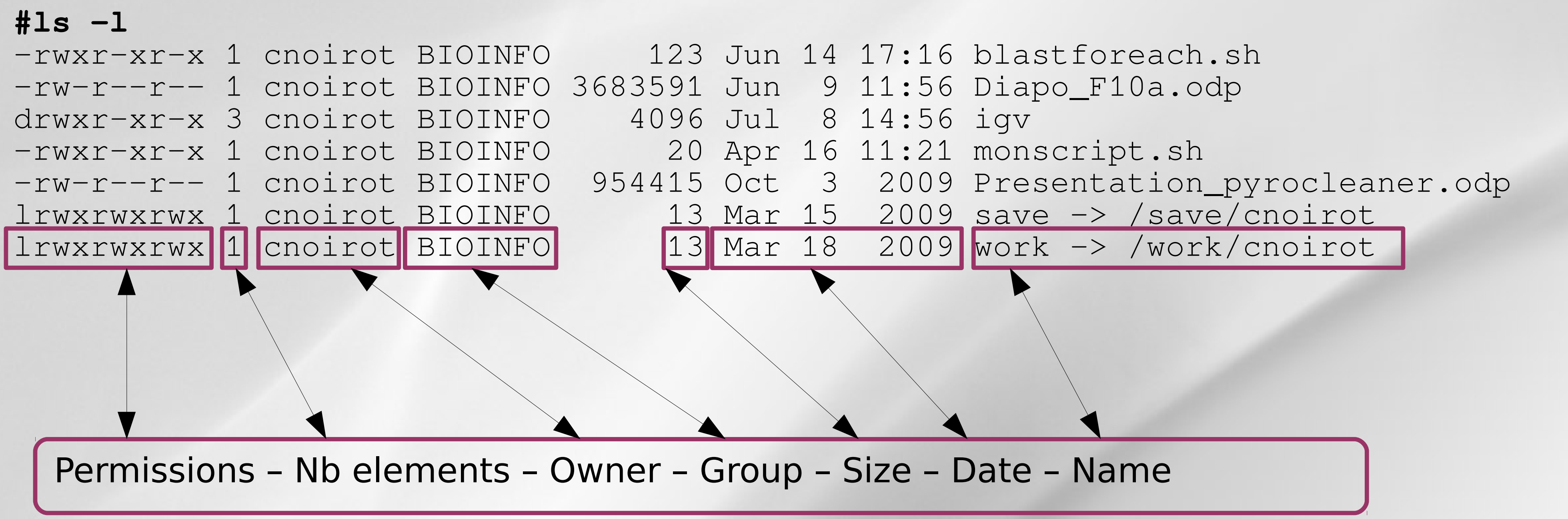
ls [-options] [dir_name]

- **-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)

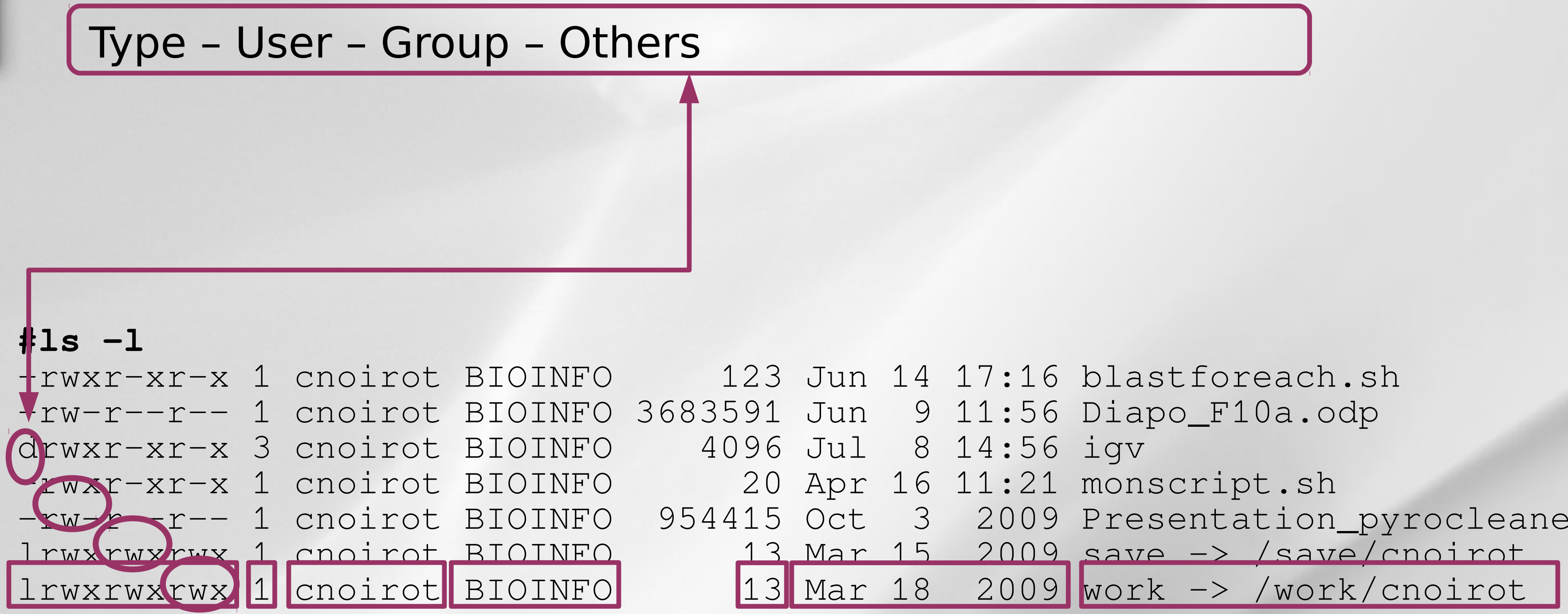


```
#ls -l
-rwxr-xr-x 1 cnoirot BIOINFO     123 Jun 14 17:16 blastforeach.sh
-rw-r--r-- 1 cnoirot BIOINFO 3683591 Jun  9 11:56 Diapo_F10a.odp
drwxr-xr-x 3 cnoirot BIOINFO    4096 Jul  8 14:56 igv
-rwxr-xr-x 1 cnoirot BIOINFO      20 Apr 16 11:21 monscript.sh
-rw-r--r-- 1 cnoirot BIOINFO  954415 Oct  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 : only directories

    #find . -type f : only files

    #find / -size +1000k : if size > 1Mo
```

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

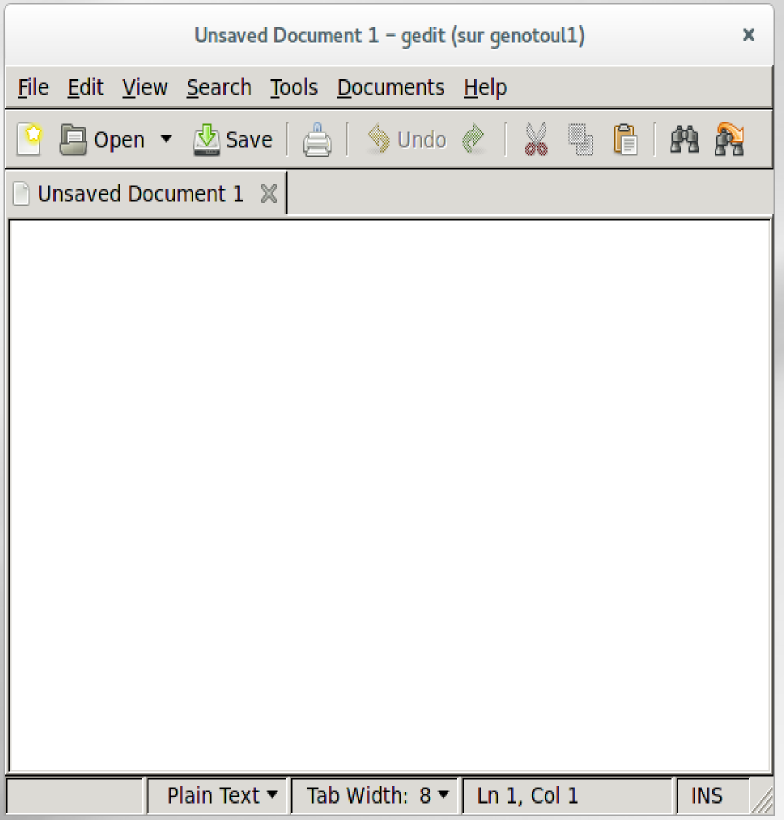
less file_name : display up and down

Plateforme Bioinformatique Midi-Pyrénées

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


Unsaved Document 1 – gedit (sur genotoul1)
File Edit View Search Tools Documents Help
Open
Save
Undo
Unsaved Document 1

Plain Text
Tab Width: 8
Ln 1, Col 1
INS

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

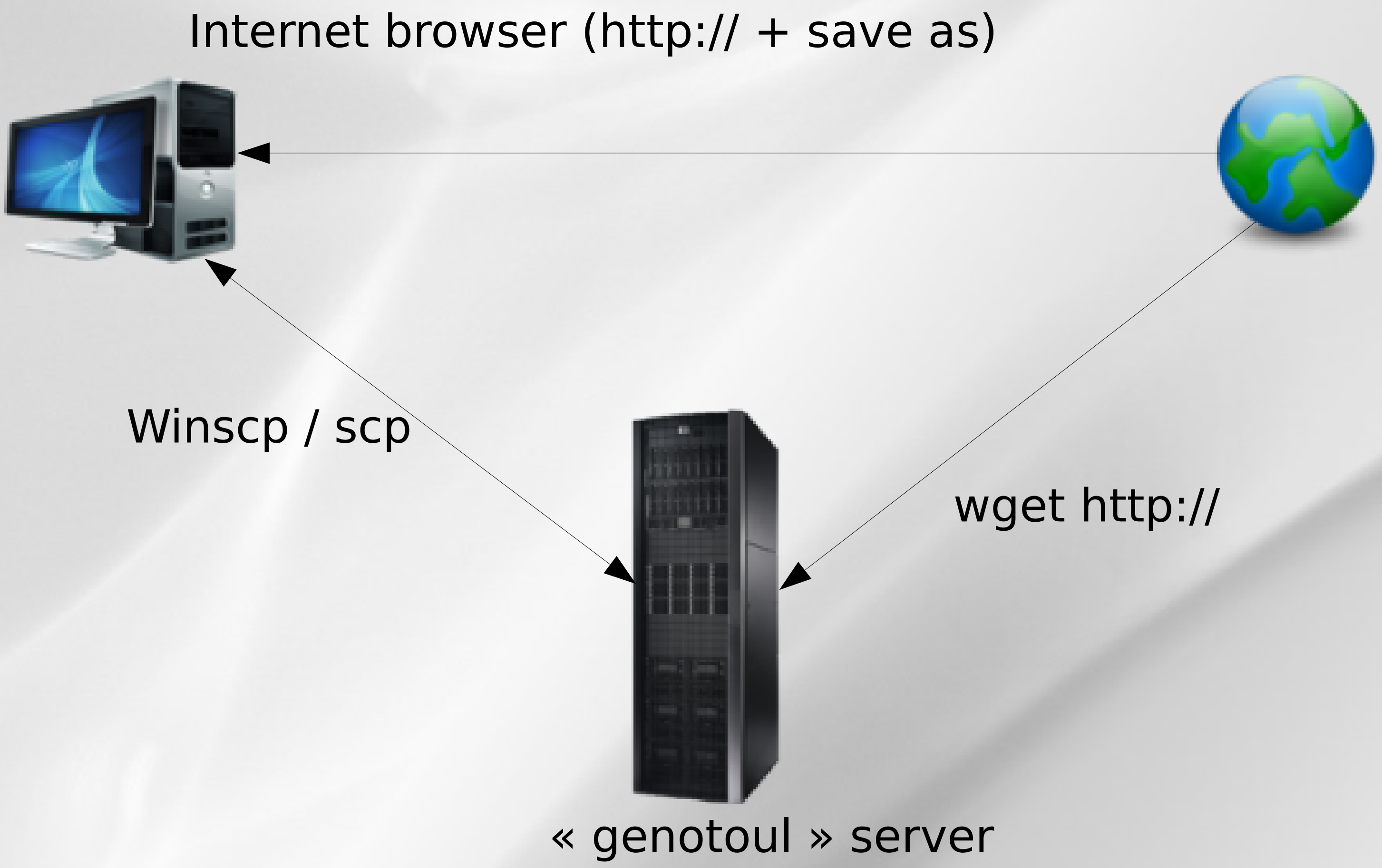
Plan

- Downloading / transferring
- Compressing / uncompressing

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

Downloading / transferring

Several possible cases



Plateforme Bioinformatique Midi-Pyrénées

Downloading / transferring

Directly from internet to genotoul

File download from Internet to « genotoul server »:

- Copy the URL of the file to download

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

Downloading / transferring

Transfer between genotoul and desktop computer

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

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

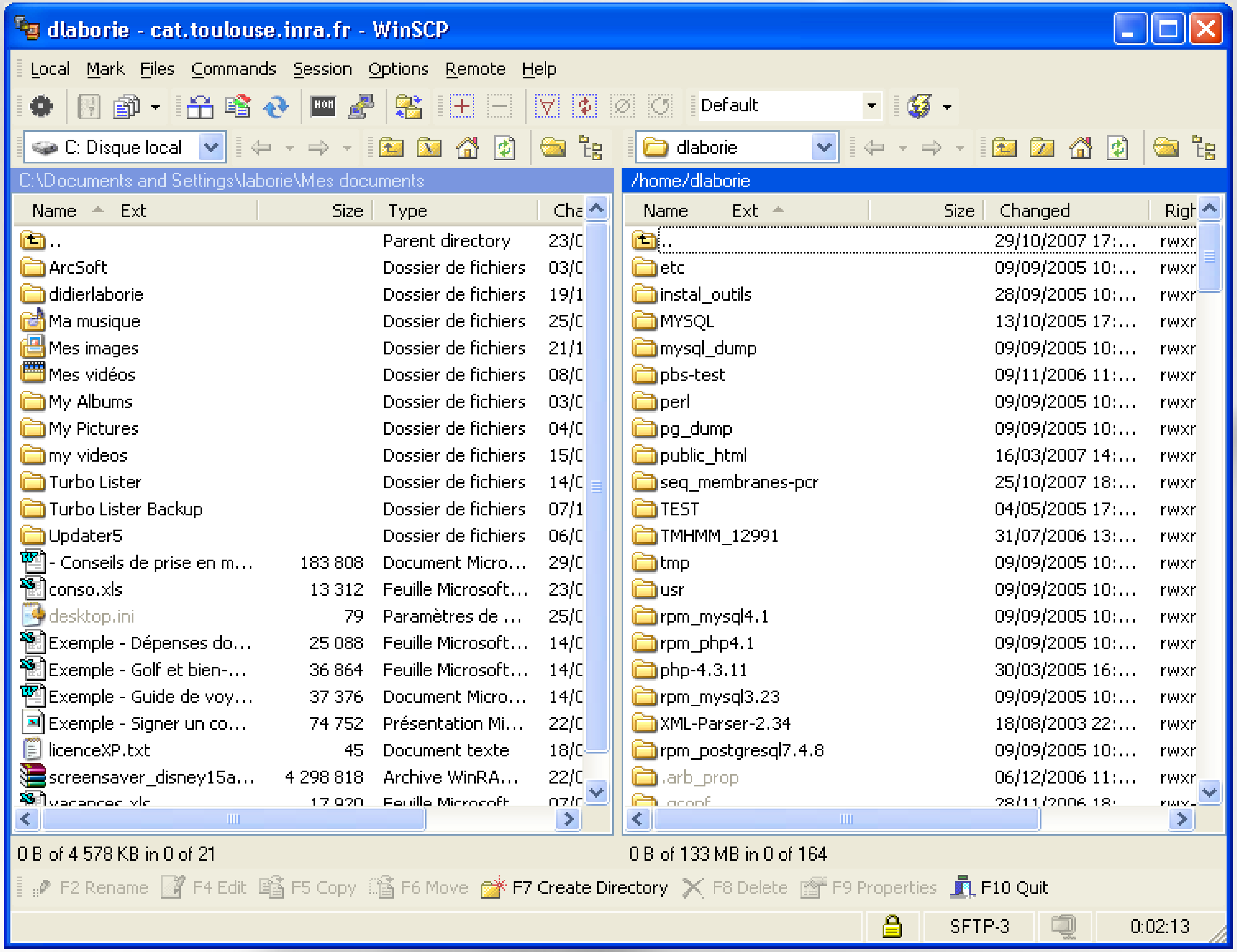
```
    #scp source_name bleuet@genotoul:destination_name
      (copy from desktop to "genotoul")
```

Plateforme Bioinformatique Midi-Pyrénées

geno toul bioinfo

Downloading / transferring

WinSCP / FileZilla : copy via graphical interface



dlaborie - cat.toulouse.inra.fr - WinSCP

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
```

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 : char. count

#wc -w file_name : words count

#wc -l file_name : lines count
```

Data extraction

Filters (1)

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

```
#cat nom_fic1 nom_fic2 > nom_fic_3
```

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

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

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

```
#tail -n 100 file_name (100 last 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 : **"<", ">"**, **"|"**, **">>"**

Plateforme Bioinformatique Midi-Pyrénées

geno toul bioinfo

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