FMI
Friedrich Miescher Institute
for Biomedical Research

Using Galaxy for the FMI NGS Pipeline

Hans-Rudolf Hotz (hrh@fmi.ch)

Friedrich Miescher Institute for Biomedical Research
Basel, Switzerland

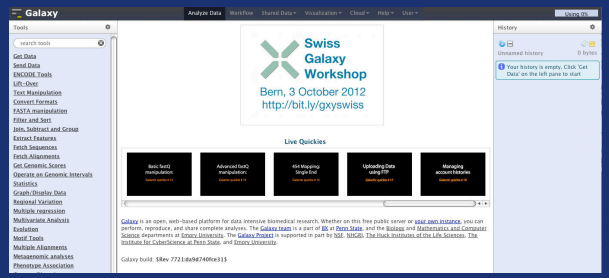
Galaxy at FMI is...

- in production for more than 5 years (3 years for NGS)
- used to reduce the workload of the core facilities
- used to bridge the gap

the “average” lab scientist

?

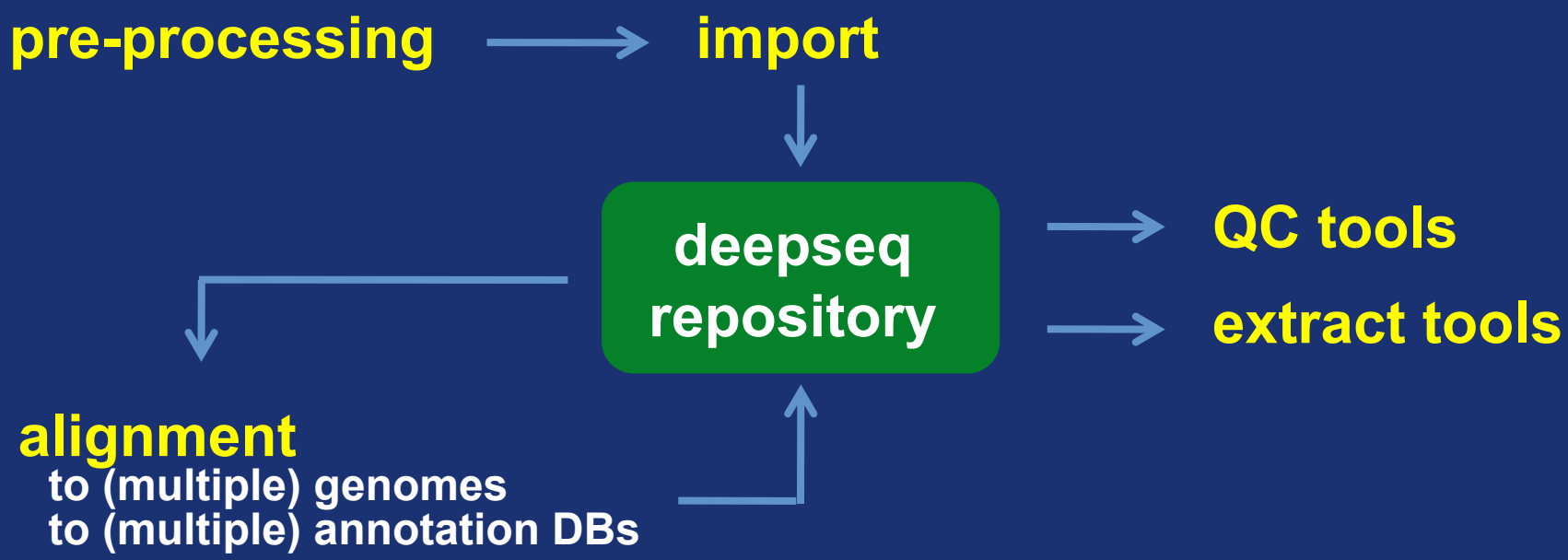
the “modern” lab scientist



FMI
Friedrich Miescher Institute for Biomedical Research

The (current) FMI Deep Sequencing Pipeline

(developed by Michael Stadler and Dimos Gaidatzis)



FMI

Friedrich Miescher Institute for Biomedical Research

The FMI Deep Sequencing is

....just a bunch of Perl scripts *(currently)*

→ which can be easily added to Galaxy

....just a simple file system *(currently)*

....which cannot be added to Galaxy.
(Galaxy uses its own data directory)

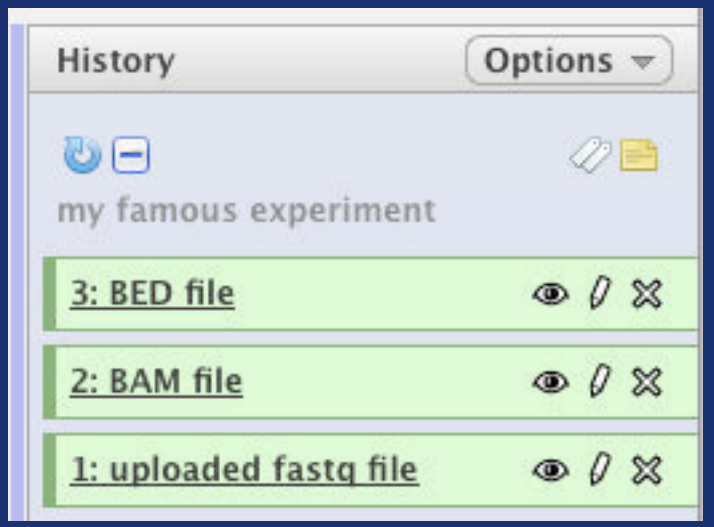
→ we don't have to, we just have to give Galaxy access to the file system
(without using "Data Libraries")

FMI
Friedrich Miescher Institute
for Biomedical Research

a simple NGS workflow

- your famous aligner

- your famous extract tool



FMI
Friedrich Miescher Institute
for Biomedical Research

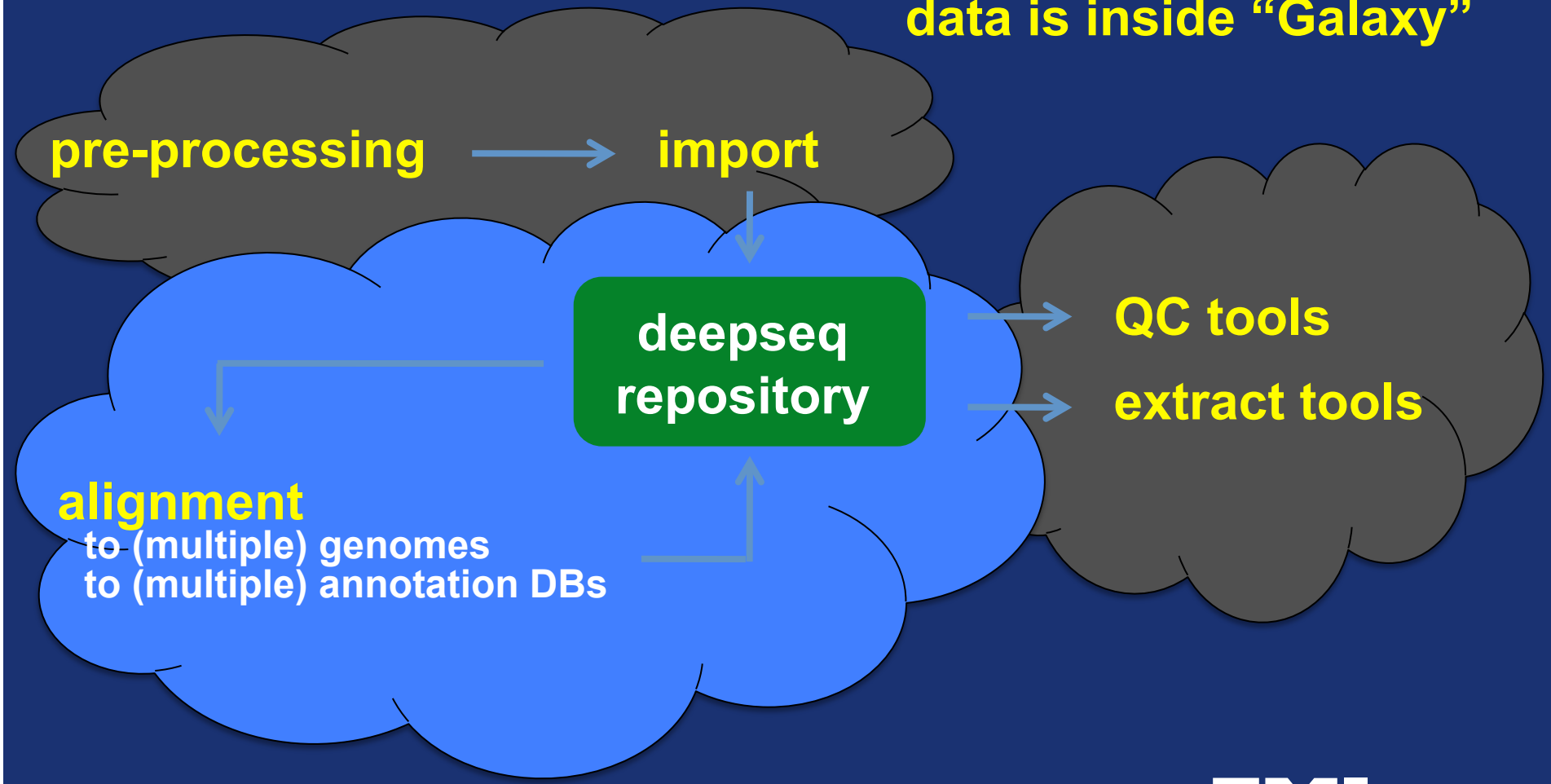
a simple NGS workflow

- do you need the first result (ie the alignment) as a new history item?
- does your tool require a Galaxy history item as input?

- the ‘famous aligner’ has a wrapper storing the BAM file in the central NGS repository and creating just a log file for Galaxy
- your ‘famous extract tool’ knows the location of the NGS repository

FMI
Friedrich Miescher Institute
for Biomedical Research

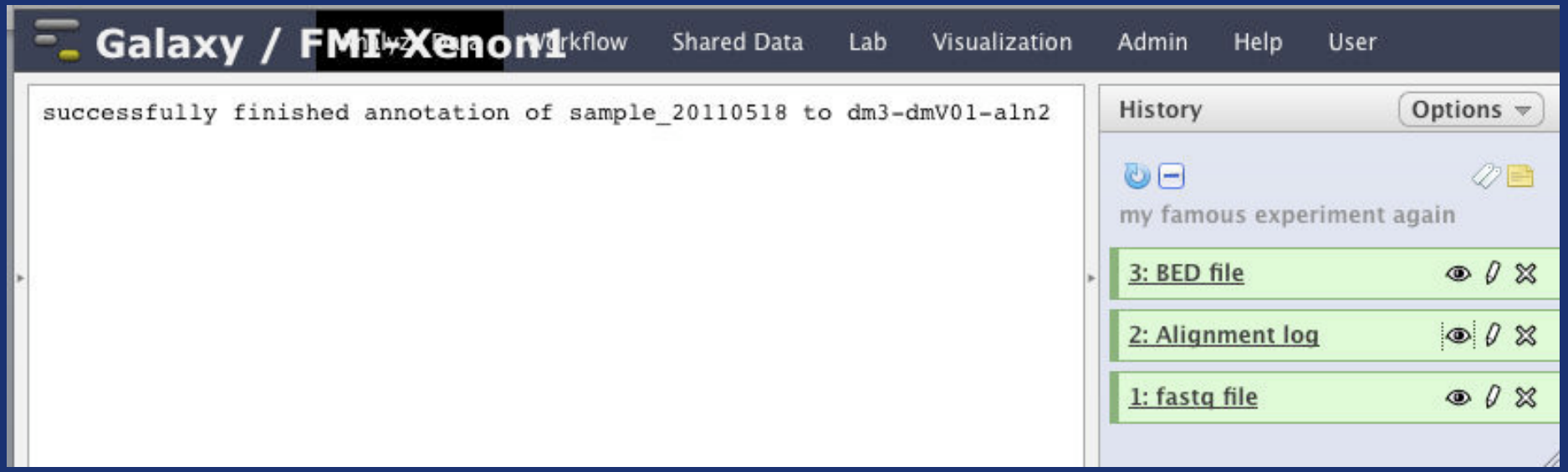
The FMI Deep Sequencing Pipeline



FMI
Friedrich Miescher Institute
for Biomedical Research

storing data outside of Galaxy

makes it easier to share with non-Galaxy users



FMI
Friedrich Miescher Institute
for Biomedical Research

makes it easier to share with non-Galaxy users

```
successfully finished annotation of
sampleId_20110518 to dm3-dmV01-aln2
```

and now the command line geek can do

```
[geek@xenon1 ~]$ extractData.pl -f -s p -m
100 -i mySampleId_20110518 dm3-dmV01-aln2
genome |frag2bed.pl -t -q -U - | head -5
track name='mySampleId_20110518'
chr2L   10493   10528   sq39319 1       +
chr2L   10736   10764   sq74484 1       +
chr2L   11442   11477   sq1340  1       +
chr2L   13799   13834   sq84955 1       +
[geek@xenon1 ~]$
```

makes it easier to share with non-Galaxy users

command line

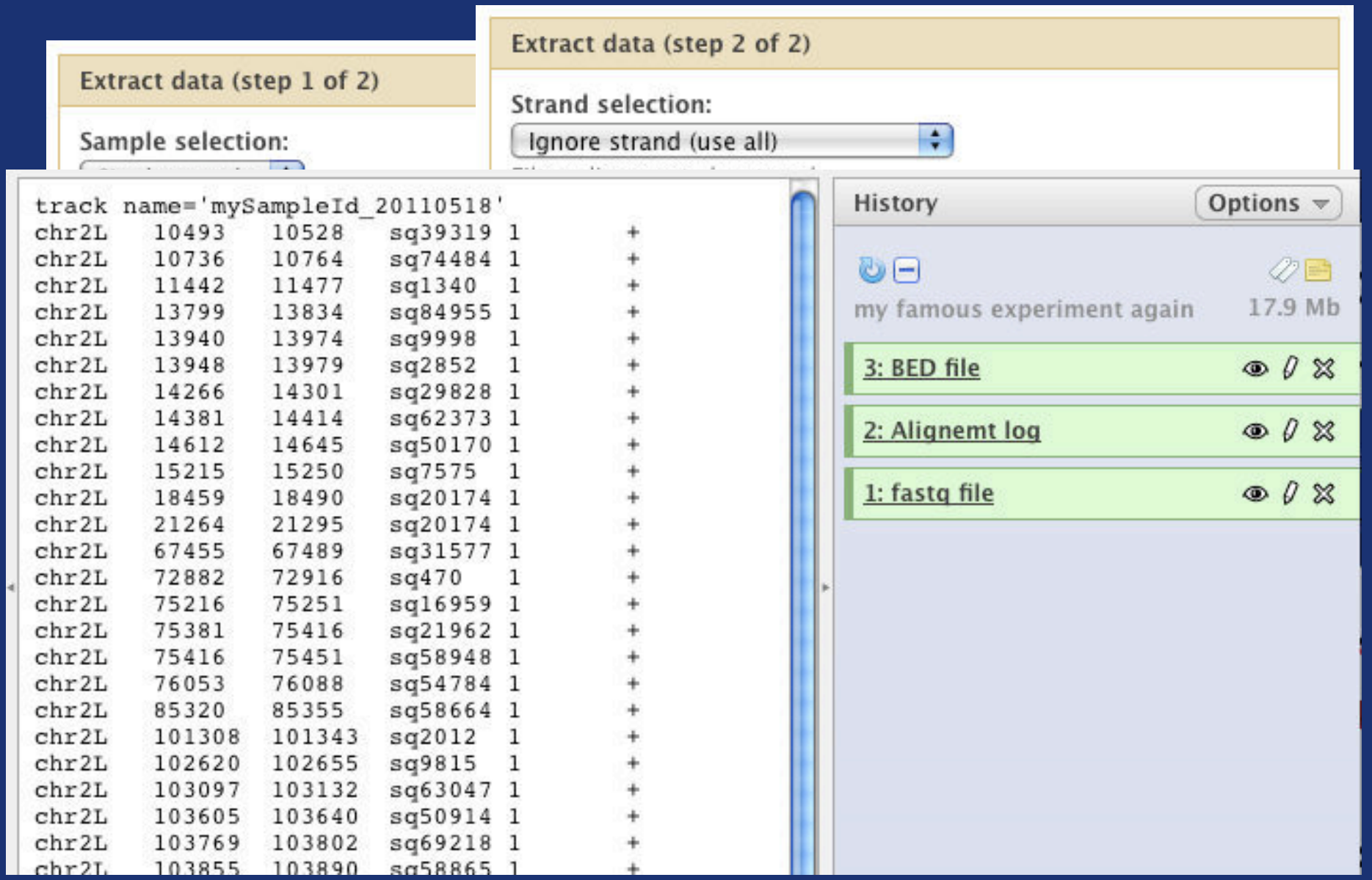
```
extractData.pl -f -s p -m 100 -i
mySampleId_20110518 dm3-dmV01-aln2 genome |
frag2bed.pl -t -q -U -
```

Galaxy tool definition file

```
 #elif ($summary.mode=="bed")#extractData.pl
-f $strand $maxhits $ignCnts
$sampleSelect.sampleId $genome-$annot-aln2
genome | frag2bed.pl -t -q $summary.ucsc -
> $output
```

FMI
Friedrich Miescher Institute
for Biomedical Research

and doing the same in Galaxy



Acknowledgment

Michael Stadler Lukas Burger

Dimos Gaidatzis

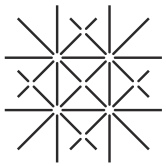
Tim Roloff Stefan Grzybek

Anita Lerch

Thomas Übermeier Jan Welker

....and all the people from the "Galaxy"

Swiss Institute of Bioinformatics



Friedrich Miescher Institute for Biomedical Research