Biogenouest

ifb
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

EMBRC FRANCE
CENTRE NATIONAL DE RESSOURCES BIOLOGIQUES MARINES

OCEANOMICS

IDEALG
seaweed for the future

ABiMS

30/05/2016

Galaxy

Initiation

Misharl Monsoor
Loraine Guéguen

Credit to Gildas Le Corguillé - V1.07

UPMC SORBONNE UNIVERSITÉS

cnrs

INTRODUCTION / PROBLEMATIC

Introduction

- Setup TP
  - http://galaxy.sb-roscoff.fr
  - Account
    - login@sb-roscoff.fr
    - *********

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted_hits.bam  junctions.bed    qsub.err  unmapped.bam
deletions.bed      logs             qsub.out
insertions.bed     prep_reads.info  tmp
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$
[login@n0 mapping
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-ro
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v12                                          .R1-2.fq
-GTF ../input/par
-num-threads 8

[login@n0 mapping                                          t.qsub
Your job 5338969
[login@n0 mapping
accepted_hits.bar                                          l.bam
deletions.bed          logs                qsub.out
insertions.bed         prep_reads.info     tmp
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

Introduction

Select your level:

Introduction

Level I

« I want to know the gene expression »

Introduction

Level 2

« I want to map my reads on a reference genome and count them »

Level 3

« I want to launch the tools tophat2 and cufflinks.
I have fastq files and my genome in fasta and gtf. »

Introduction

Level 4

« I want 1TB for my project. I will launch tophat2 through SSH on the cluster in multi-thread mode.

Next I want to submit the bam file to my genome with cufflinks.

Except that, I will manage :P”

Introduction

Level 5

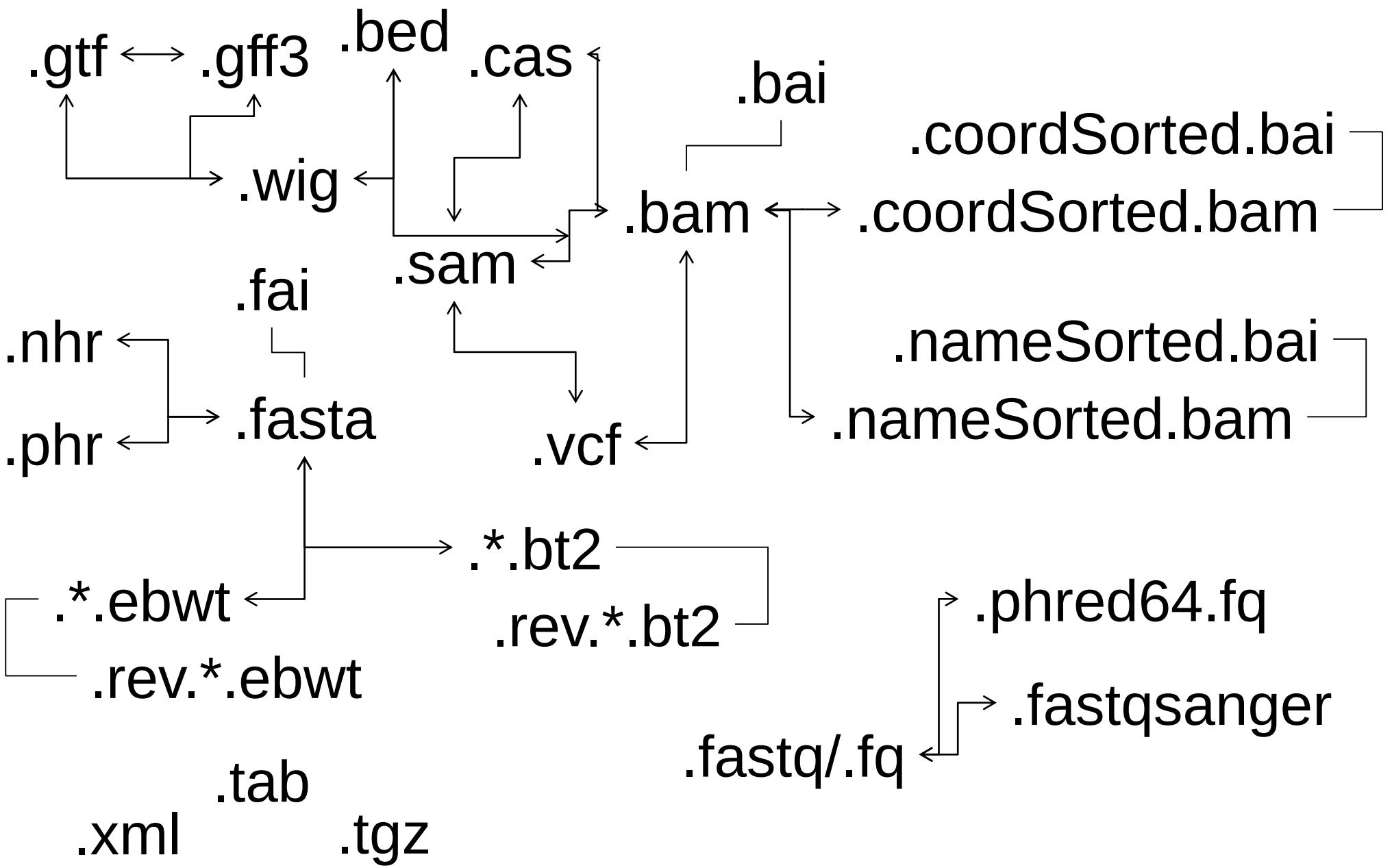
« I have a bunch of cool tools!
But I'm the only one who can launch them.

Comments? »

Introduction

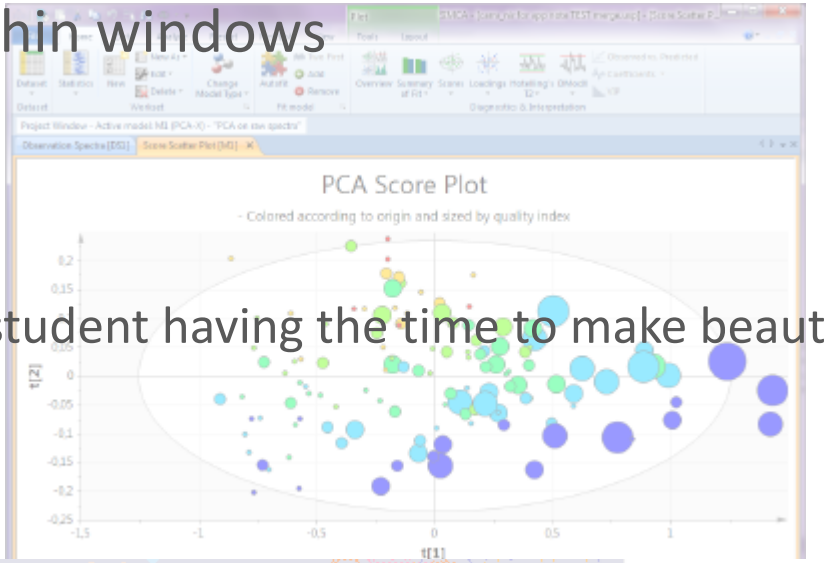
.gtf .gff3 .bed .cas .bai .coordSorted.bai .wig .bam .coordSorted.bam .sam .fai .nhr .nameSorted.bai .fasta .nameSorted.bam .phr .vcf .*.bt2 .*.ebwt .rev.*.bt2 .phred64.fq .rev.*.ebwt .fastqsanger .fastq/.fq .tab .xml .tgz

Introduction

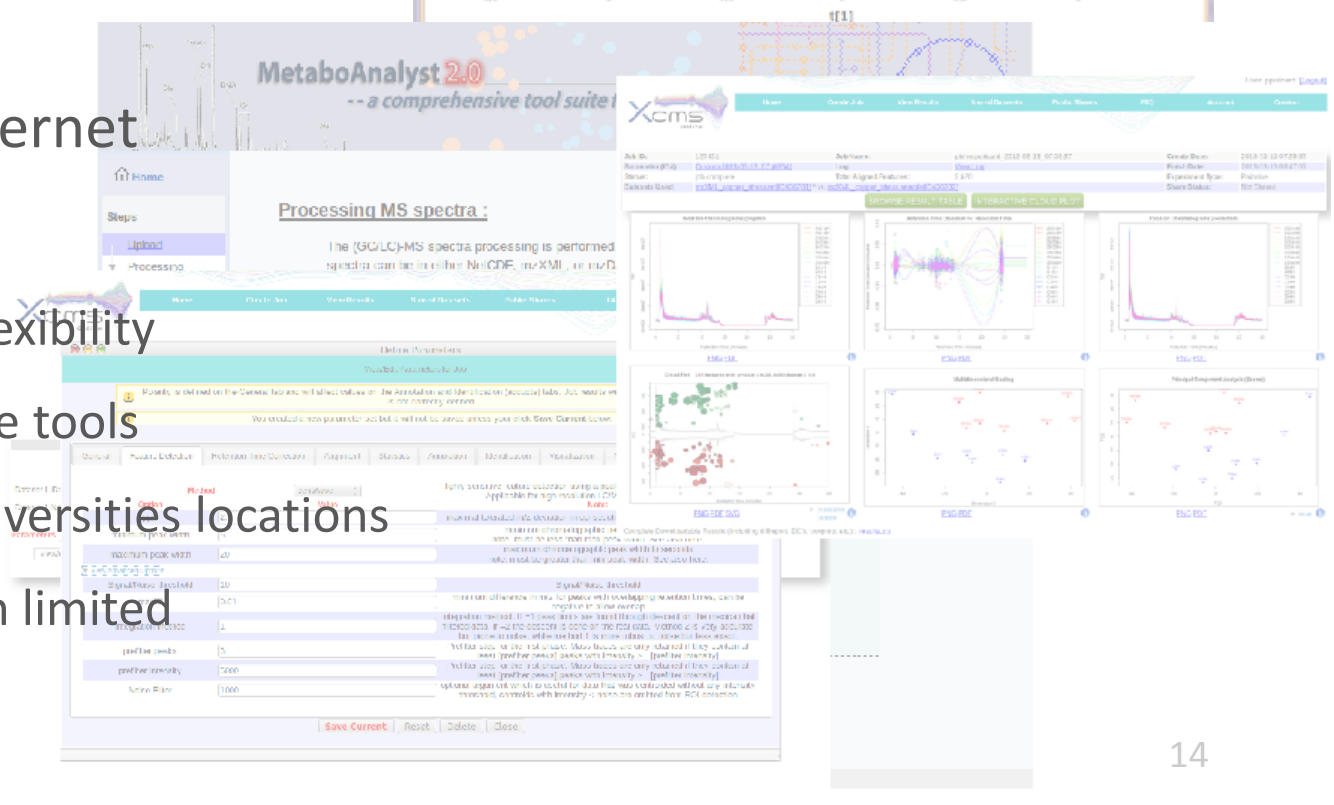


Introduction

- Graphical interface click-button tools within windows
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – don't count on it ! Have you ever seen a PhD student having the time to make beautiful green buttons?
  - – paying for it!



- Tools available on the internet
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – A small part of the available tools
  - – distributed on different universities locations
  - – the submission size is often limited
  - – must not be paranoid



Introduction

- Command line tools
  - + represent almost the majority of scientific tools
  - + good parameters completeness
  - + can be executed on high performance computers
  - + g33ks love it, since automatable, workflowsable, ...
  - – minimum linux knowledge is required
  - – cruel lack of ergonomics

INTRODUCTION / GALAXY


Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation
GC-MS
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation
NMR
Preprocessing
Normalisation
Quality Control
Statistical Analysis
COMMON TOOLS
Data Handling
Text Manipulation
Filter and Sort
Join, Subtract and Group
xcms.xcmsSet
version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
hide
Execute
Authors Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton08@imperial.ac.uk and Christopher Conley cjconley@ucdavis.edu
If you use this tool, please cite: Smith,C.A. et al.(2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
For details about this tool, please go to http://www.bioconductor.org/packages/release/bioc/html/xcms.html
Galaxy integration ABIMS TEAM, Station biologique de Roscoff.
Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
History
search datasets
Sacuri Zip
19 shown
289.7 MB
19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)
18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata
17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv
16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv
15: xset.group.retcor.group.fillPeaks.RData
14: xset.group.retcor.group.Rplots.pdf
13: xset.group.retcor.group.RData
12: xset.group.retcor.BPCs_corrected.pdf
11:

Introduction / Galaxy

- Galaxy it's …
  - No need to execute a command line through a terminal
  - Programming or scripting skills are not required
  - Submission of jobs is transparent through a high performance computer cluster
  - Secure histories and data manager
  - A data and protocols sharing system
  - Tool-boxes of several bioinformatics fields
    - NGS
    - Chemistry
    - Metabolomics
    - Image analysis
    - Statistics
    - Etc …
  - A web-based interface

Introduction / Galaxy

Why Galaxy ?

- Accessibility
- Reproductibility
- Transparency

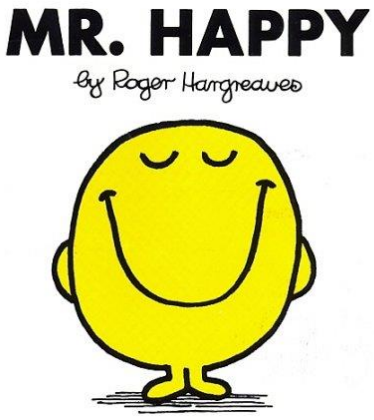
Introduction / Galaxy

MR. GEEK

```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```

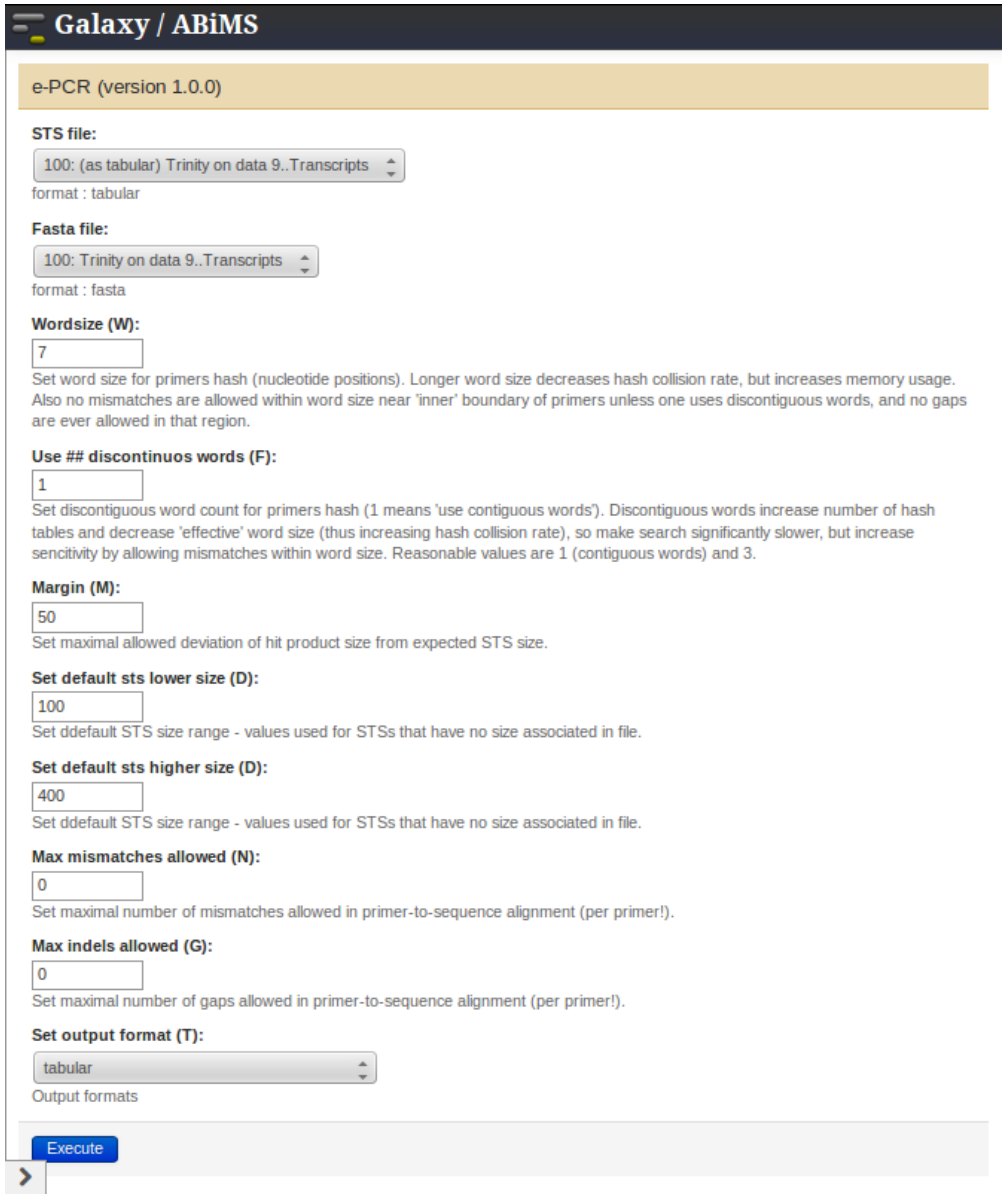
Introduction / Galaxy

```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/Ill1R1-1.fq ../input/Ill1R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```



Introduction / Galaxy

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
          -m ##     Margin (default 50)
          -w ##     Wordsize  (default 7)
          -n ##     Max mismatches allowed (default 0)
          -g ##     Max indels allowed (default 0)
          -f ##     Use ## discontiguos words, slow if
                              ##>1
          -o ##     Set output file
          -t ##     Set output format:
                    1 - classic, range (pos1..pos2)
                    2 - classic, midpoint
                    3 - tabular
                    4 - tabular with alignment in
comments
                              (slow)
          -d##-##   Set default size range
                              (default 100-350)
          -p +-     Turn hits postprocess on/off
          -v ##     Verbosity flags
          -a a|f    Use presize alignmens (only if
                              gaps>0), slow
                     a - Allways or f - as Fallback
          -x +-     Use 5'-end lowercase masking of
                              primers (default -)
          -u +-     Uppercase all primers (default -)

[...]
```

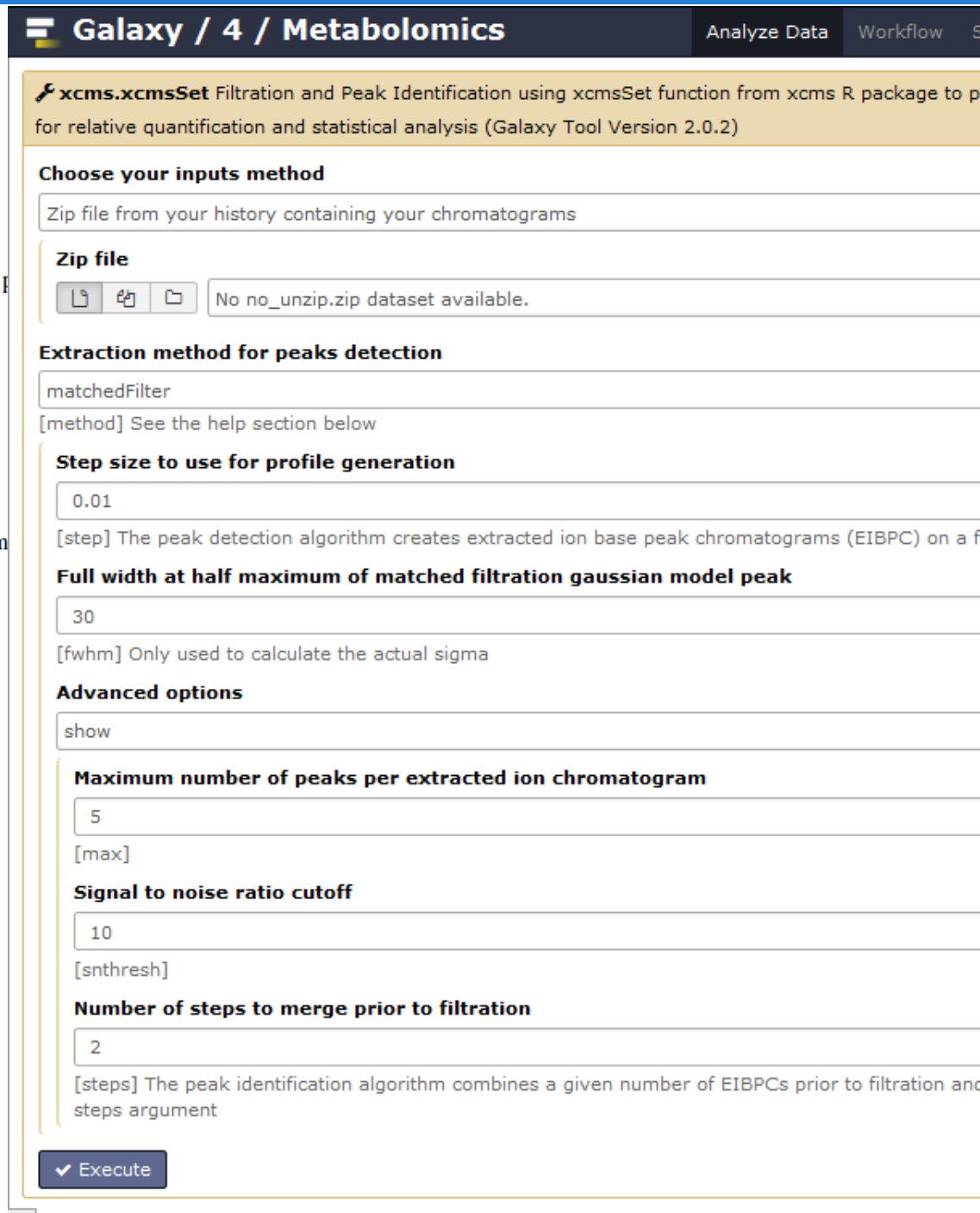


Introduction / Galaxy

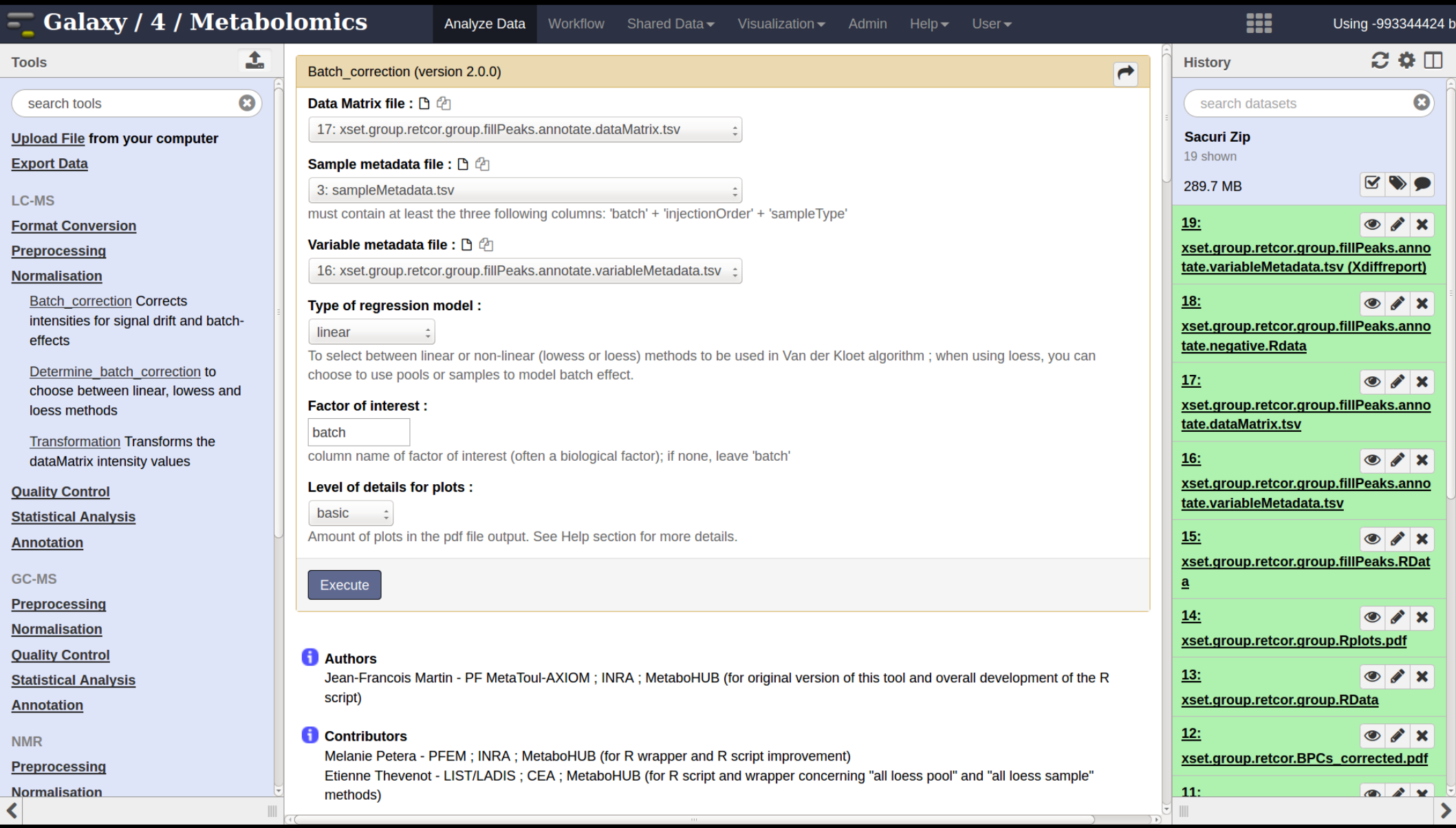
xcmsSet.matchedFilter(object, fwhm = 30, sigma = f

Arguments

`object`	`xcmsRaw` object
`fwhm`	full width at half maximum of matched filtration gaussian model p used to calculate the actual sigma, see below.
`sigma`	standard deviation (width) of matched filtration model peak
`max`	maximum number of peaks per extracted ion chromatogram
`snthresh`	signal to noise ratio cutoff
`step`	step size to use for profile generation
`steps`	number of steps to merge prior to filtration
`mzdiff`	minimum difference in m/z for peaks with overlapping retention tim
`index`	return indicies instead of values for m/z and retention times
`sleep`	number of seconds to pause between plotting peak finding cycles
`scanrange`	scan range to process

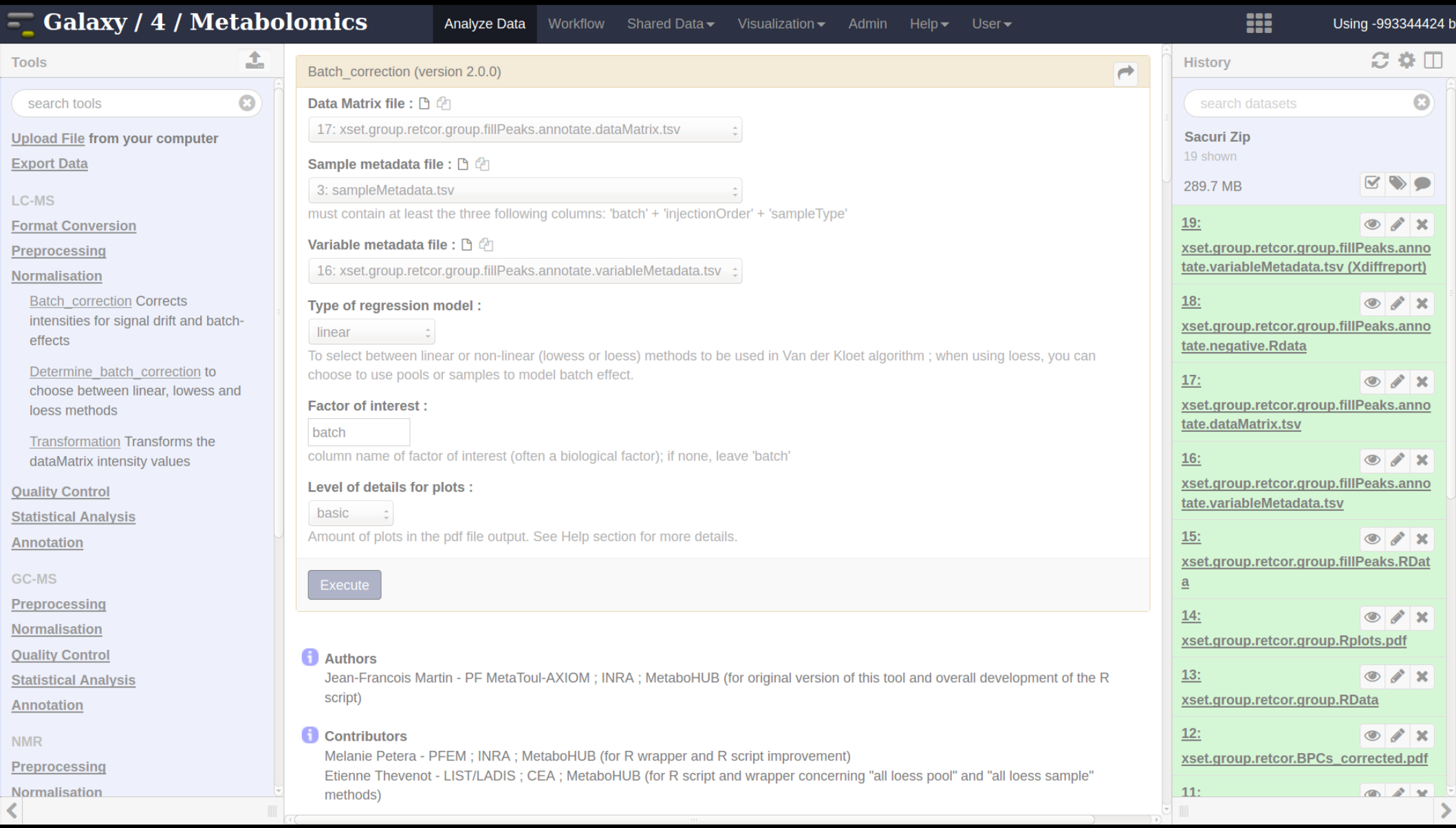


Galaxy interface



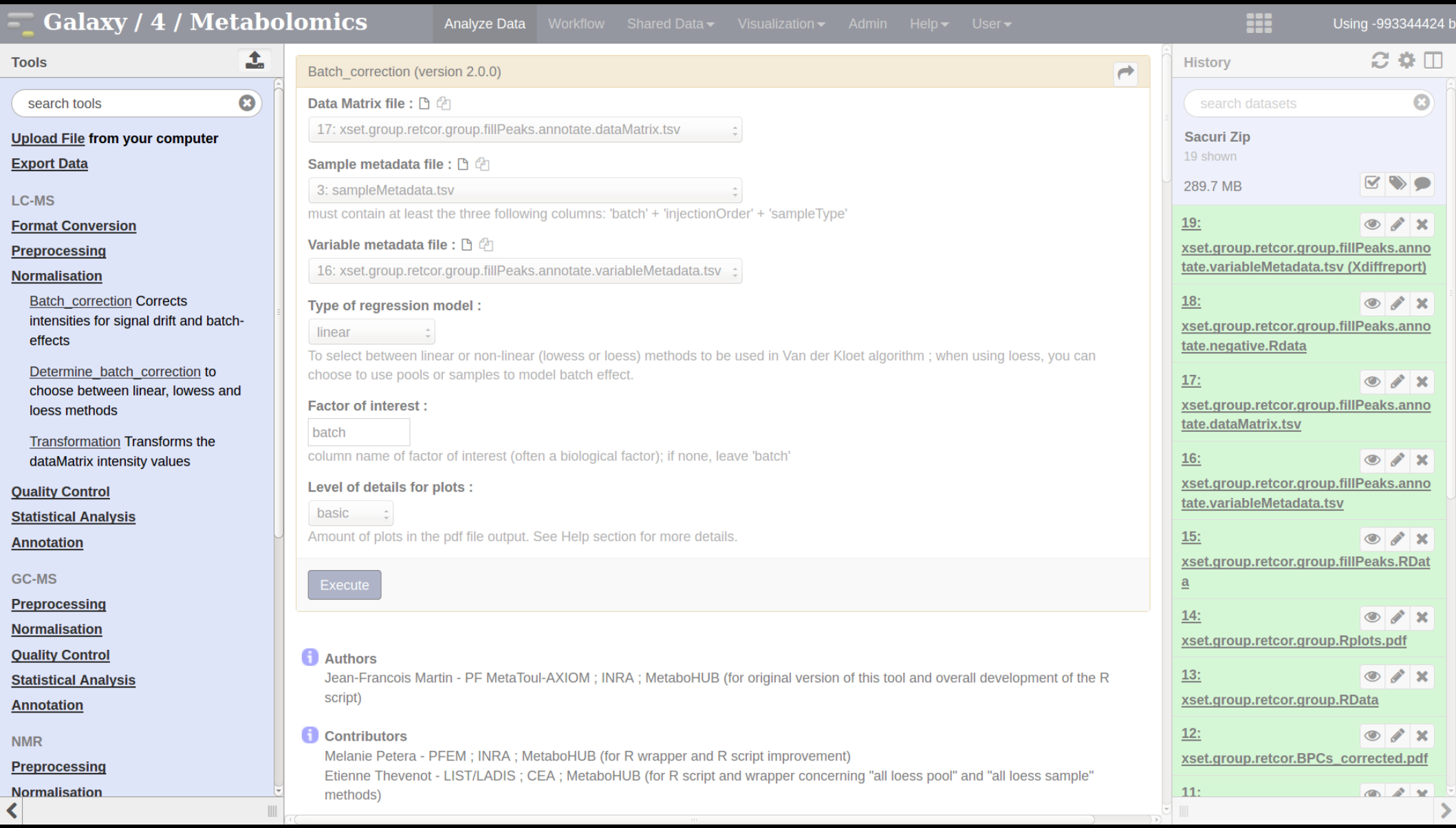
Galaxy interface

Menu



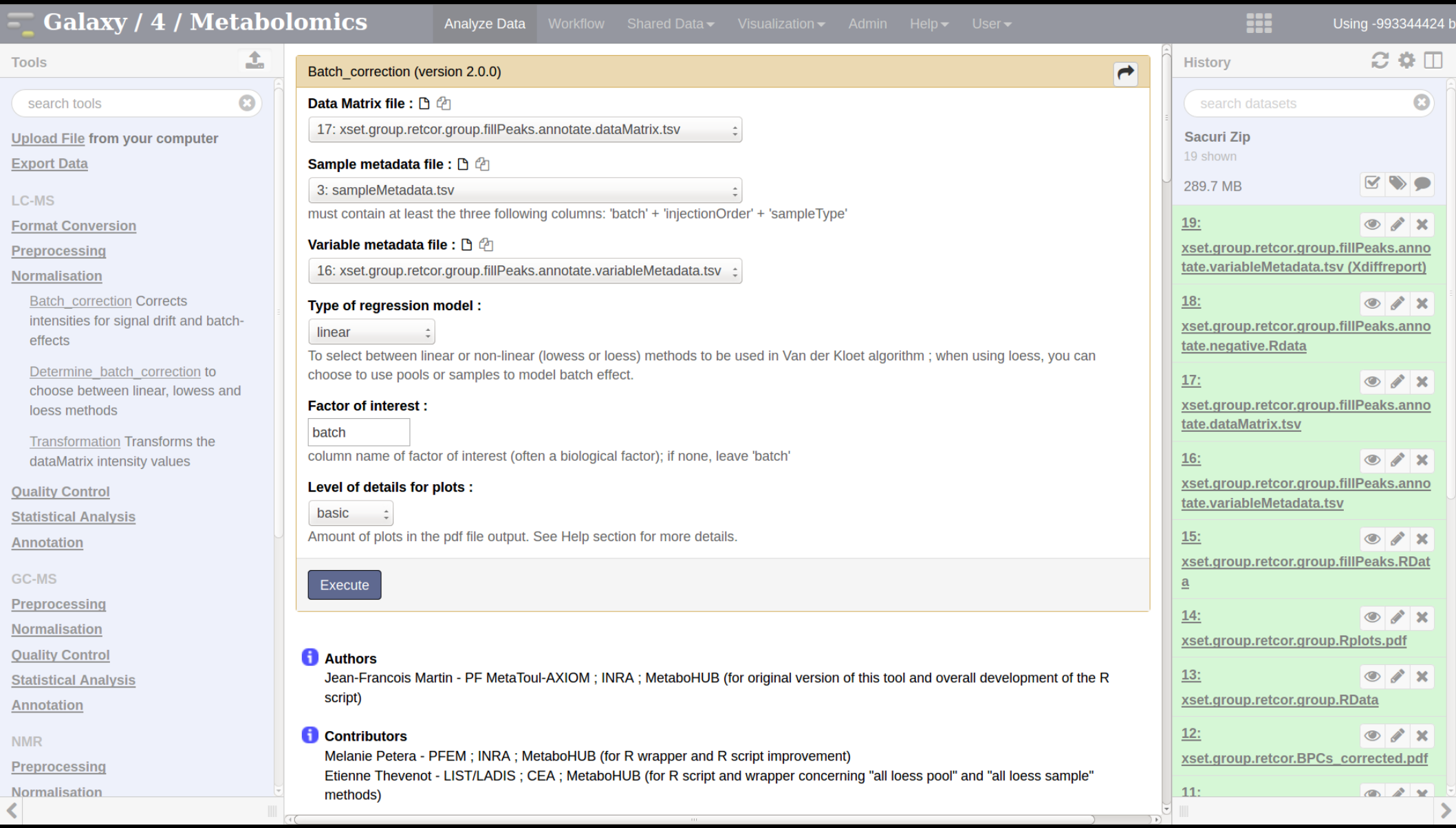
Galaxy interface

Tool list



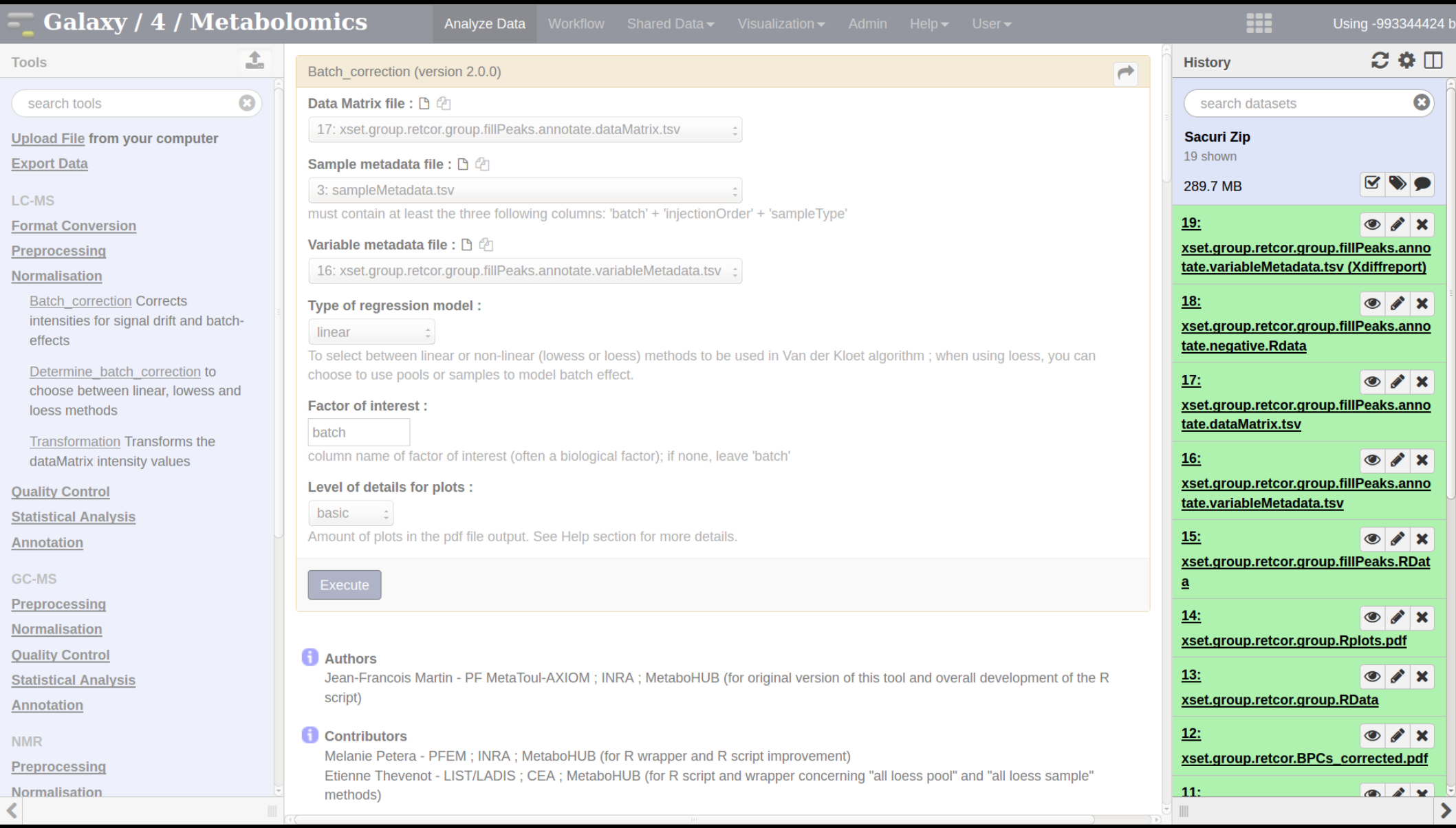
Galaxy interface

Web forms / visualization / diverse information



Galaxy interface

History



GET HELP

Get help


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 12%
Tools
search tools
Get Data
COMMON TOOLS
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Text Manipulation
Graphics
Statistics
EMBOSS 5 Suite
SEARCHING TOOLS
NCBI BLAST+
Diamond
Primer/Microsatellite
NGS TOOLS
NGS: BedTools
NGS: Mapping
NGS: Picard
NGS: QC and manipulation
NGS: SAM Tools
GATK Tools
GATK2 Tools
DNA-SEQ TOOLS
DNA-Seq tools
Welcome to galaxy.sb-roscoff.fr
Warning
17-05-16: For performance and maintenance reasons, we will disable the tool Export2Dir. You can always export your whole history in a file or download one file using the old fashion "floppy disk" icon. Export2Dir could return on day after a little refactoring?
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.
History
search datasets
sartools deseq2
7 shown
353.2 KB
7: SARTools DESeq2 R objects (.RData)
6: SARTools DESeq2 R log
5: SARTools DESeq2 figures
4: SARTools DESeq2 tables
3: SARTools DESeq2 report
2: targetT048.txt
1: t048.zip

Get help

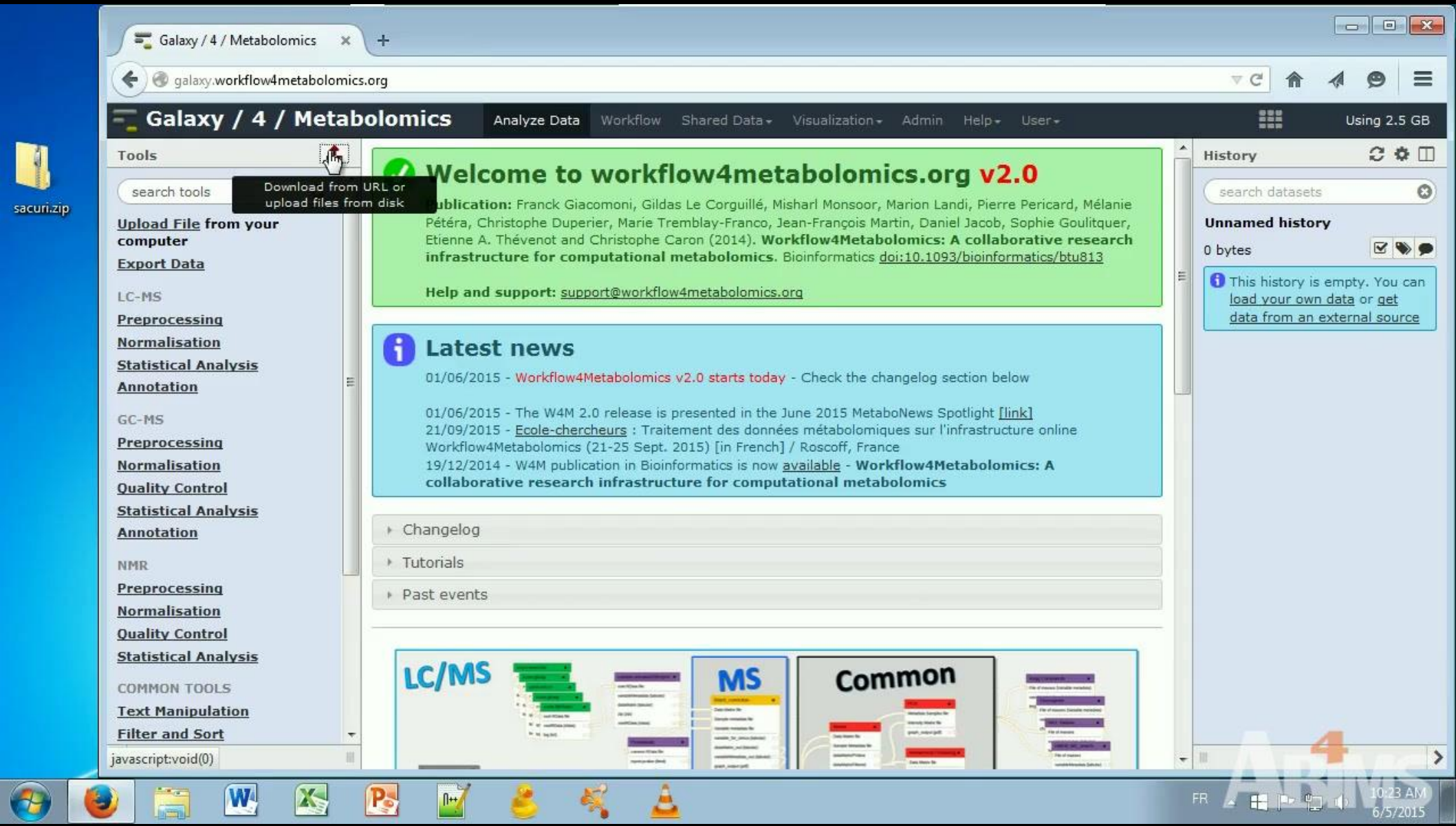

Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 12%
Tools
search tools
Get Data
COMMON TOOLS
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Text Manipulation
Graphics
Statistics
EMBOSS 5 Suite
SEARCHING TOOLS
NCBI BLAST+
Diamond
Primer/Microsatellite
NGS TOOLS
NGS: BedTools
NGS: Mapping
NGS: Picard
NGS: QC and manipulation
NGS: SAM Tools
GATK Tools
GATK2 Tools
DNA-SEQ TOOLS
DNA-Seq tools
Welcome to galaxy.sb-roscoff.fr
Warning
17-05-16: For performance and maintenance reasons, we will disable the tool Export2Dir. You can always export your whole history in a file or download one file using the old fashion "floppy disk" icon. Export2Dir could return on day after a little refactoring?
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.
History
search datasets
sartools deseq2
7 shown
353.2 KB
7: SARTools DESeq2 R objects (.RData)
6: SARTools DESeq2 R log
5: SARTools DESeq2 figures
4: SARTools DESeq2 tables
3: SARTools DESeq2 report
2: targetT048.txt
1: t048.zip

DATA IMPORT

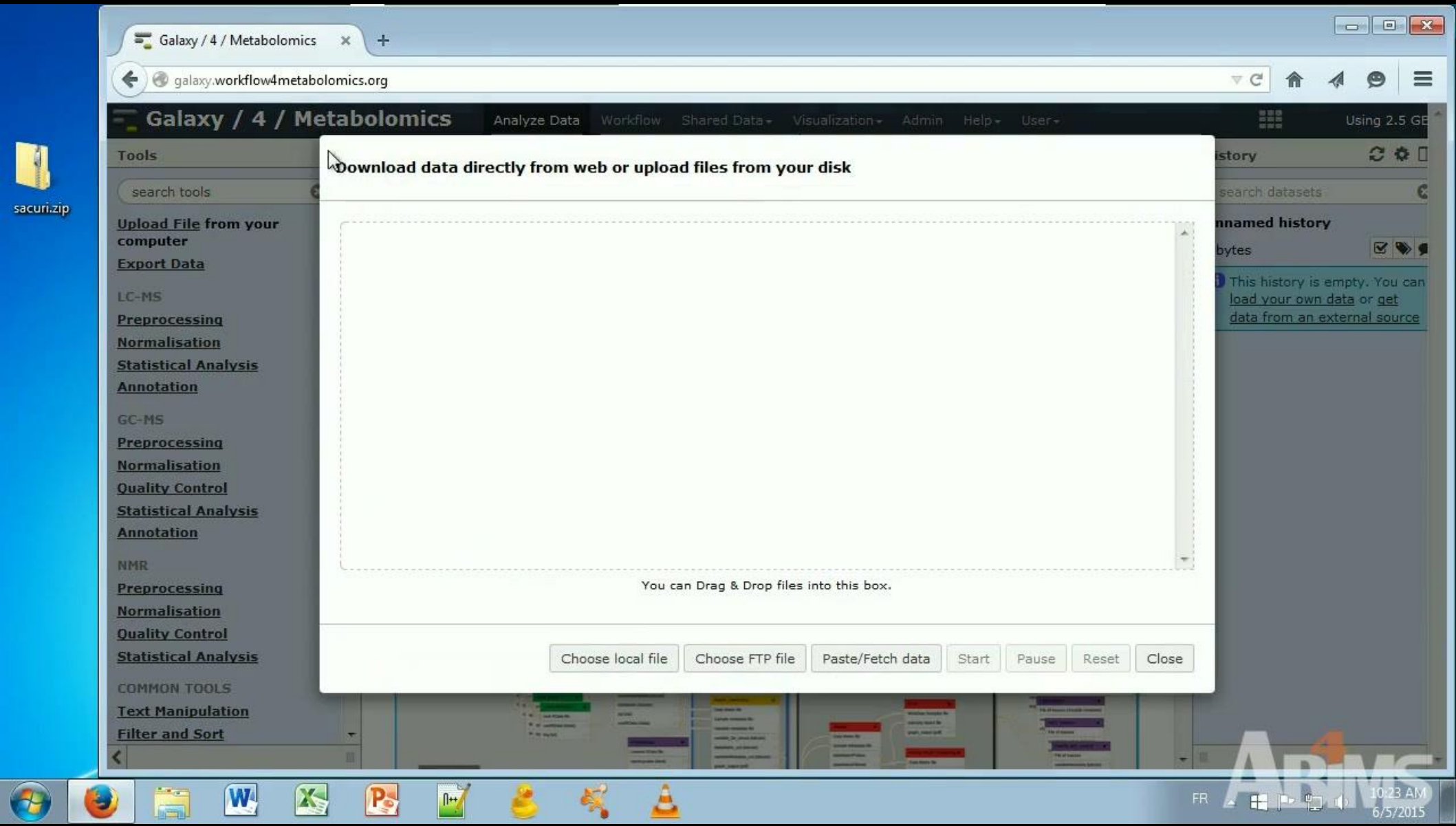
DATA IMPORT

< 2 GO

Data import < 2 Go

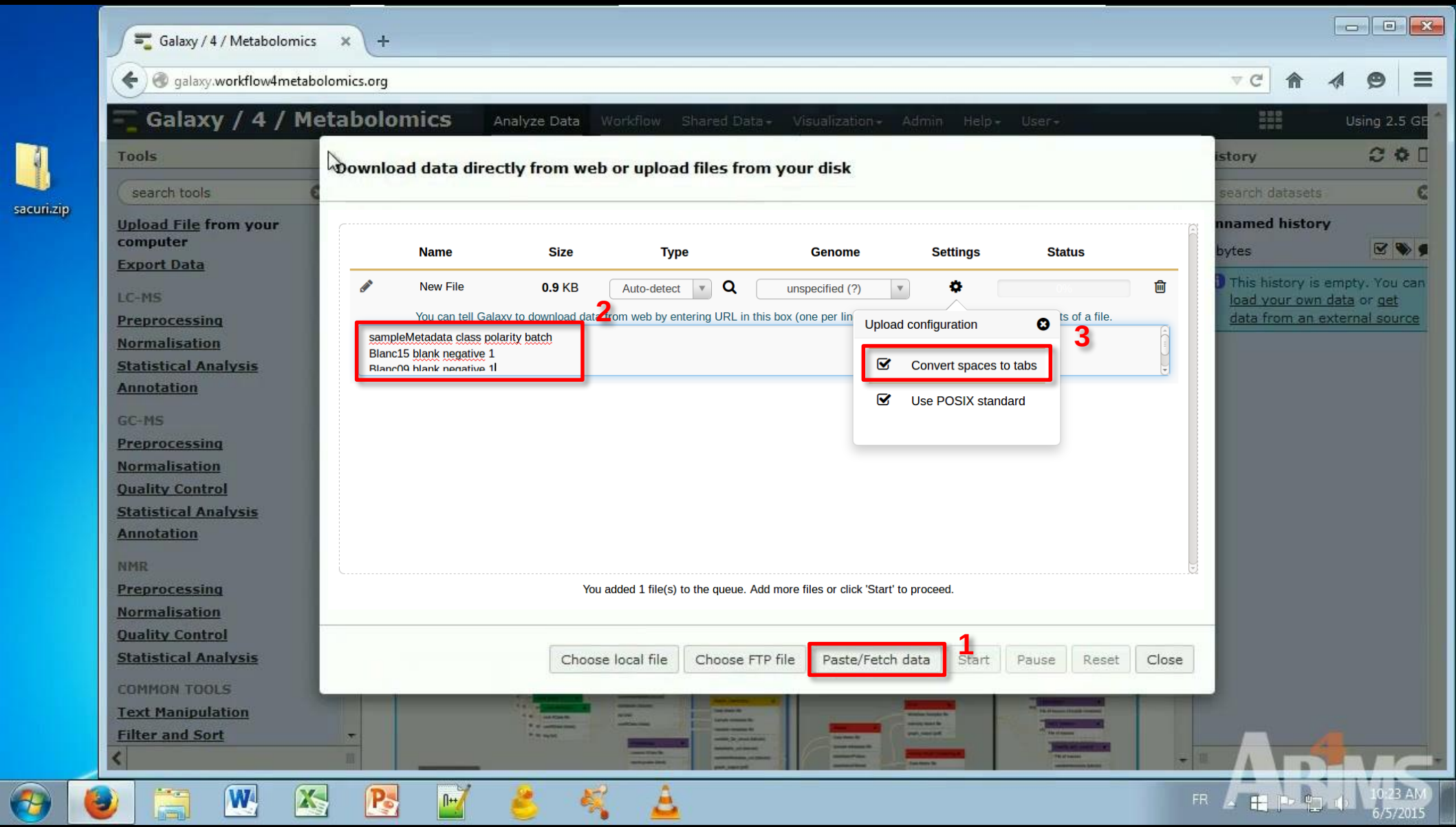


Data import < 2 Go



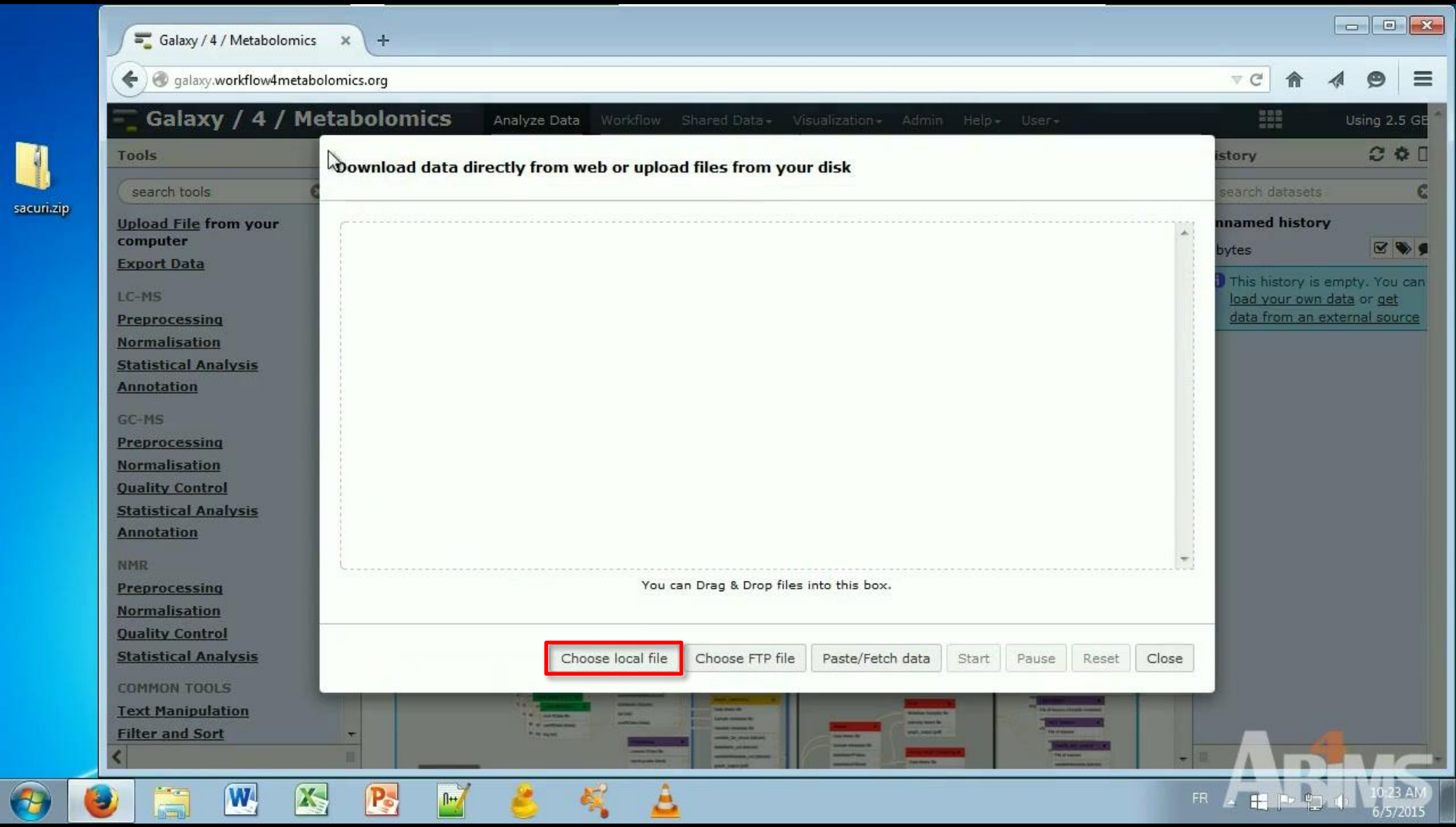
Data import < 2 Go

Copy / Paste data



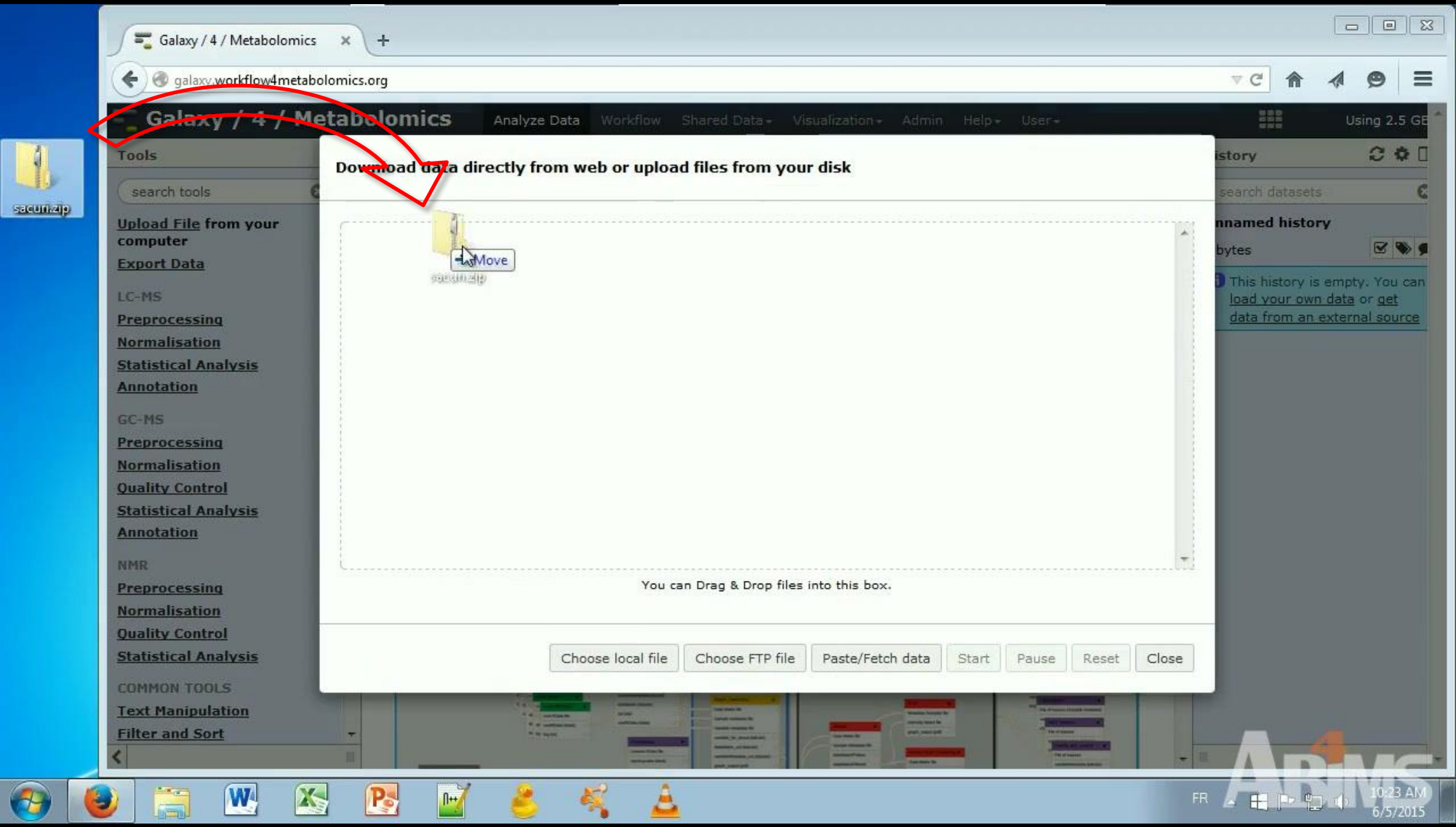
Data import < 2 Go

From local files



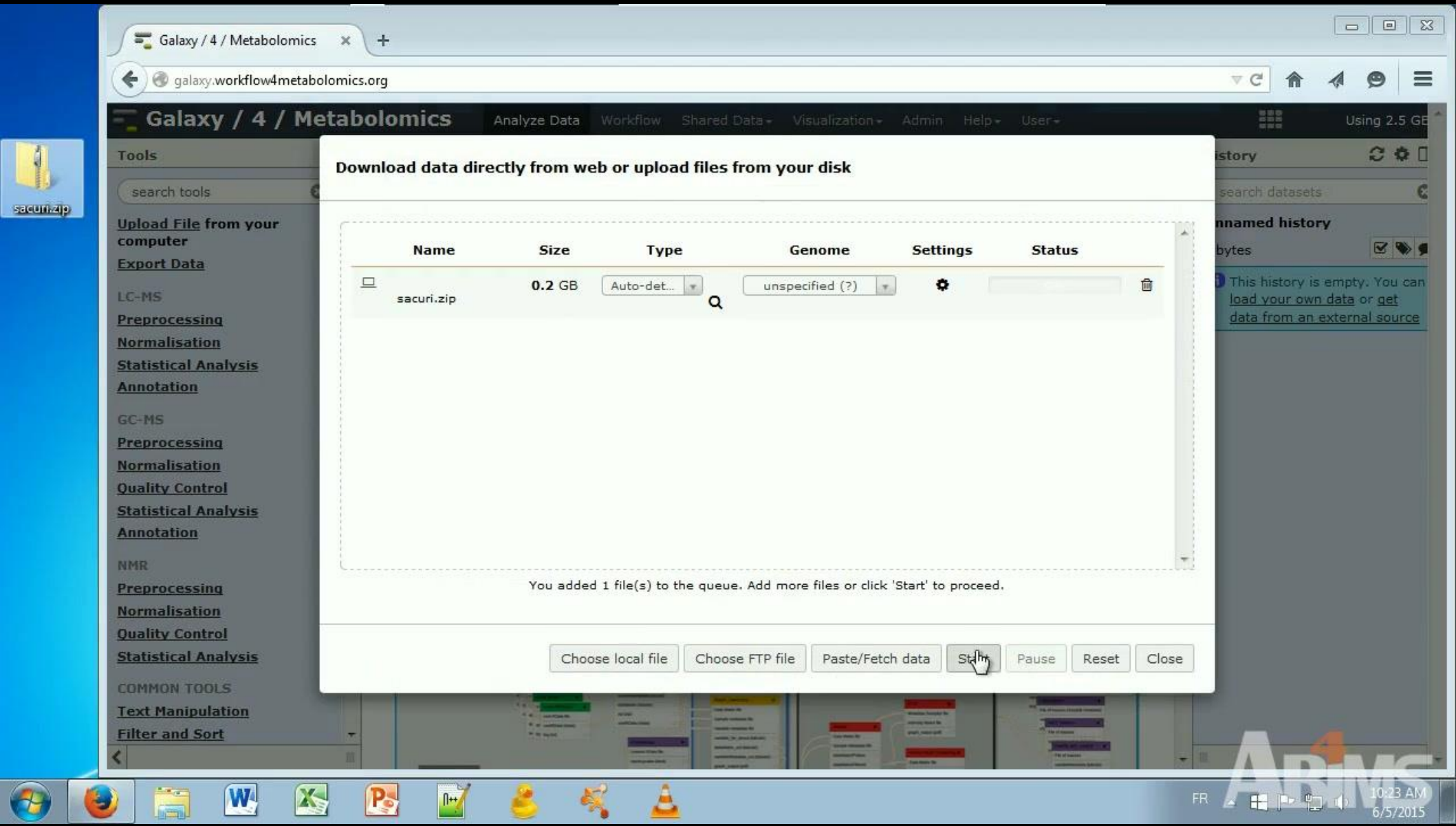
Data import < 2 Go

From local files



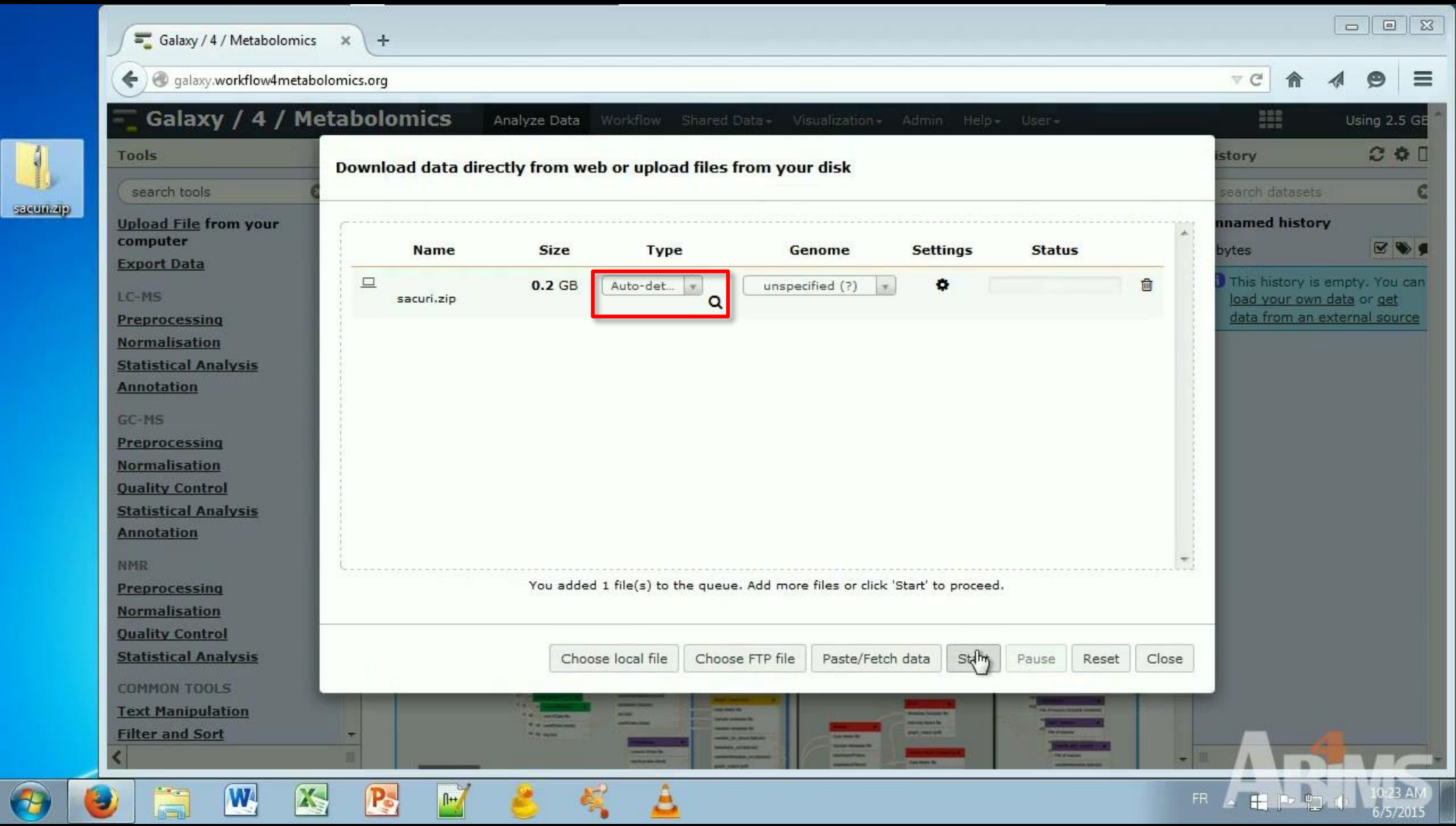
Data import < 2 Go

From local files



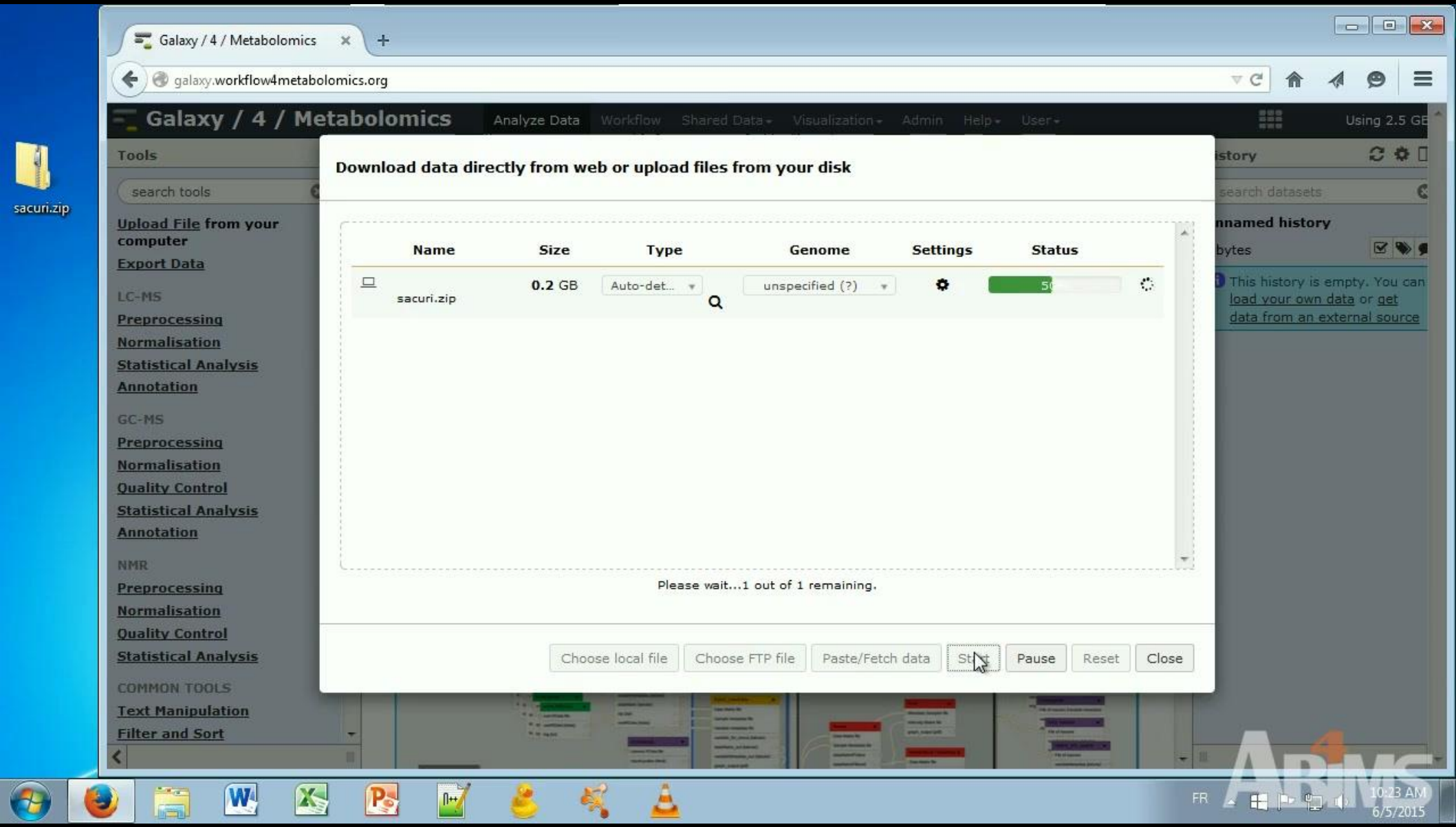
Data import < 2 Go

From local files



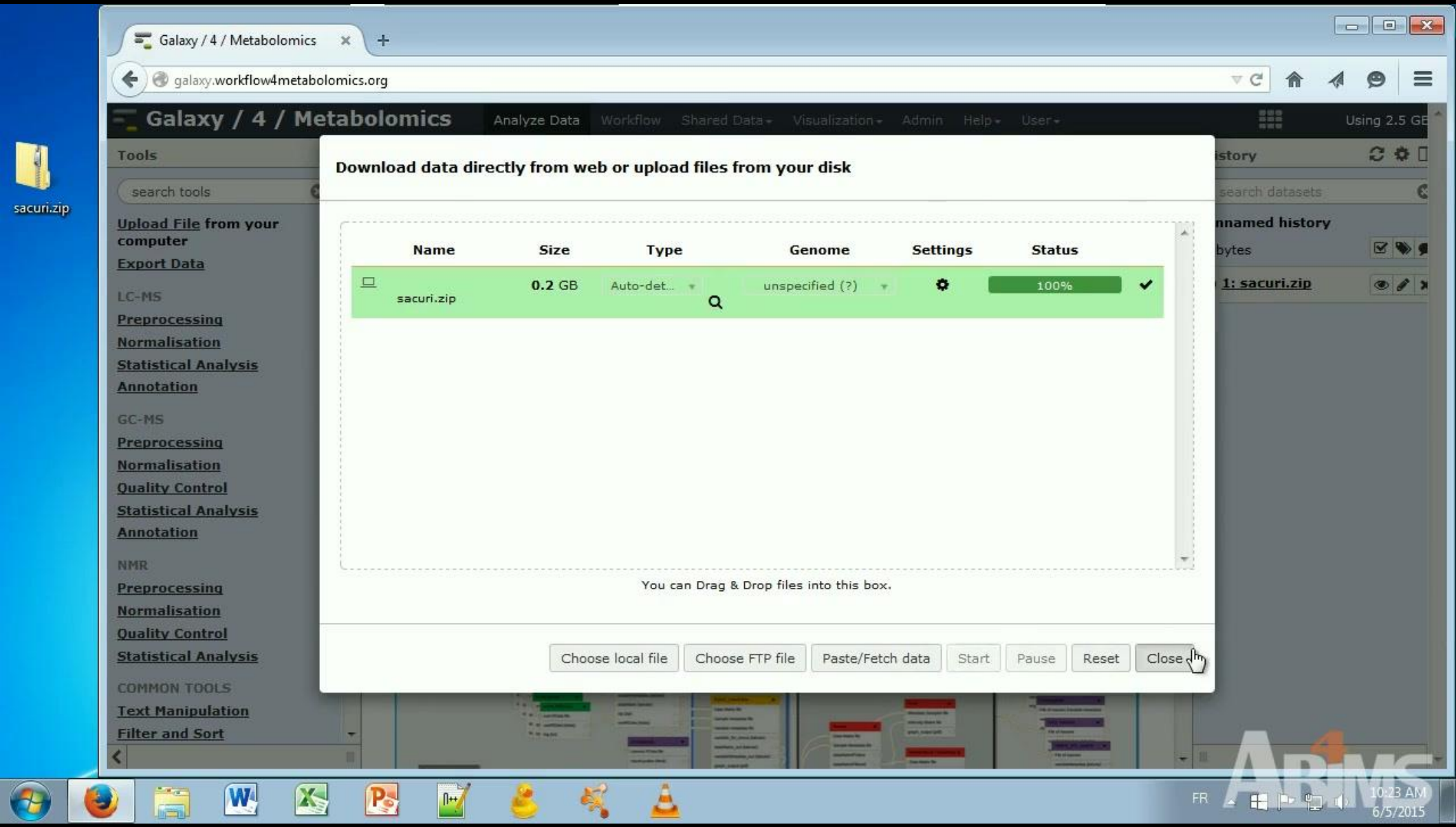
Data import < 2 Go

From local files



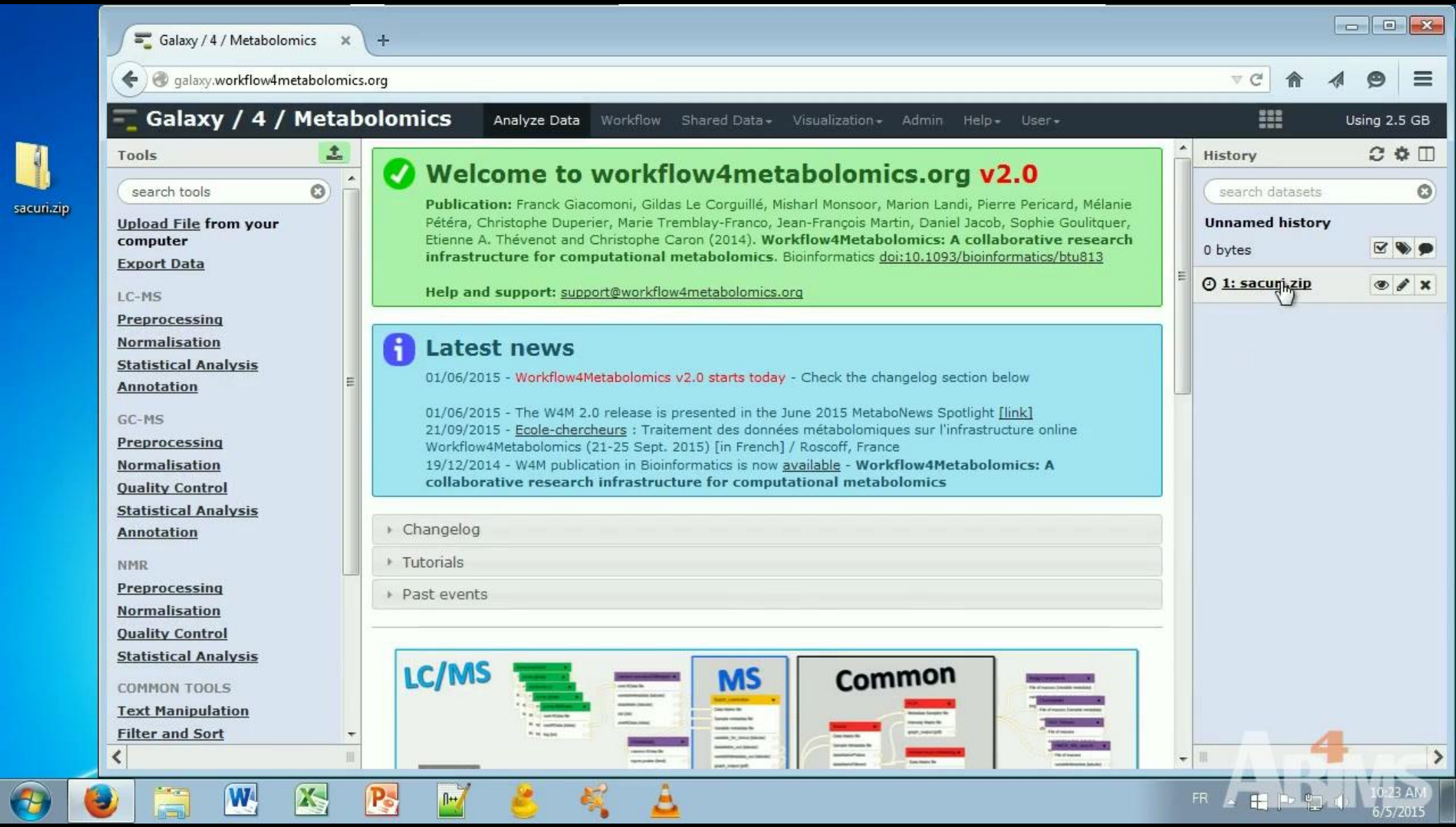
Data import < 2 Go

From local files



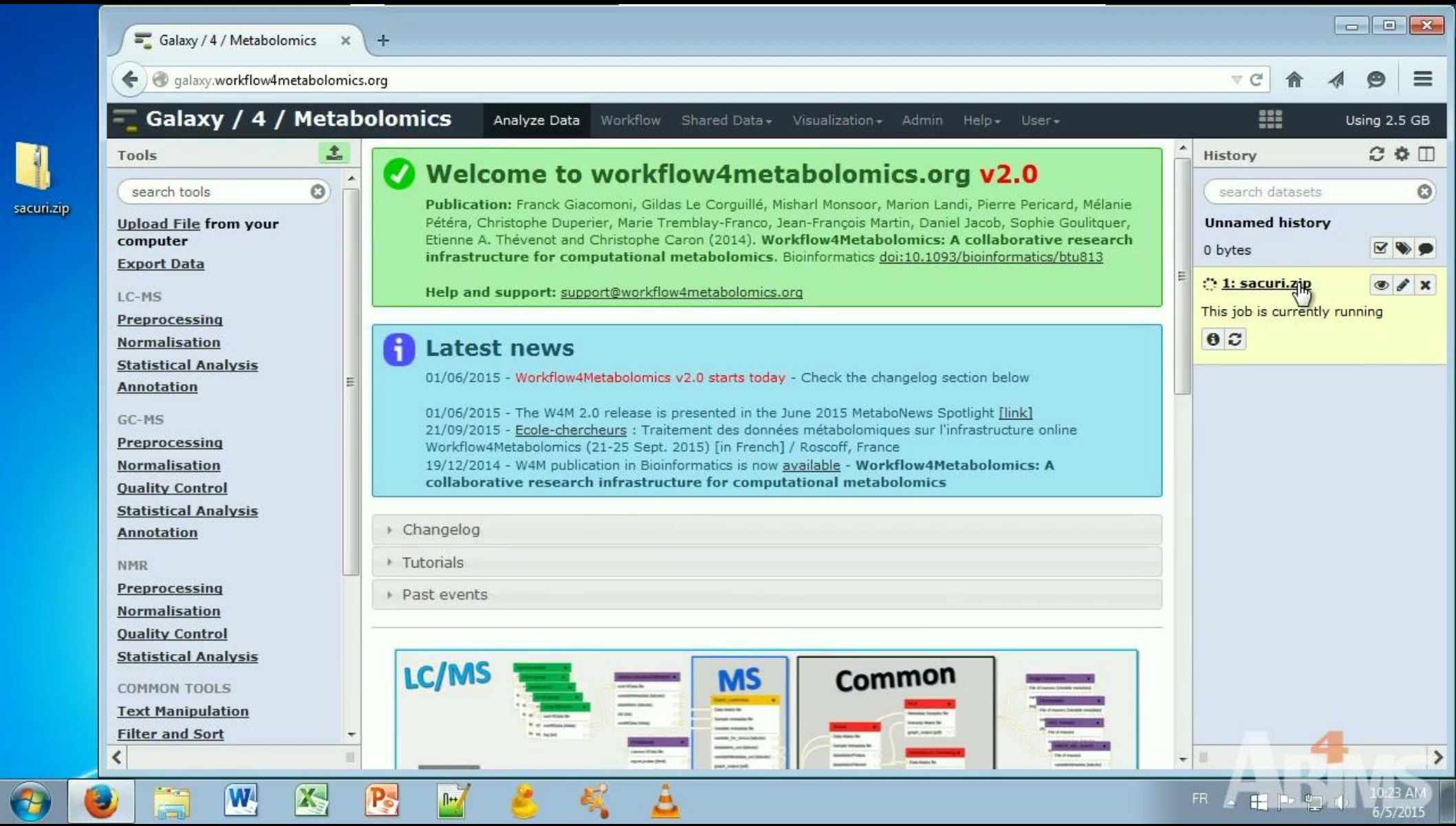
Data import < 2 Go

From local files



Data import < 2 Go

From local files



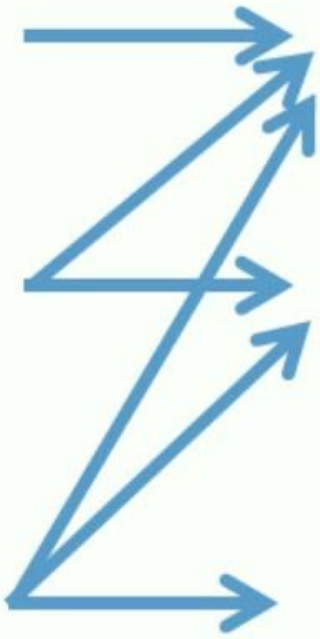
Step 1: Choose a FTP Client

DATA IMPORT

> 2 GO

STEP 1: CHOOSE A FTP CLIENT

STEP 1: CHOOSE A FTP CLIENT



Avoid:
Malwares inside

Cyberduck

WinSCP

STEP 1: CHOOSE A FTP CLIENT

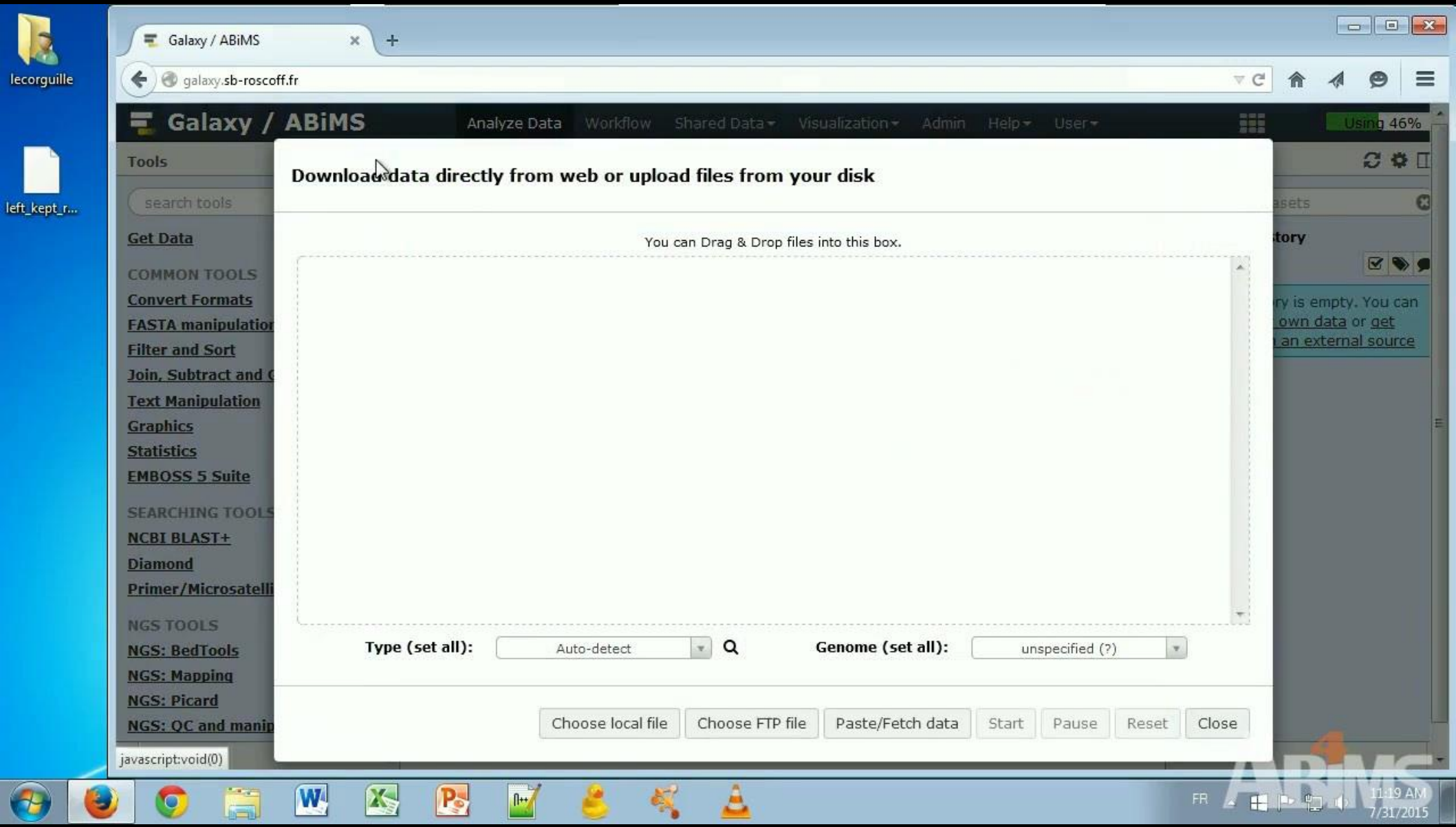
Step 2: Easy!

DATA IMPORT

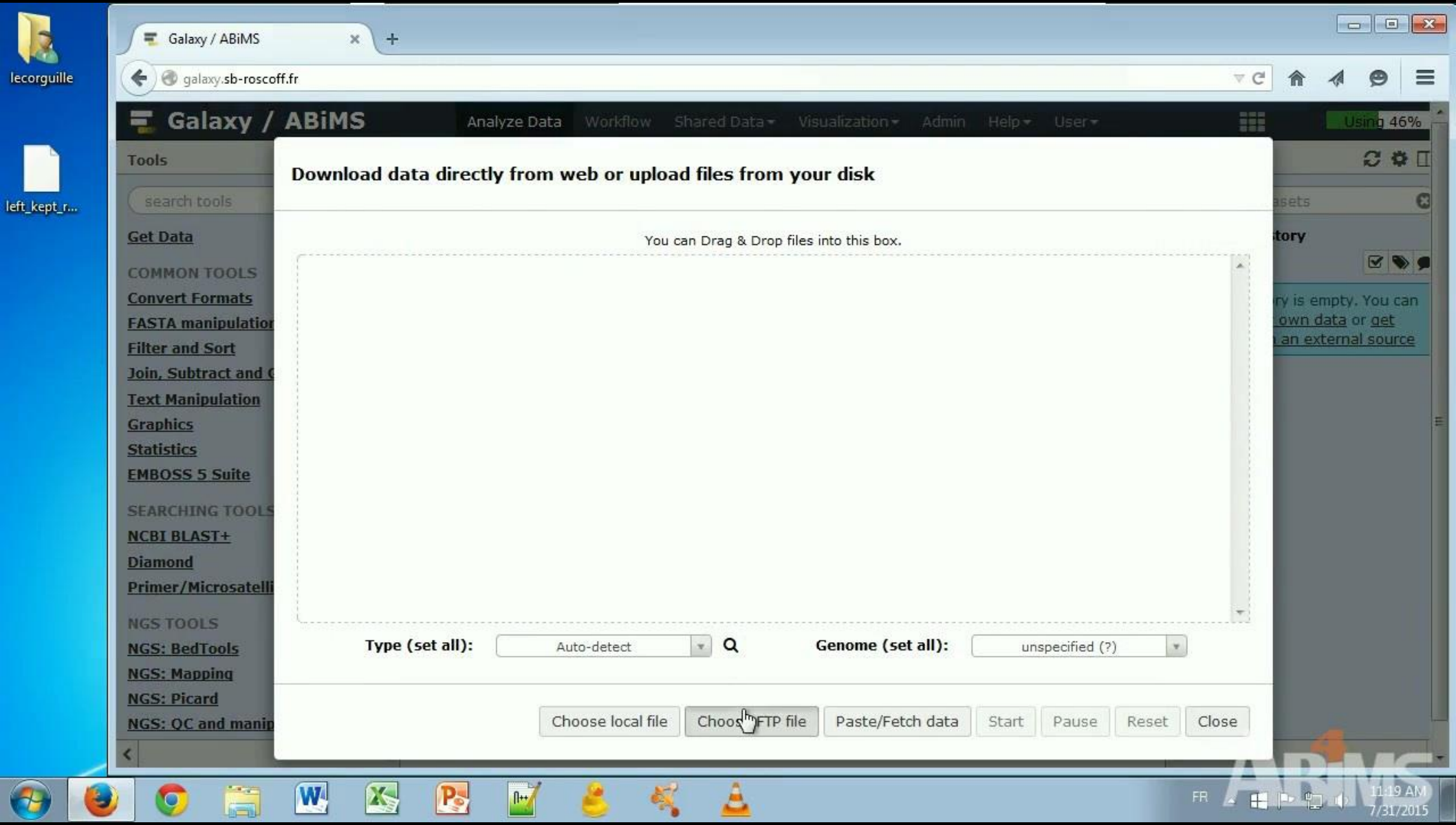
> 2 GO

Data import > 2 Go

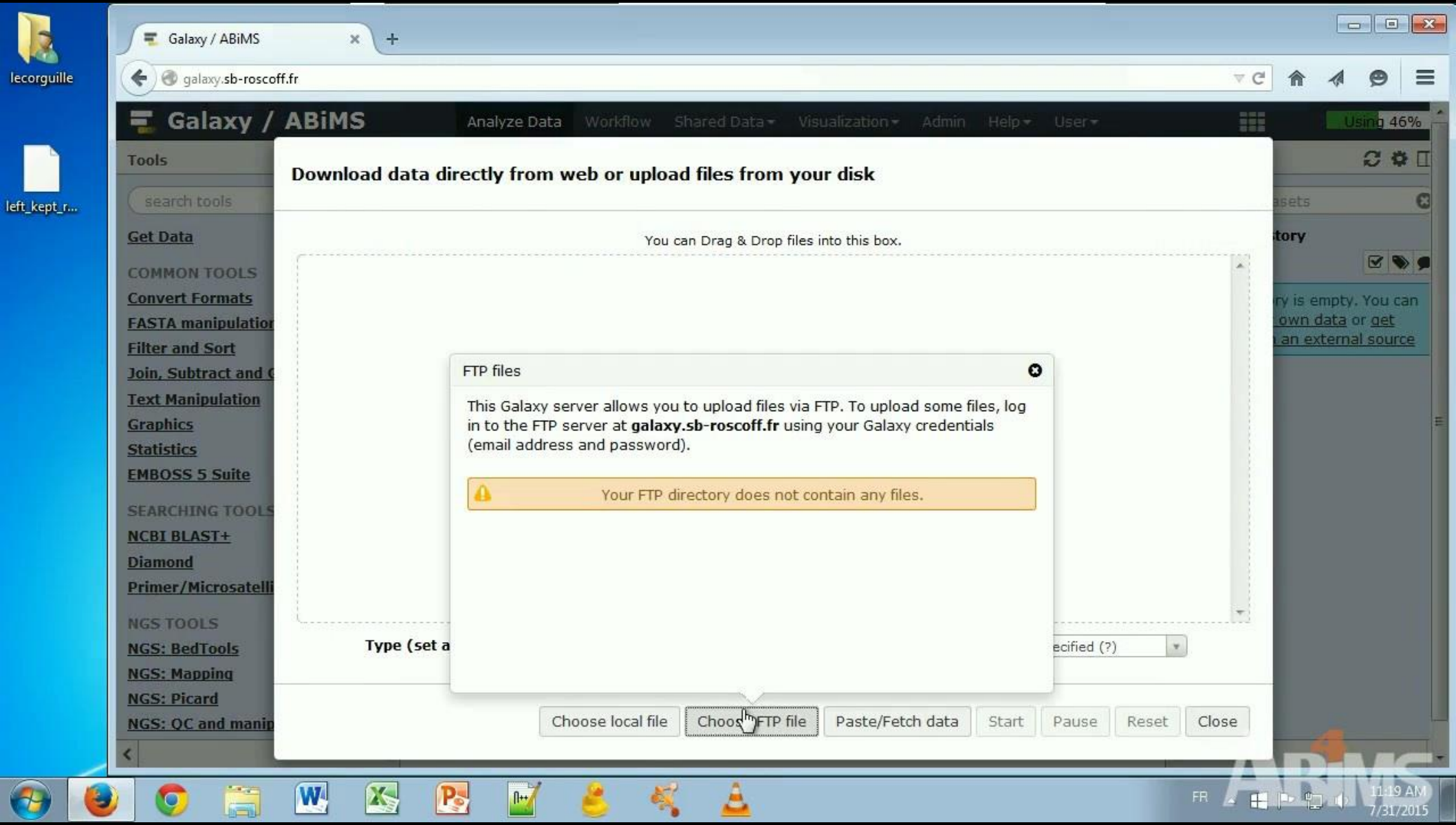
Data import > 2 Go



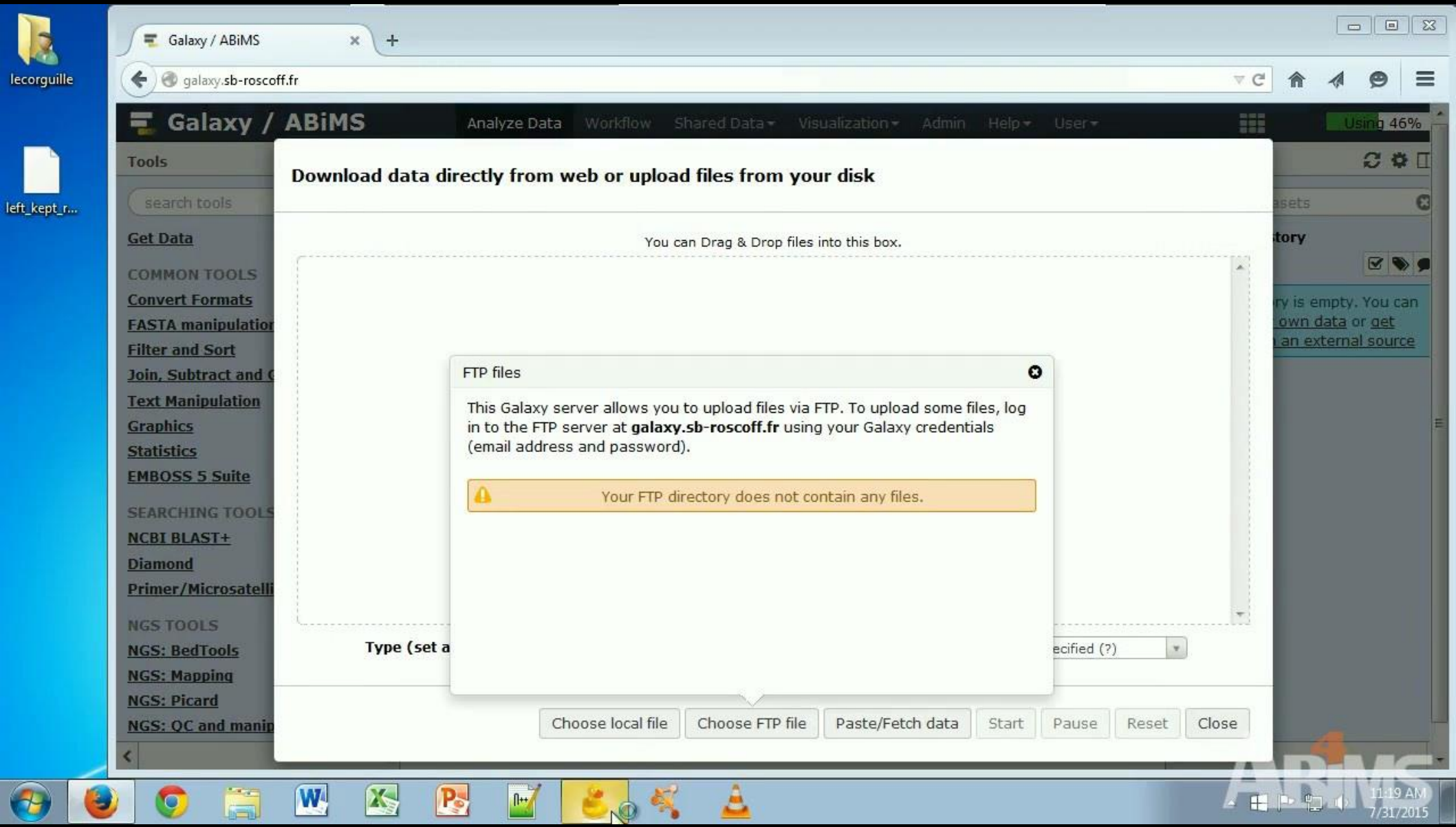
Data import > 2 Go



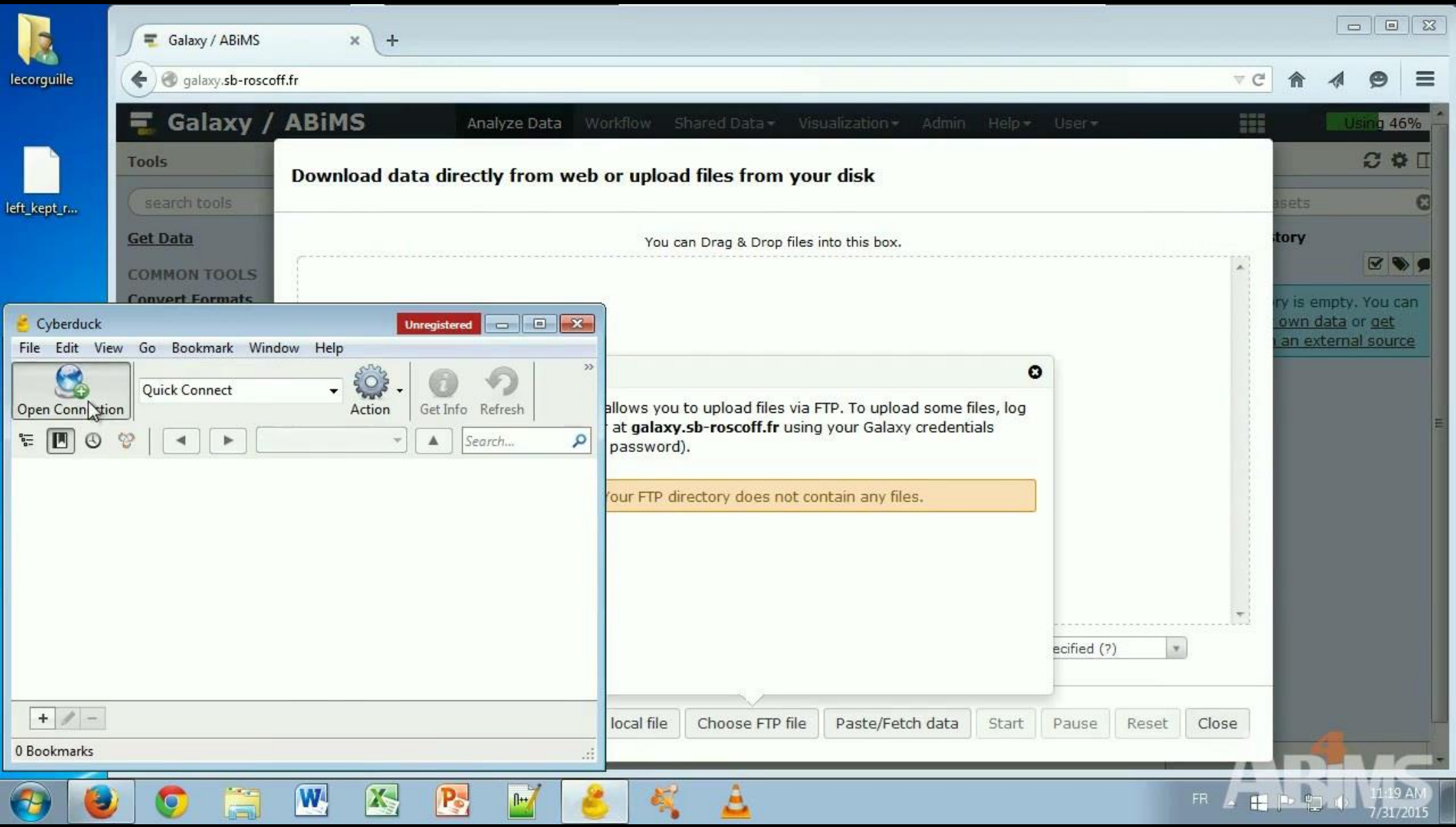
Data import > 2 Go



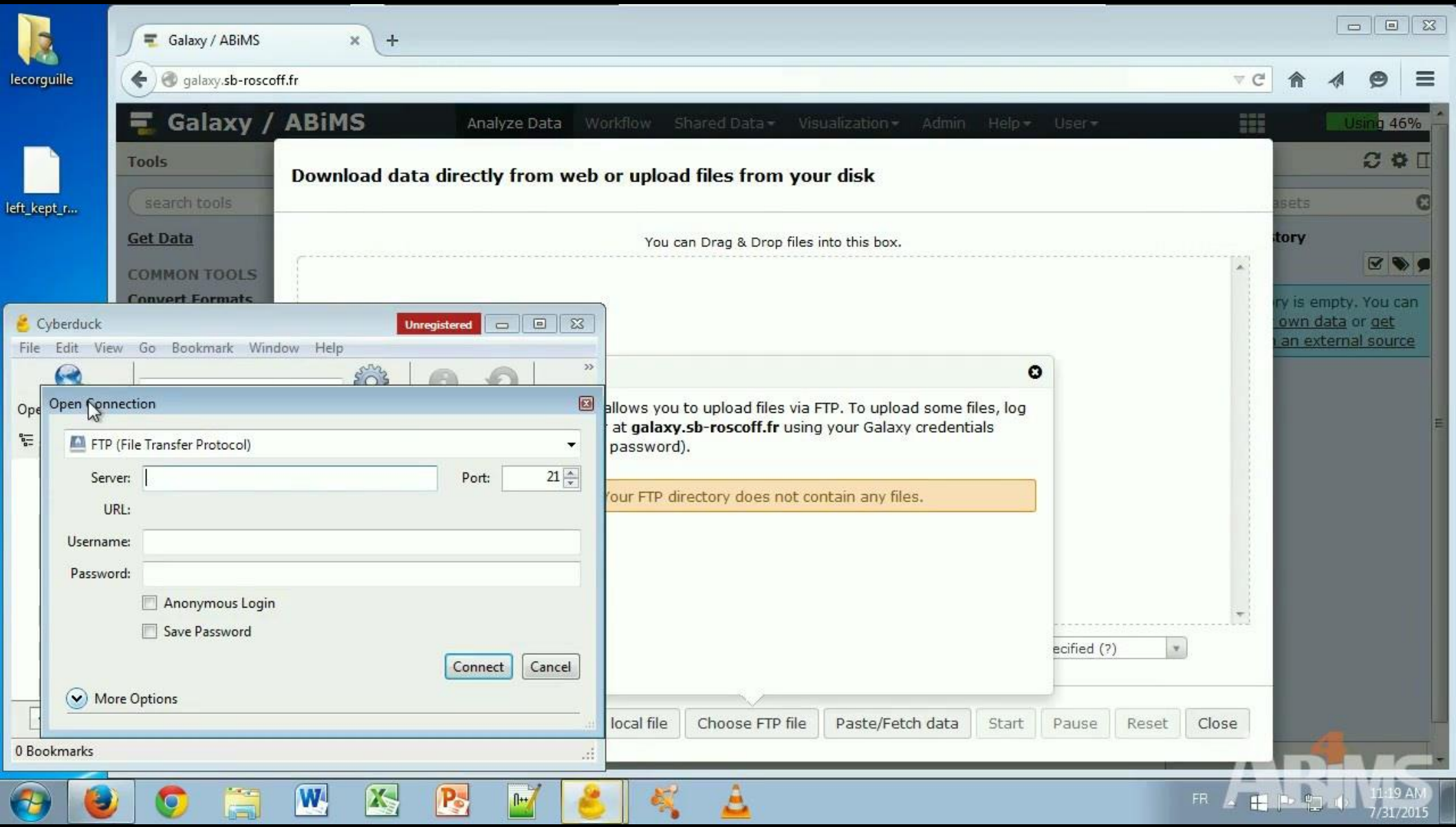
Data import > 2 Go



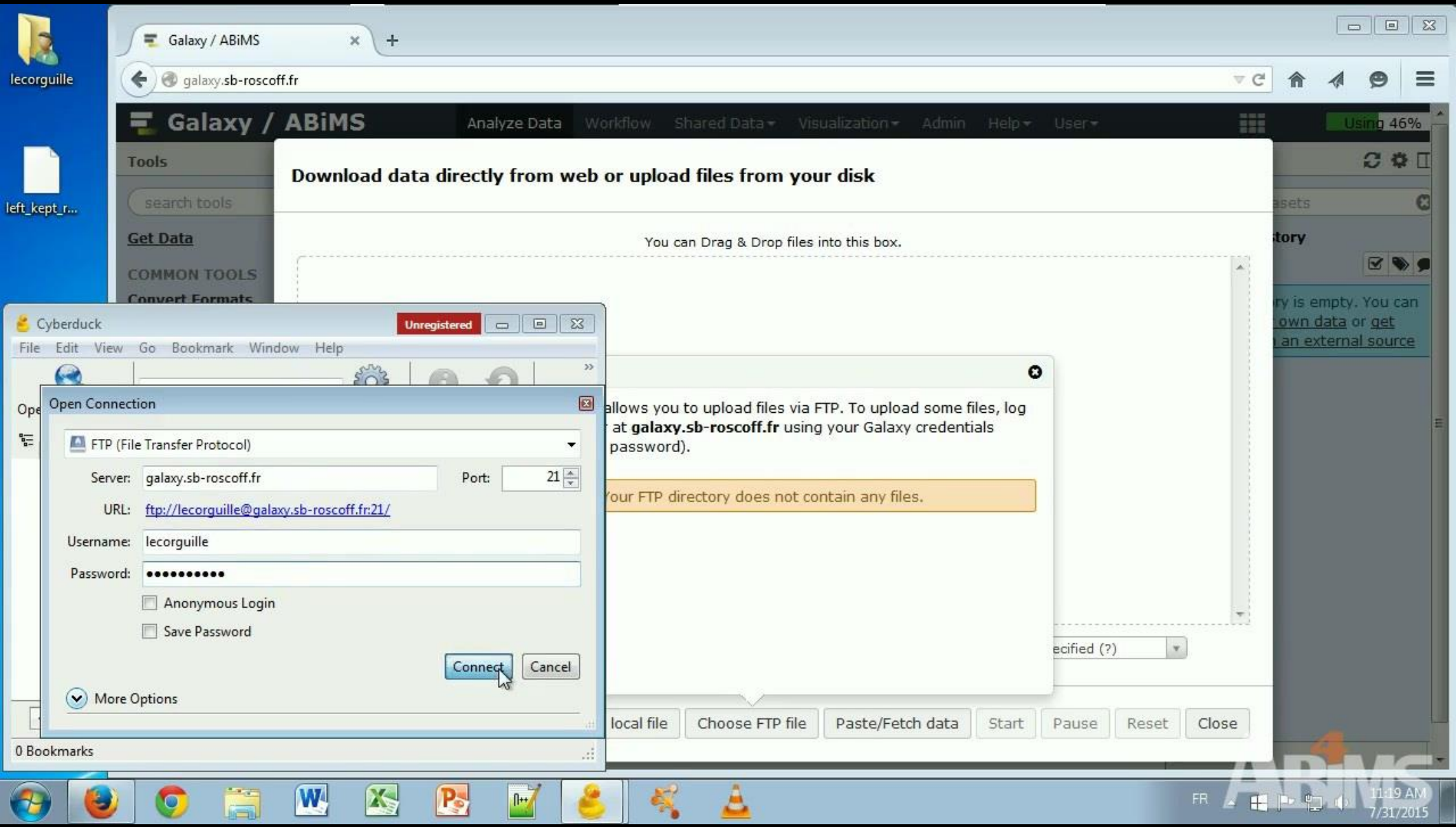
Data import > 2 Go



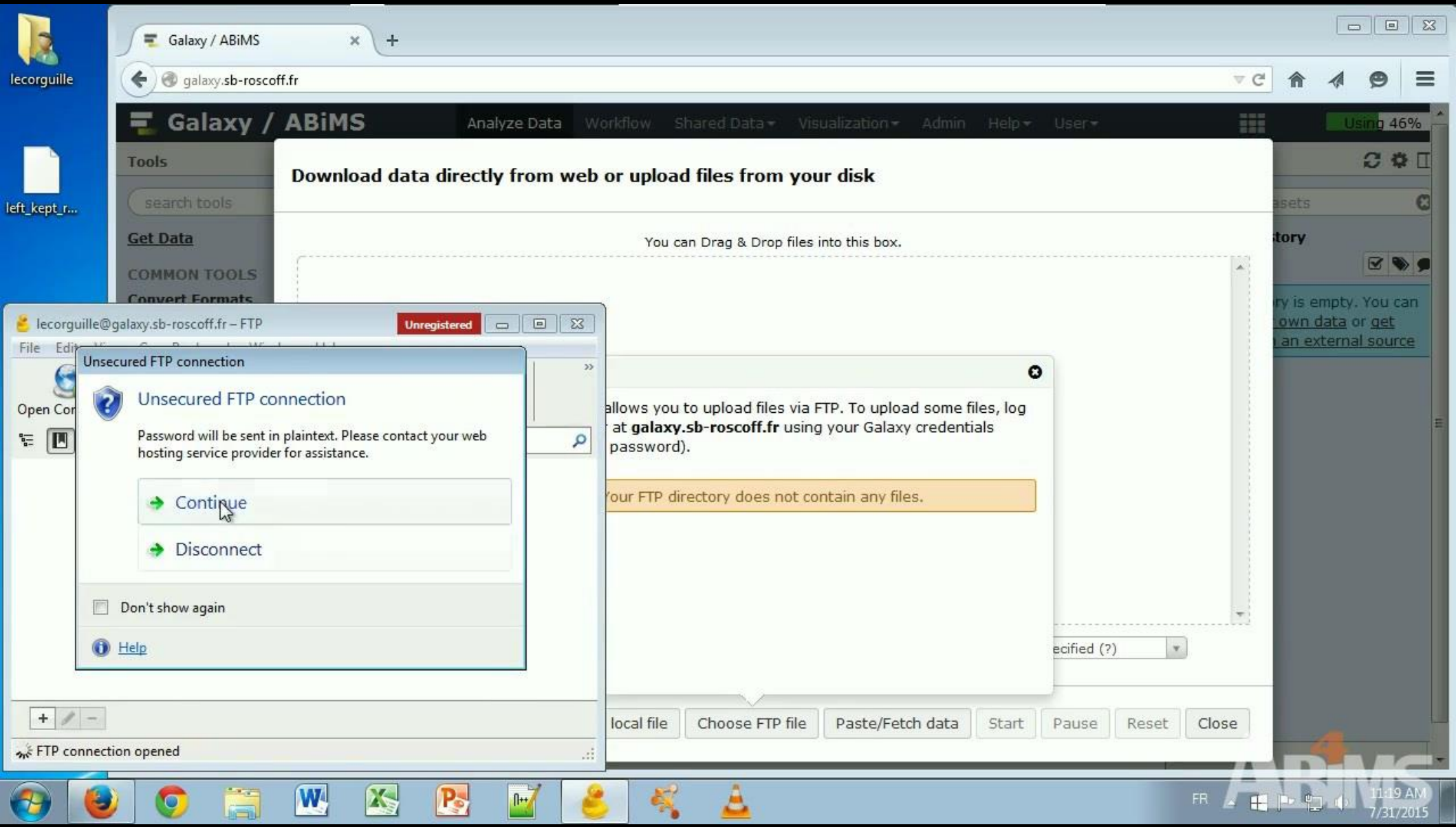
Data import > 2 Go



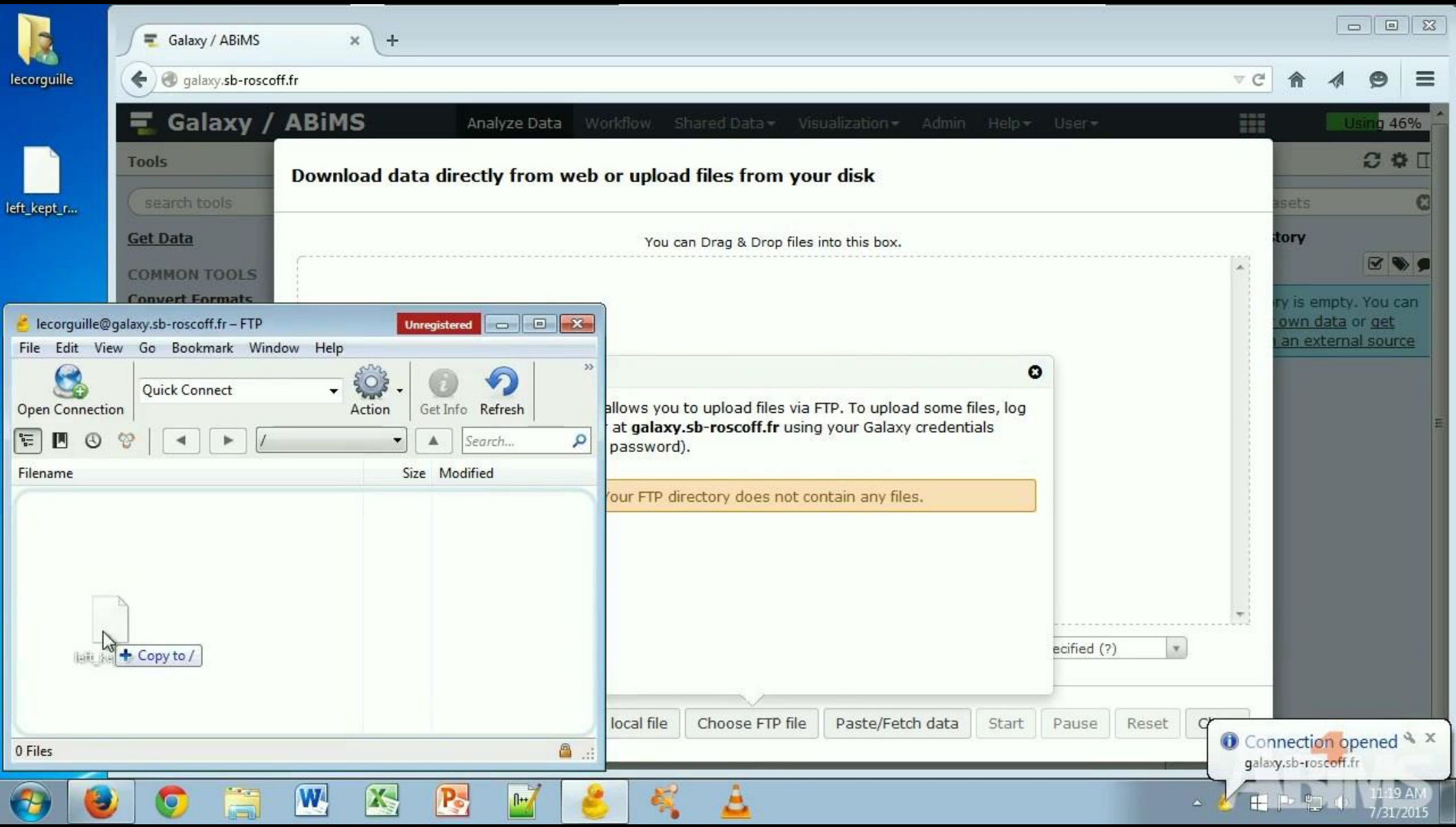
Data import > 2 Go



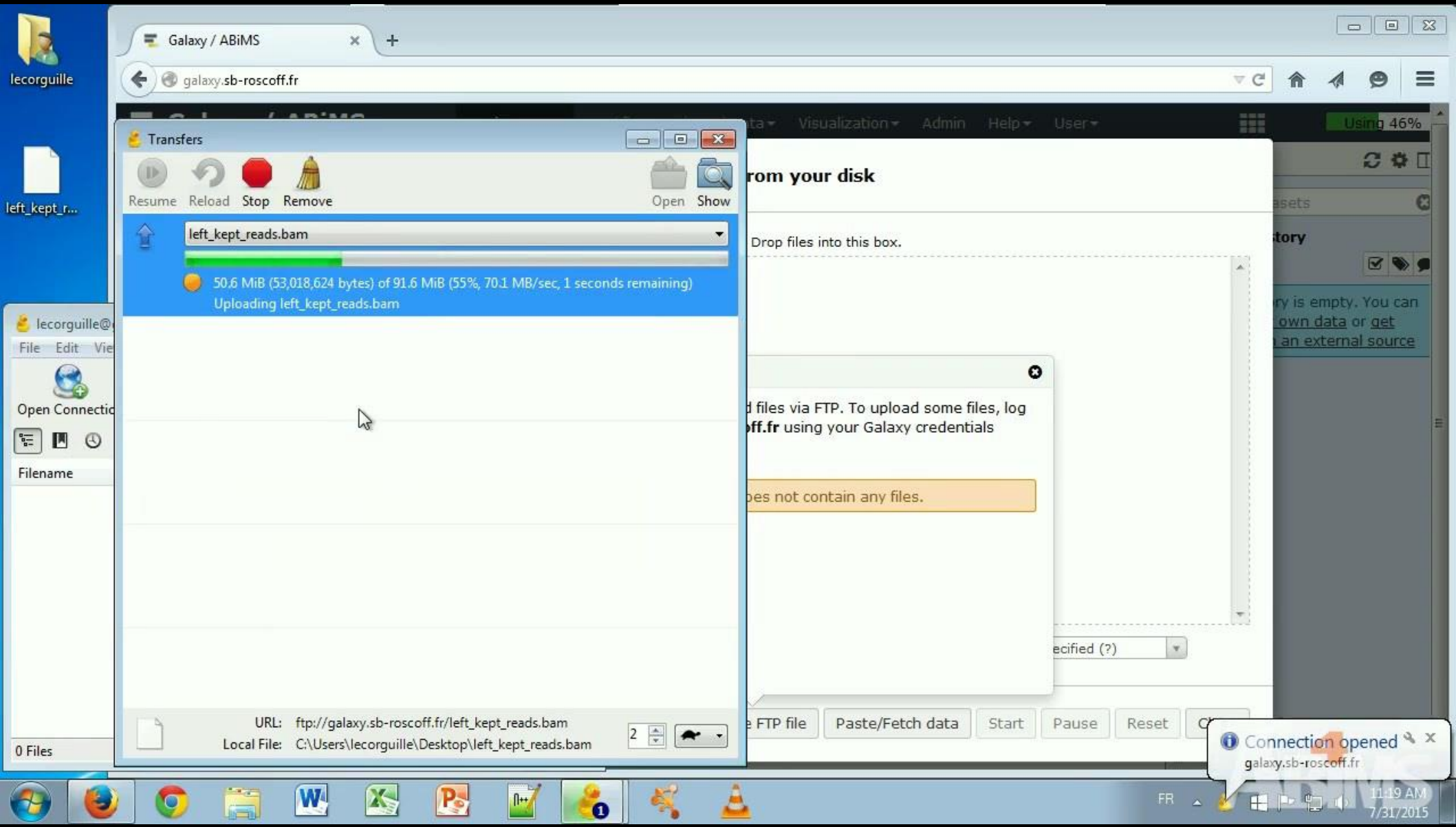
Data import > 2 Go



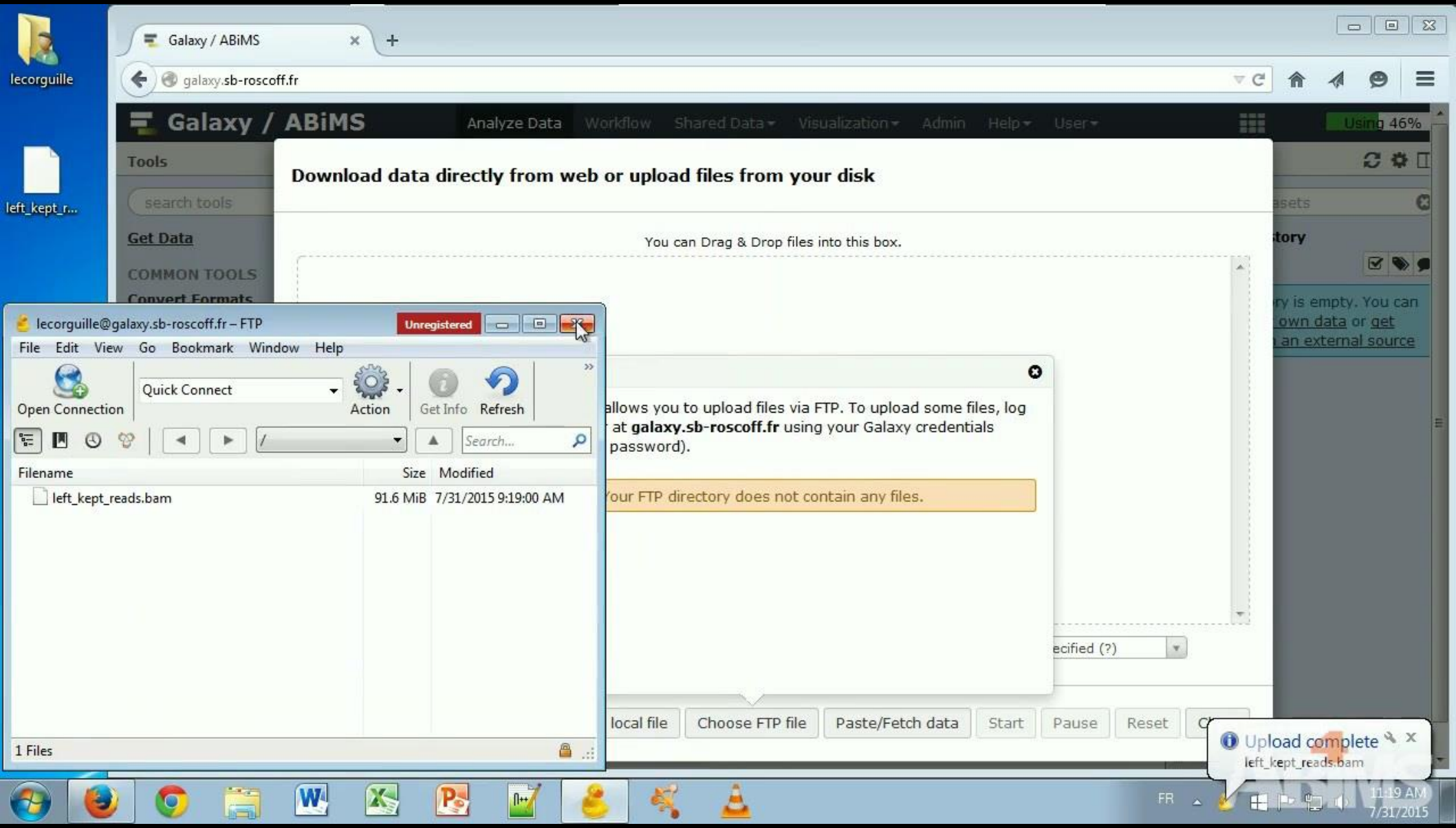
Data import > 2 Go



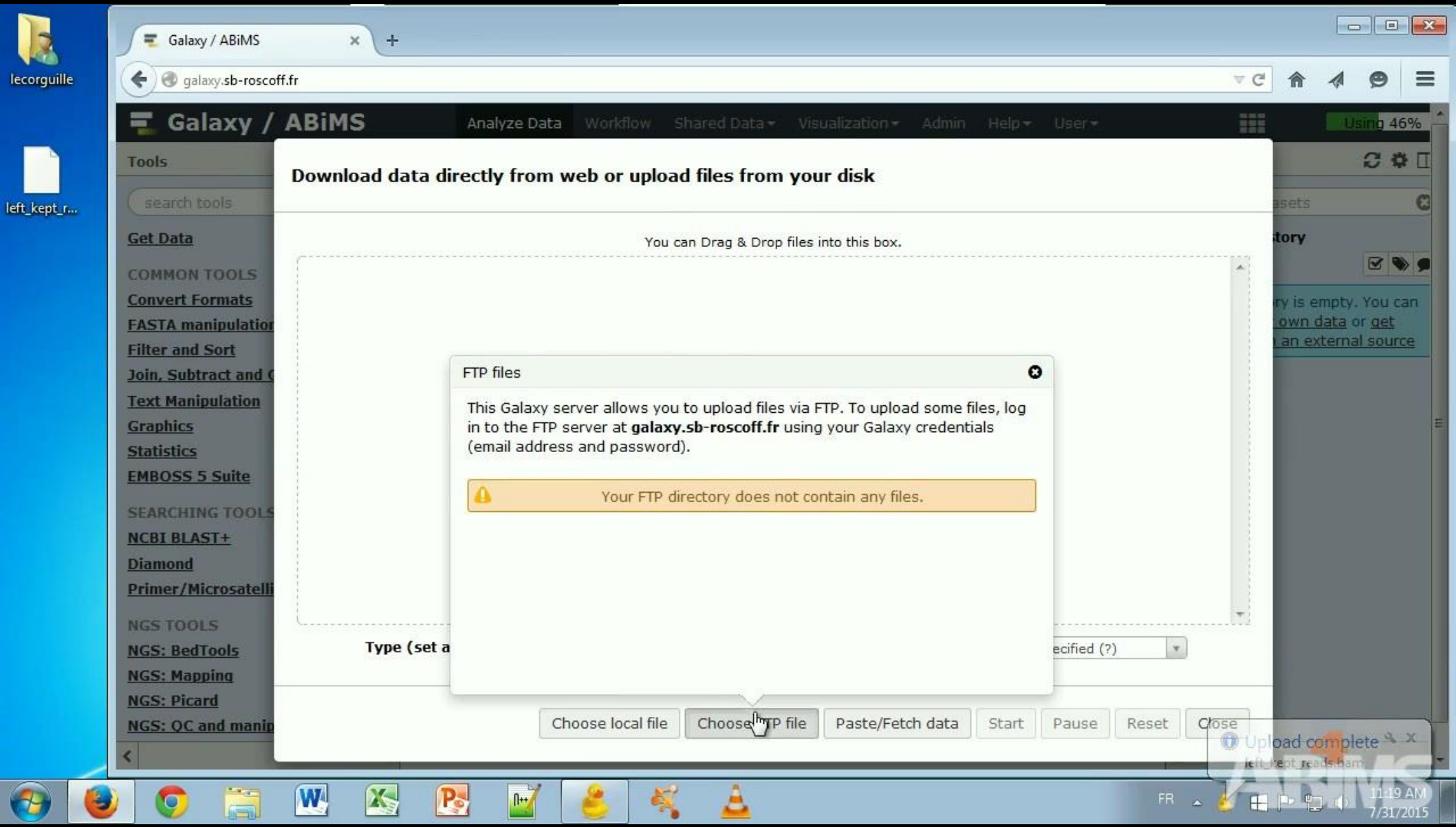
Data import > 2 Go



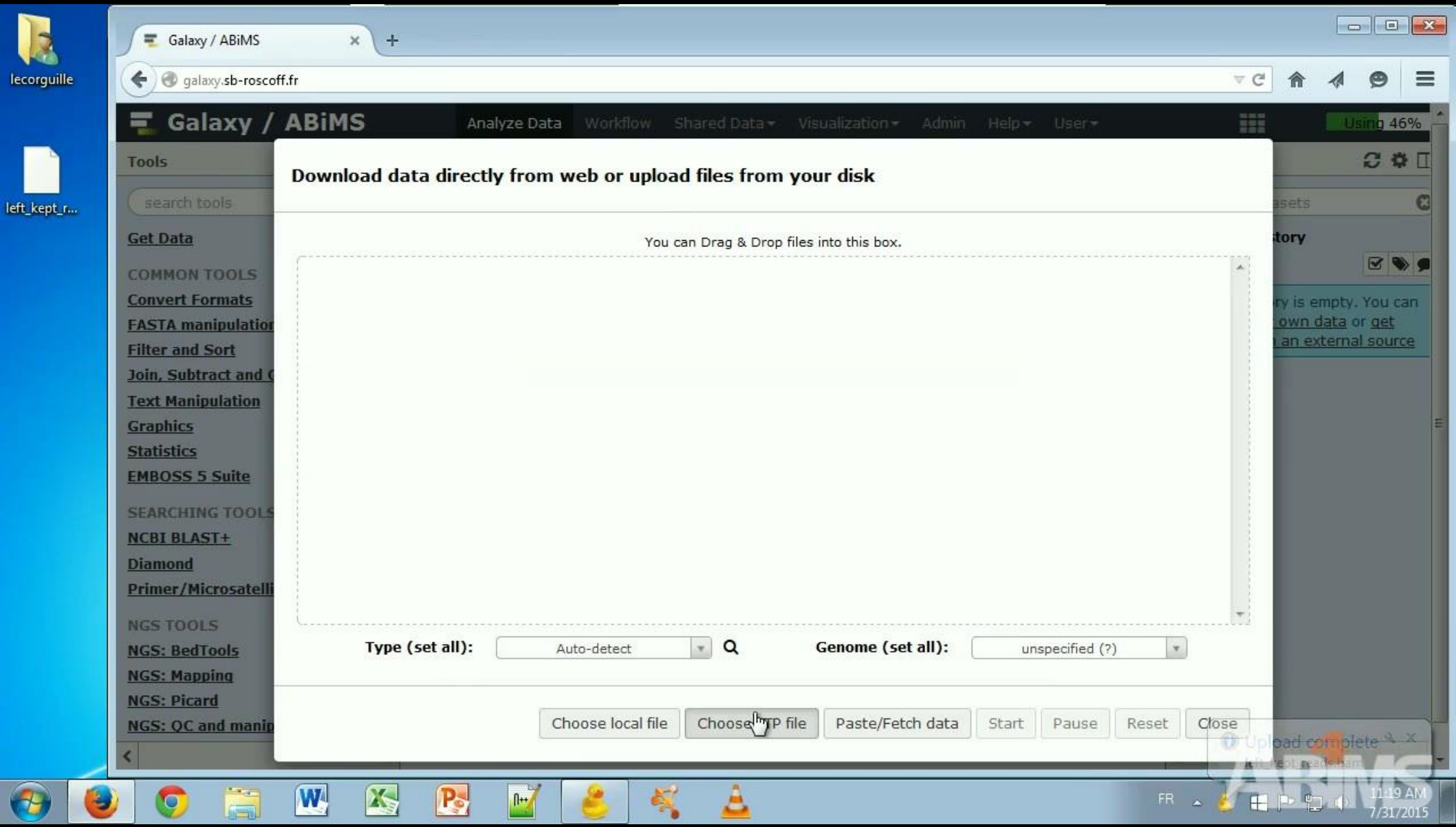
Data import > 2 Go



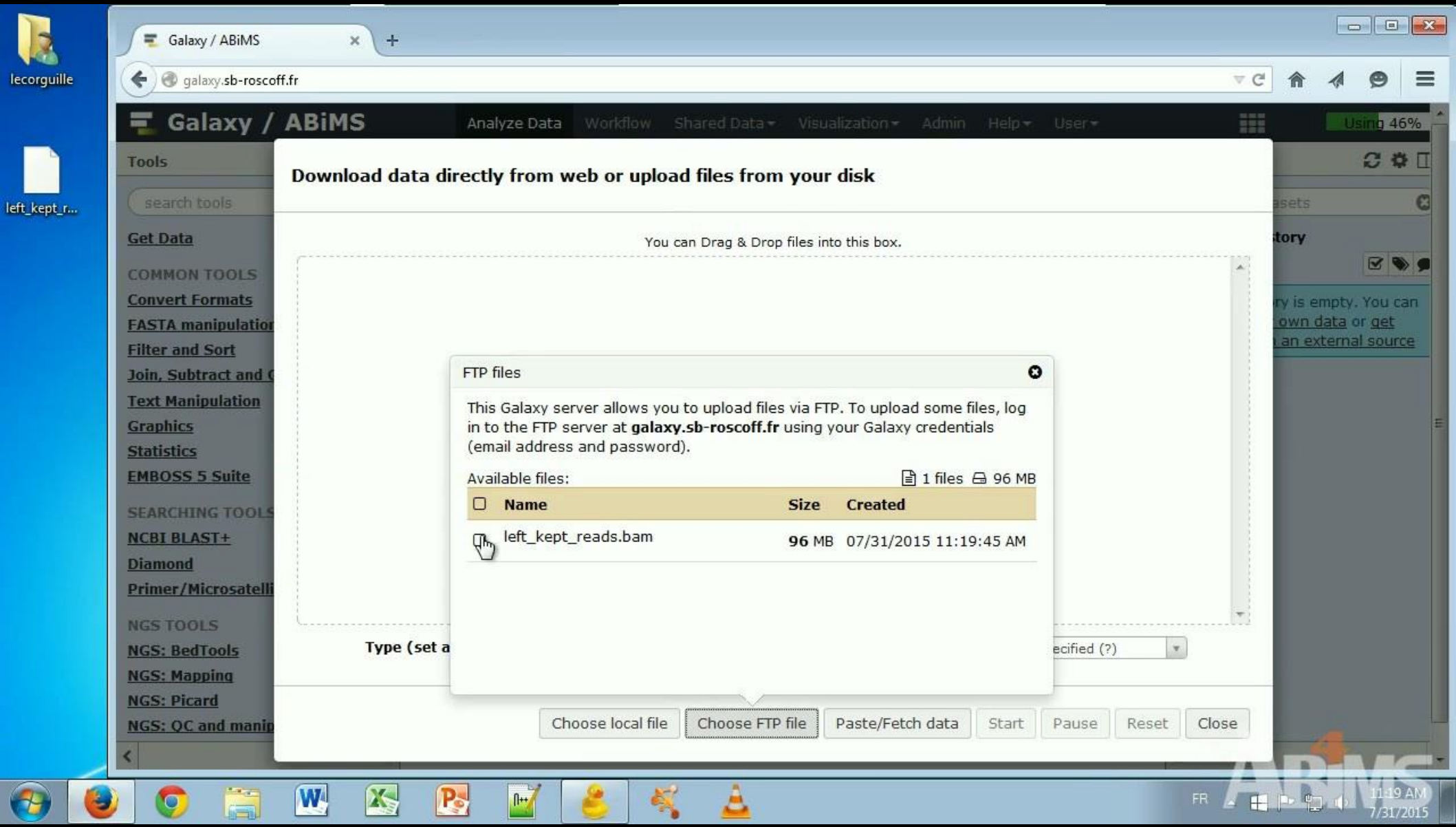
Data import > 2 Go



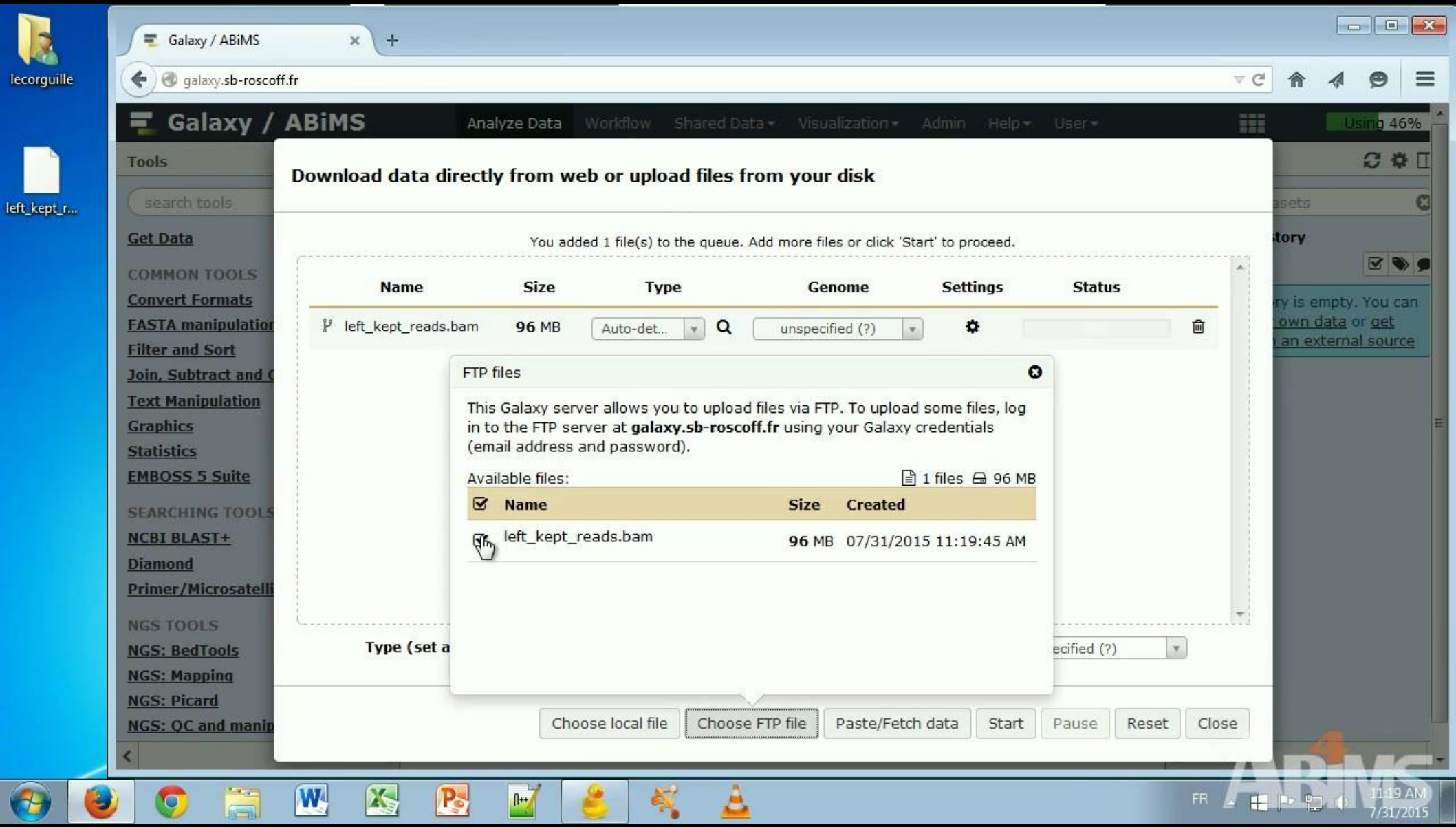
Data import > 2 Go



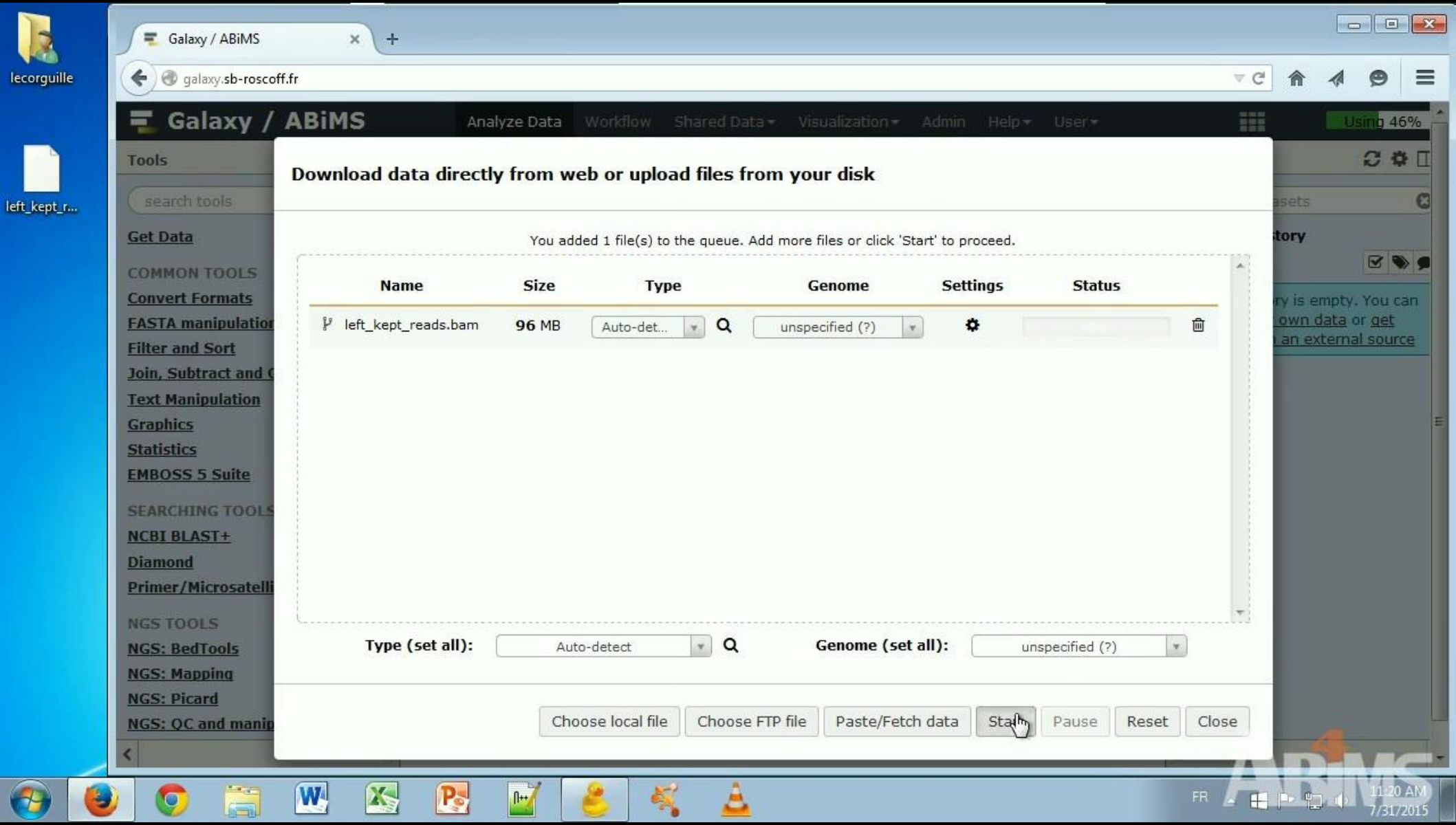
Data import > 2 Go



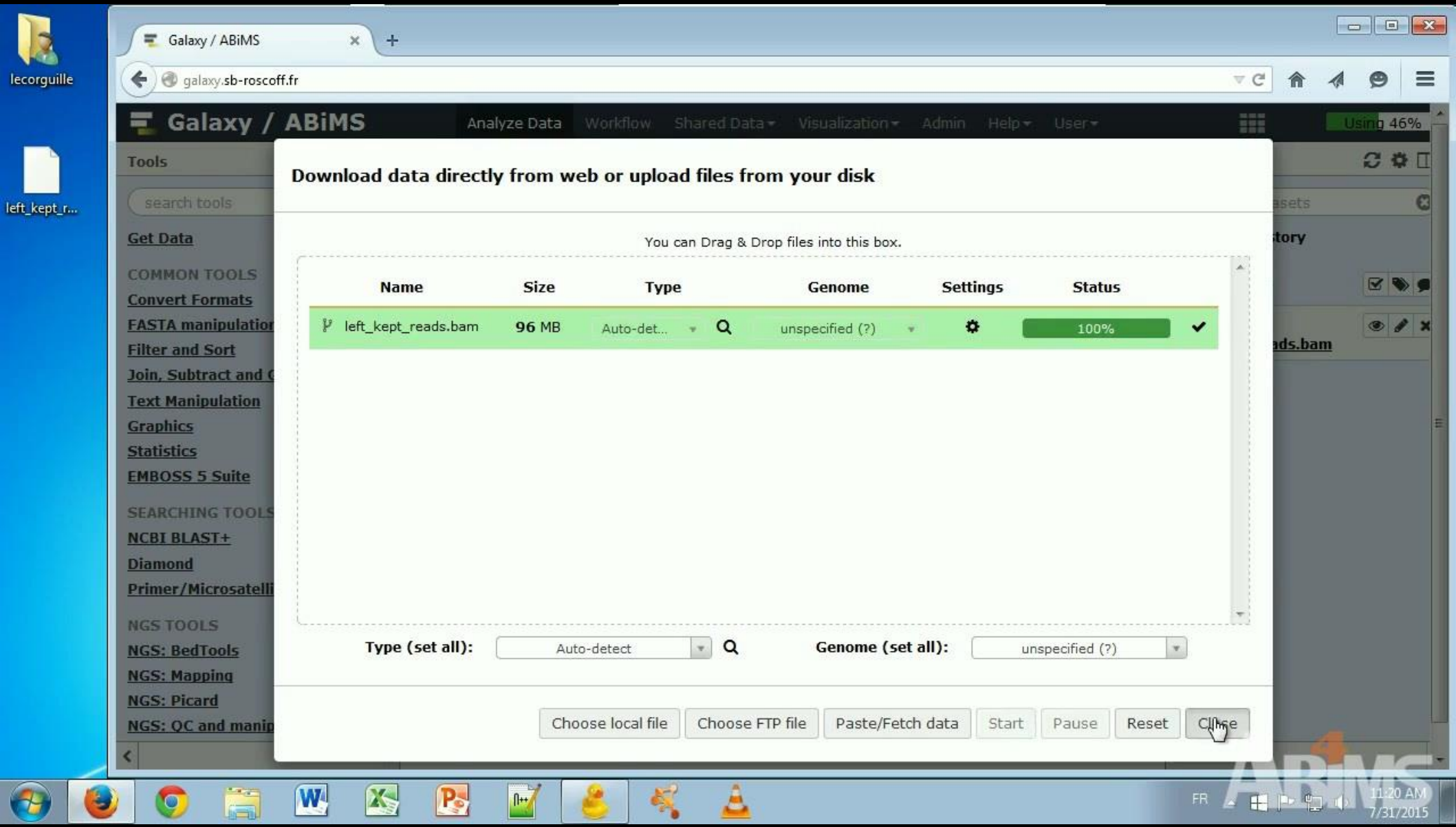
Data import > 2 Go



Data import > 2 Go



Data import > 2 Go



Data import > 2 Go

Data import > 2 Go

Data import > 2 Go

Exercise

DATA IMPORT

Data import

- Exercise
  - Fetch this file

http://tinyurl.com/GI-input0

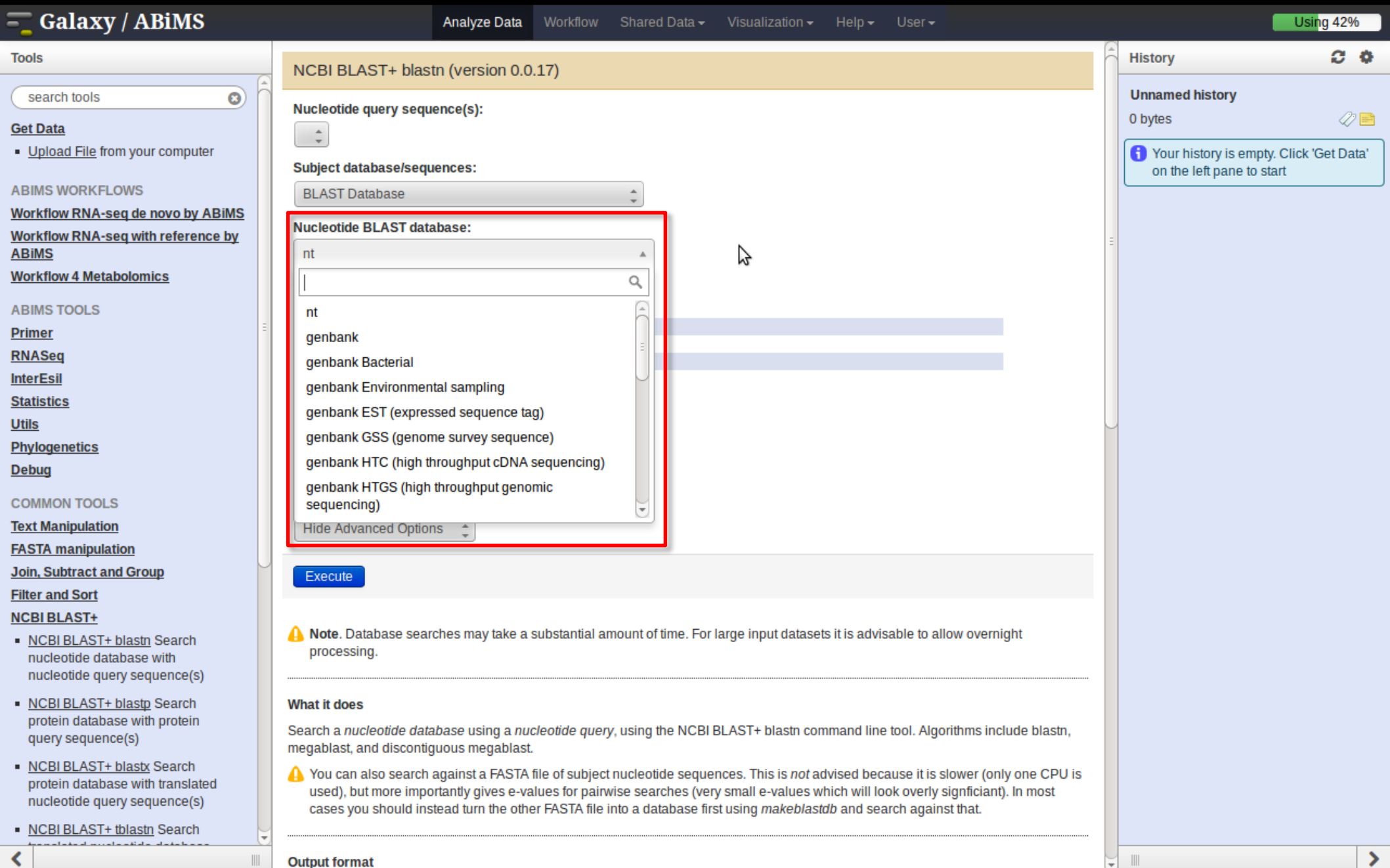
1. First, as you want
2. Then, consider that it is >2 Go

~~DATA IMPORT~~

~~Data import~~

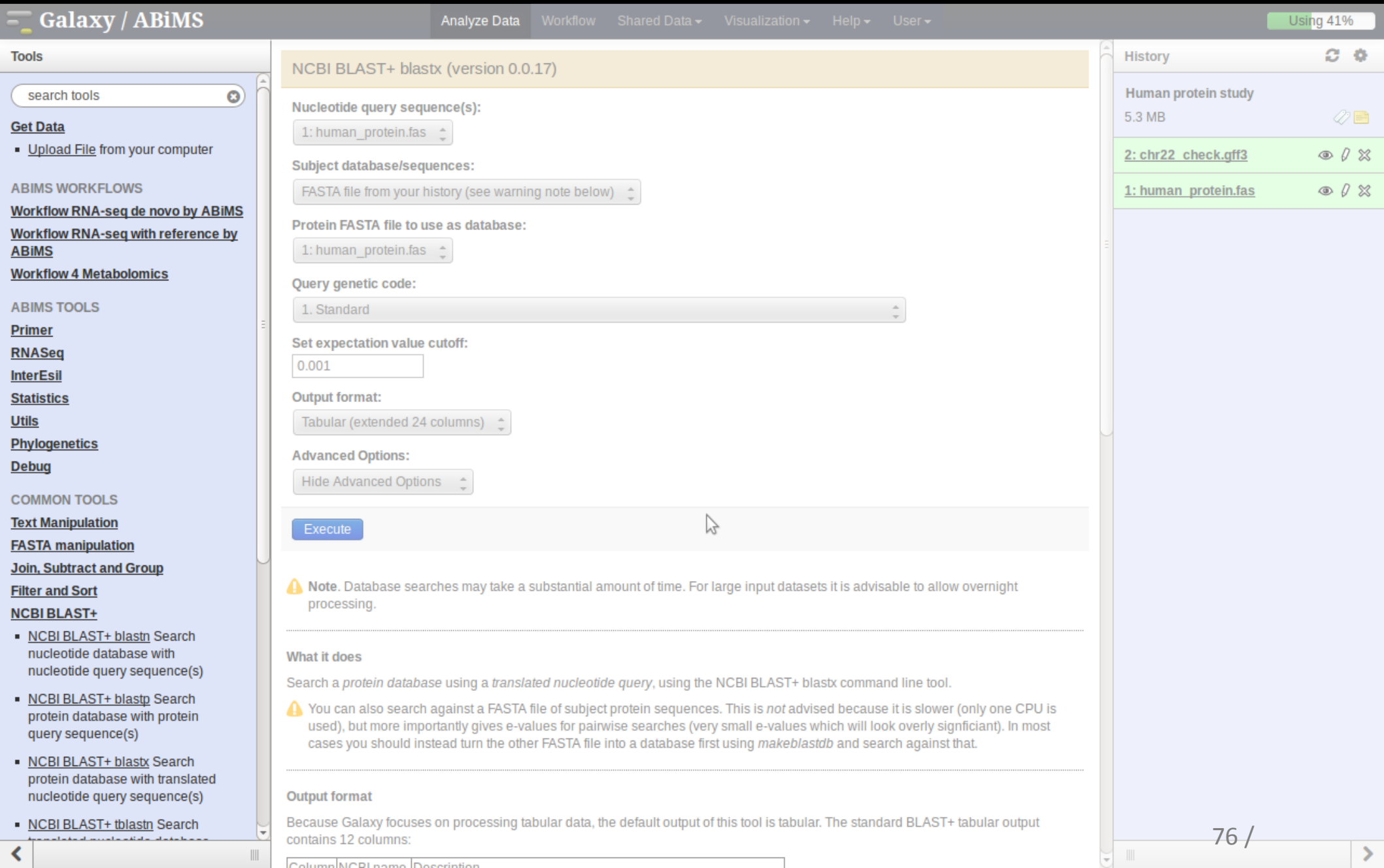
For HUGE public resources: genome, databank …

--> Make a request to the support team

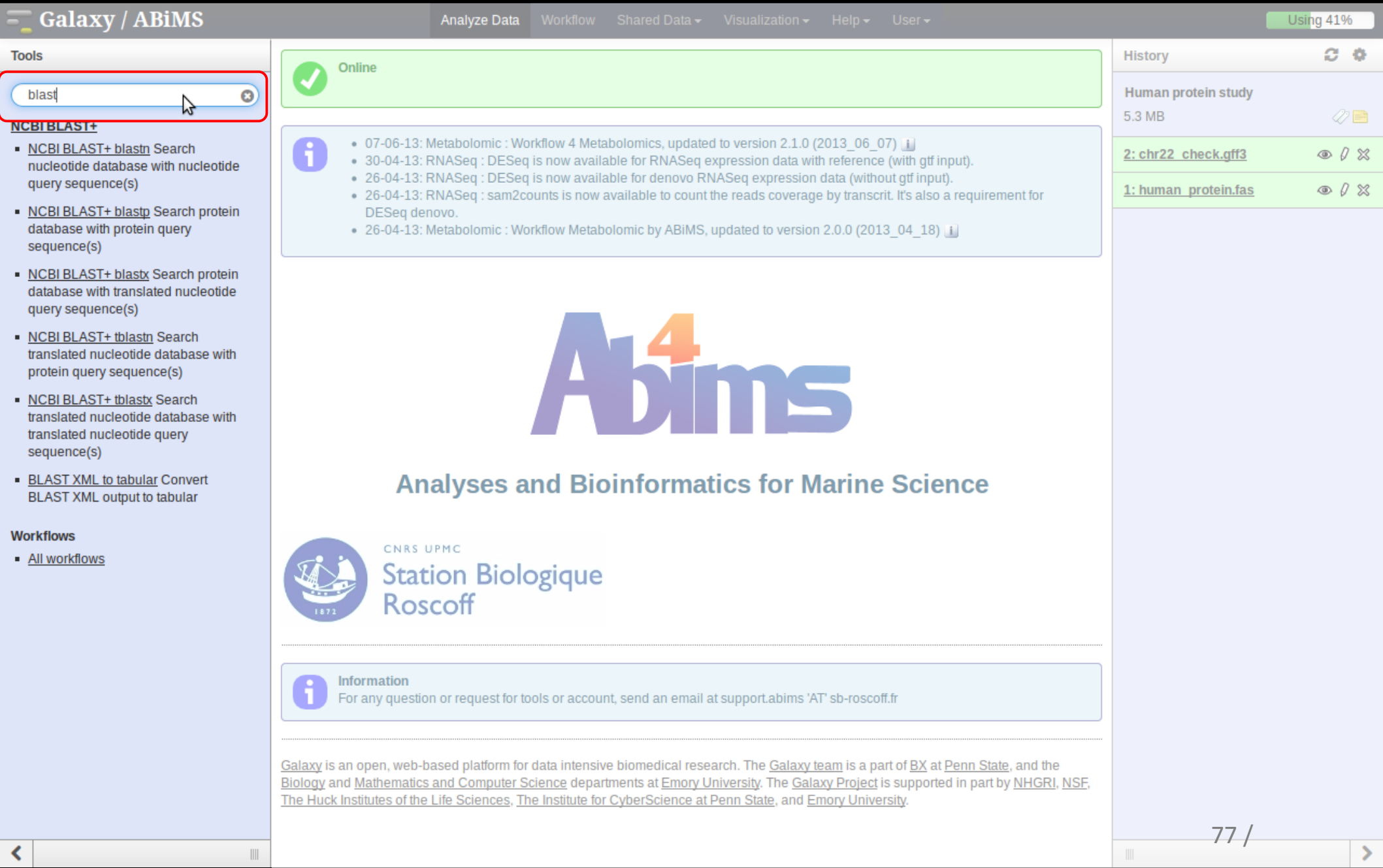


TOOLS

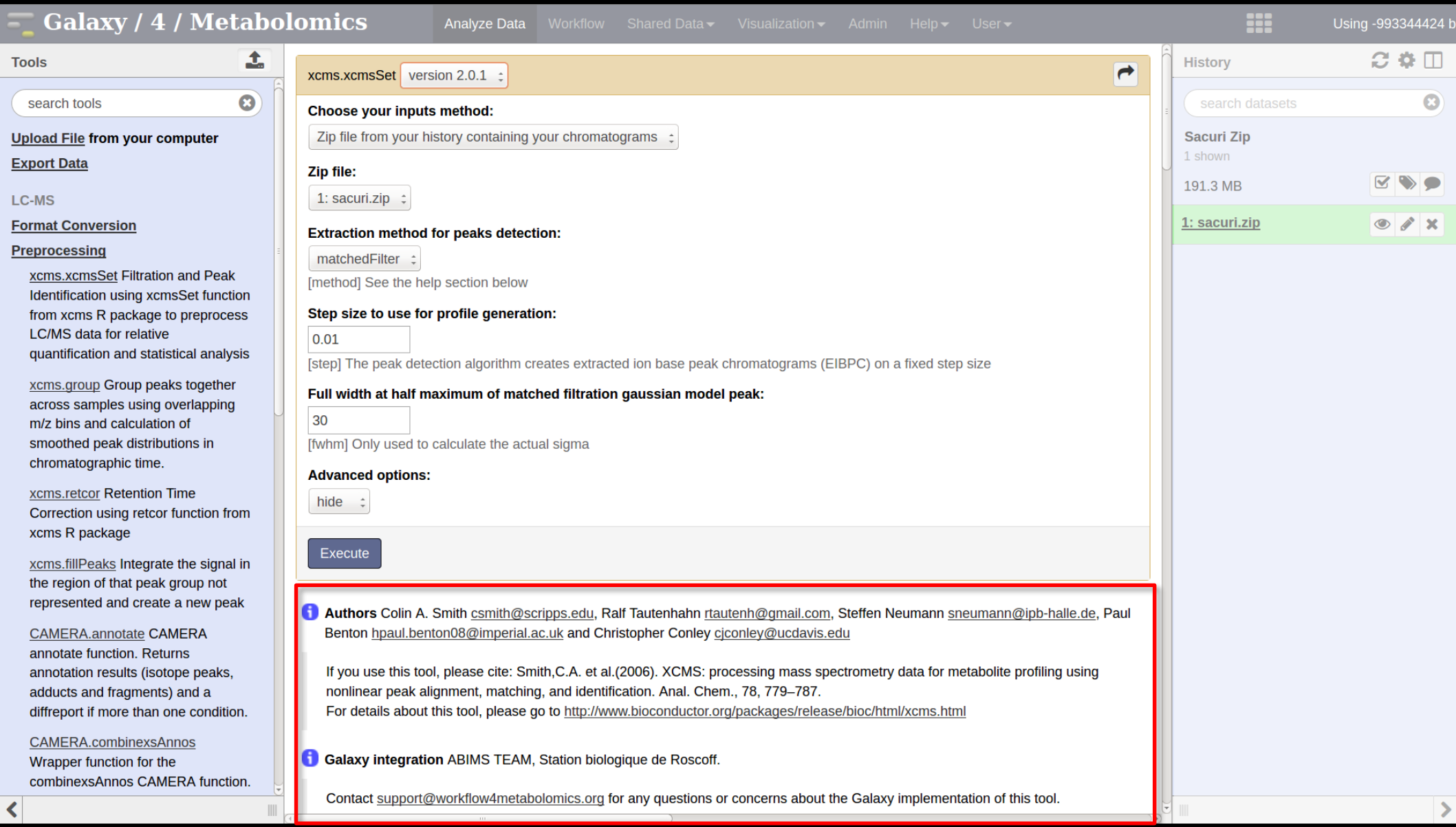
Tools - panel


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 41%
Tools
search tools
Get Data
Upload File from your computer
ABIMS WORKFLOWS
Workflow RNA-seq de novo by ABiMS
Workflow RNA-seq with reference by ABiMS
Workflow 4 Metabolomics
ABIMS TOOLS
Primer
RNASeq
InterEsil
Statistics
Utils
Phylogenetics
Debug
COMMON TOOLS
Text Manipulation
FASTA manipulation
Join, Subtract and Group
Filter and Sort
NCBI BLAST+
NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)
NCBI BLAST+ blastp Search protein database with protein query sequence(s)
NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
NCBI BLAST+ tblastn Search
NCBI BLAST+ blastx (version 0.0.17)
Nucleotide query sequence(s):
1: human_protein.fas
Subject database/sequences:
FASTA file from your history (see warning note below)
Protein FASTA file to use as database:
1: human_protein.fas
Query genetic code:
1. Standard
Set expectation value cutoff:
0.001
Output format:
Tabular (extended 24 columns)
Advanced Options:
Hide Advanced Options
Execute
Note. Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.
What it does
Search a protein database using a translated nucleotide query, using the NCBI BLAST+ blastx command line tool.
You can also search against a FASTA file of subject protein sequences. This is not advised because it is slower (only one CPU is used), but more importantly gives e-values for pairwise searches (very small e-values which will look overly signficiant). In most cases you should instead turn the other FASTA file into a database first using makeblastdb and search against that.
Output format
Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:
History
Human protein study
5.3 MB
2: chr22_check.gff3
1: human_protein.fas

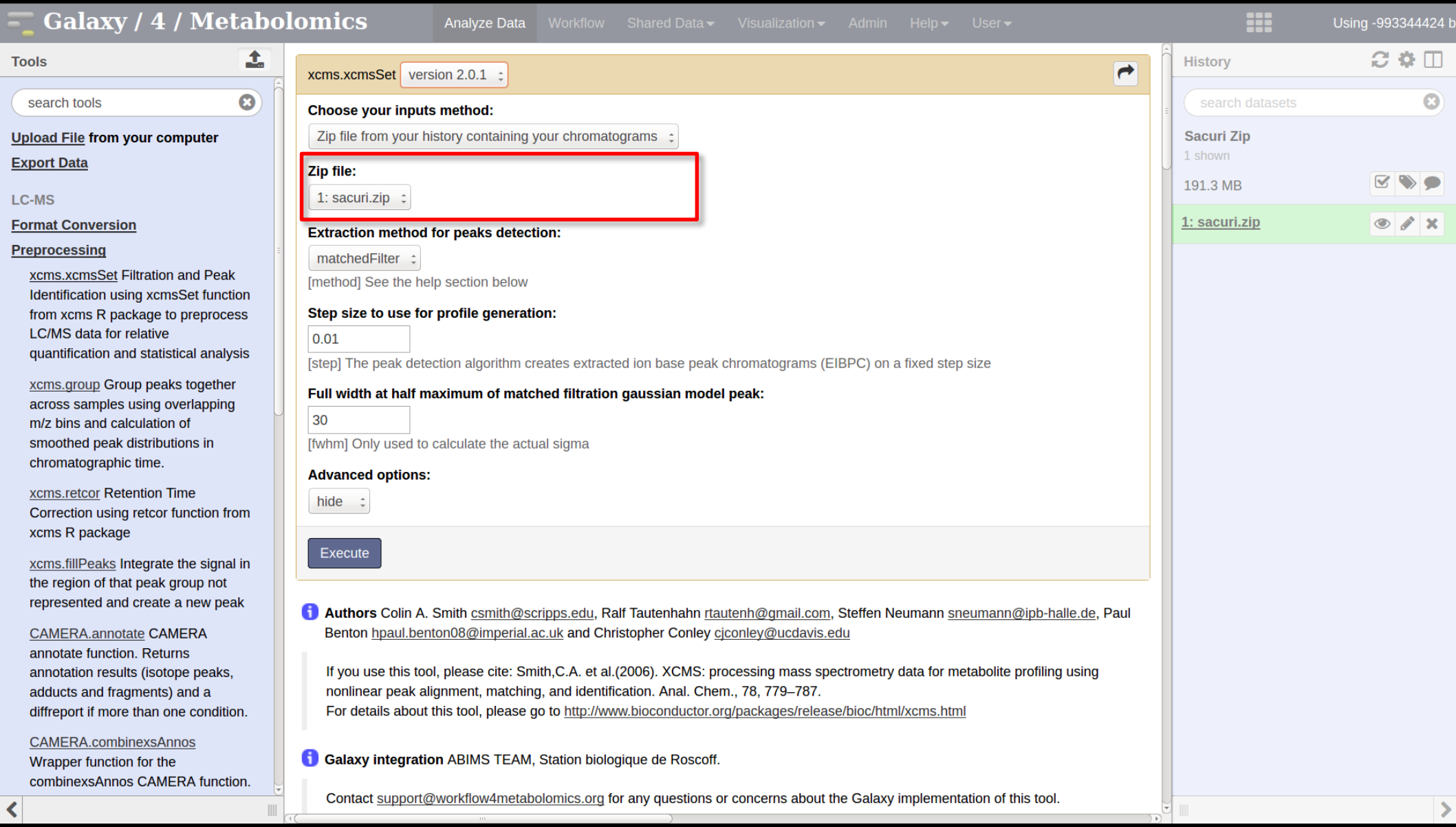
Tools - panel


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 41%
Tools
blast
NCBI BLAST+
NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)
NCBI BLAST+ blastp Search protein database with protein query sequence(s)
NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
NCBI BLAST+ tblastn Search translated nucleotide database with protein query sequence(s)
NCBI BLAST+ tblastx Search translated nucleotide database with translated nucleotide query sequence(s)
BLAST XML to tabular Convert BLAST XML output to tabular
Workflows
All workflows
Online
07-06-13: Metabolomic : Workflow 4 Metabolomics, updated to version 2.1.0 (2013_06_07)
30-04-13: RNASeq : DESeq is now available for RNASeq expression data with reference (with gtf input).
26-04-13: RNASeq : DESeq is now available for denovo RNASeq expression data (without gtf input).
26-04-13: RNASeq : sam2counts is now available to count the reads coverage by transcrit. It's also a requirement for DESeq denovo.
26-04-13: Metabolomic : Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013_04_18)
Analyses and Bioinformatics for Marine Science
CNRS UPMC
Station Biologique
Roscoff
Information
For any question or request for tools or account, send an email at support.abims 'AT' sb-roscoff.fr
Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.
History
Human protein study
5.3 MB
2: chr22_check.gff3
1: human_protein.fas

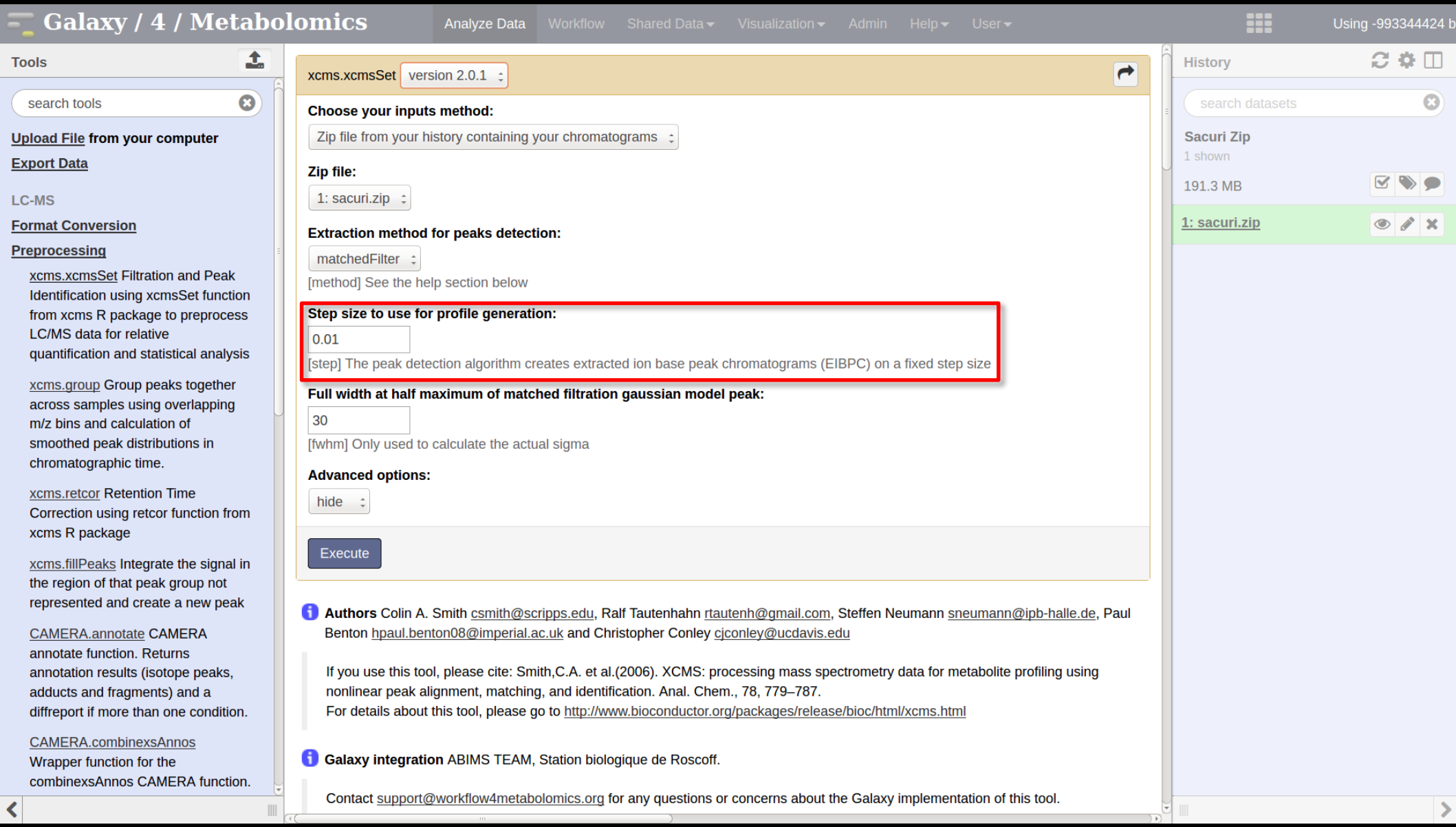
Tools - form


Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
xcms.retcor Retention Time Correction using retcor function from xcms R package
xcms.fillPeaks Integrate the signal in the region of that peak group not represented and create a new peak
CAMERA.annotate CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.
CAMERA.combinexsAnnos Wrapper function for the combinexsAnnos CAMERA function.
xcms.xcmsSet version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
hide
Execute
Authors Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton08@imperial.ac.uk and Christopher Conley cjconley@ucdavis.edu
If you use this tool, please cite: Smith,C.A. et al.(2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
For details about this tool, please go to http://www.bioconductor.org/packages/release/bioc/html/xcms.html
Galaxy integration ABIMS TEAM, Station biologique de Roscoff.
Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
History
search datasets
Sacuri Zip
1 shown
191.3 MB
1: sacuri.zip

Tools - form

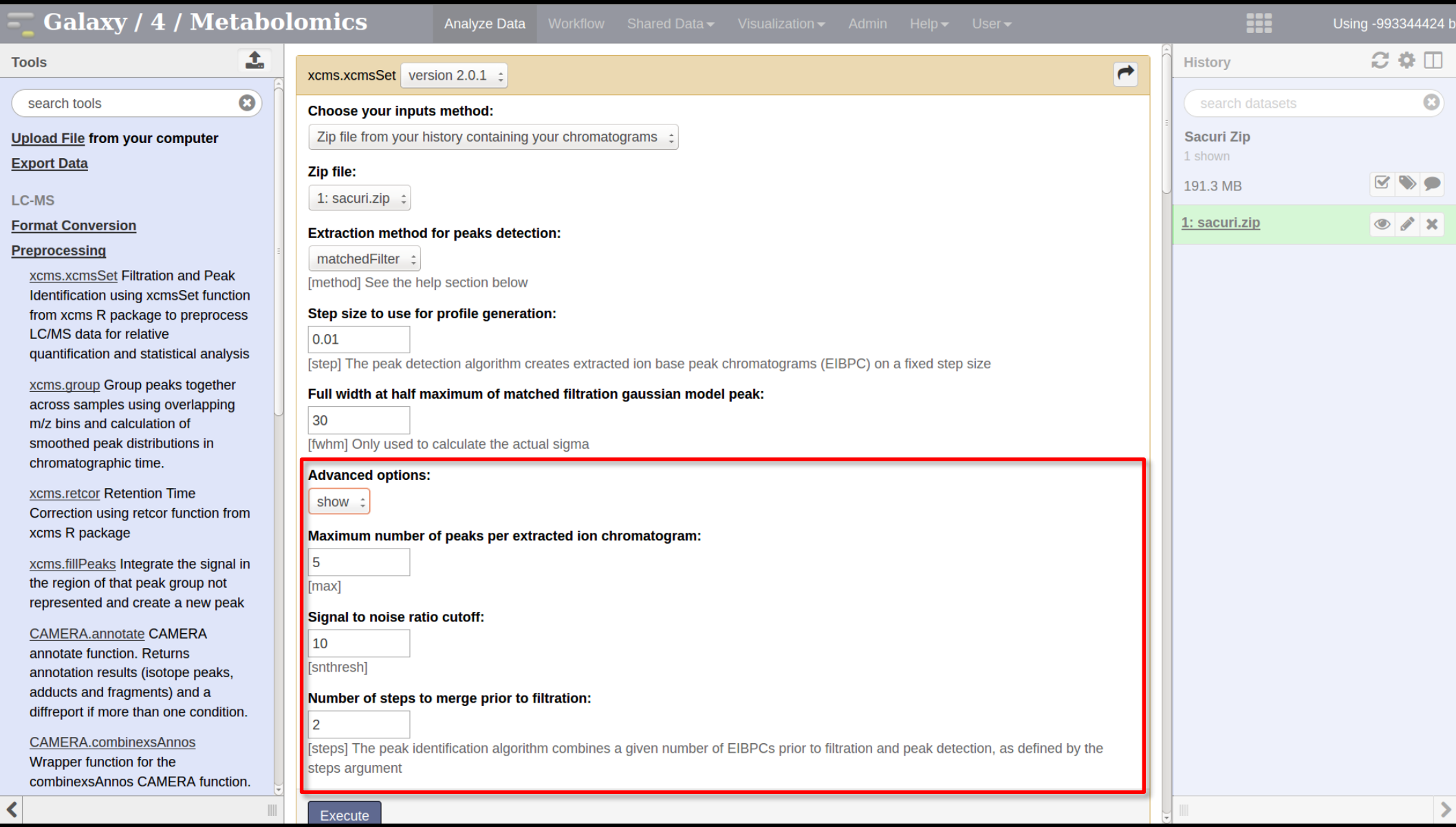

Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
xcms.retcor Retention Time Correction using retcor function from xcms R package
xcms.fillPeaks Integrate the signal in the region of that peak group not represented and create a new peak
CAMERA.annotate CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.
CAMERA.combinexsAnnos Wrapper function for the combinexsAnnos CAMERA function.
xcms.xcmsSet version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
hide
Execute
Authors Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton08@imperial.ac.uk and Christopher Conley cjconley@ucdavis.edu
If you use this tool, please cite: Smith,C.A. et al.(2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
For details about this tool, please go to http://www.bioconductor.org/packages/release/bioc/html/xcms.html
Galaxy integration ABIMS TEAM, Station biologique de Roscoff.
Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
History
search datasets
Sacuri Zip
1 shown
191.3 MB
1: sacuri.zip

Tools - form

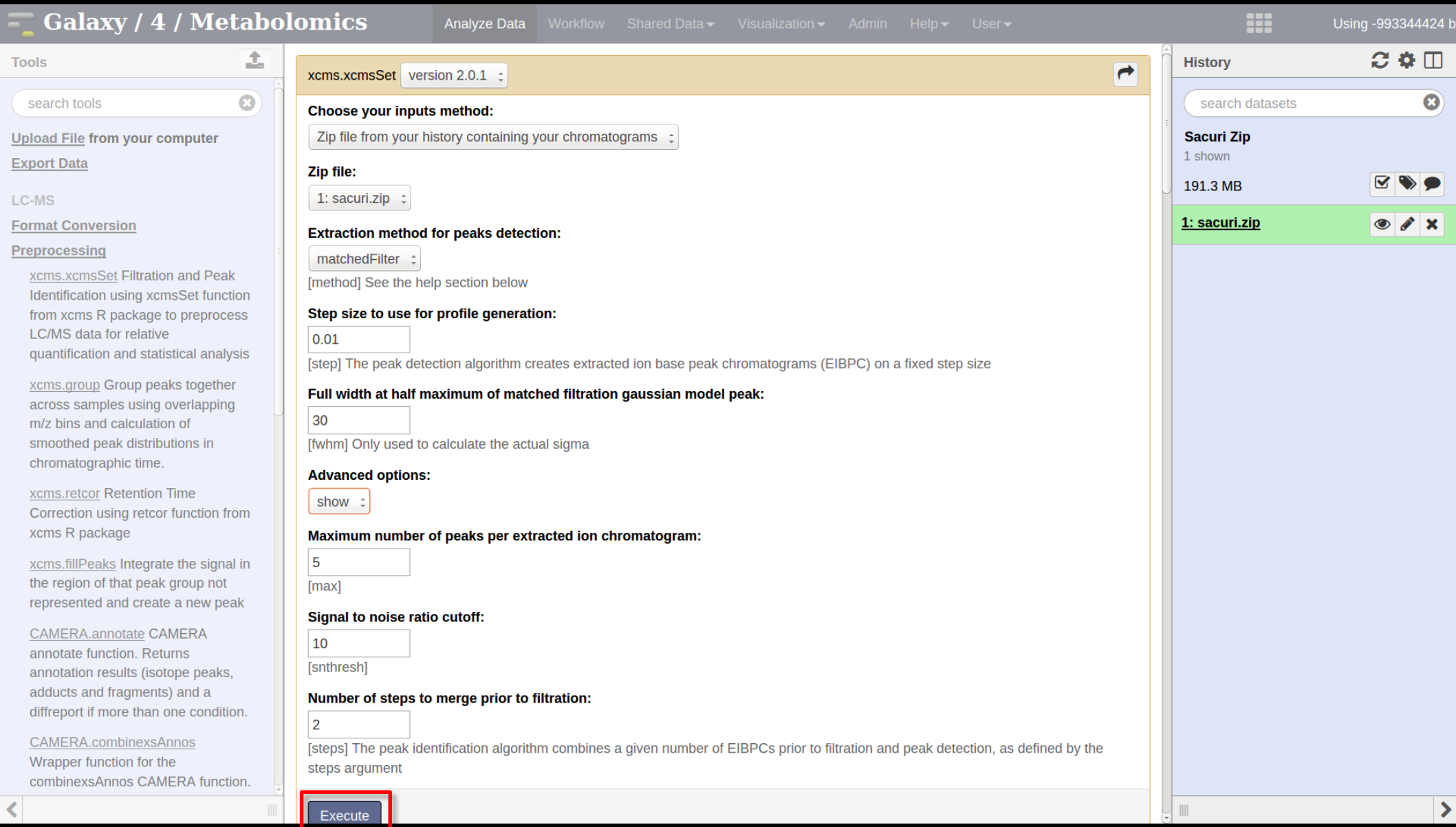

Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
xcms.retcor Retention Time Correction using retcor function from xcms R package
xcms.fillPeaks Integrate the signal in the region of that peak group not represented and create a new peak
CAMERA.annotate CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.
CAMERA.combinexsAnnos Wrapper function for the combinexsAnnos CAMERA function.
xcms.xcmsSet version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
hide
Execute
Authors Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton08@imperial.ac.uk and Christopher Conley cjconley@ucdavis.edu
If you use this tool, please cite: Smith,C.A. et al.(2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
For details about this tool, please go to http://www.bioconductor.org/packages/release/bioc/html/xcms.html
Galaxy integration ABIMS TEAM, Station biologique de Roscoff.
Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
History
search datasets
Sacuri Zip
1 shown
191.3 MB
1: sacuri.zip

Tools - form

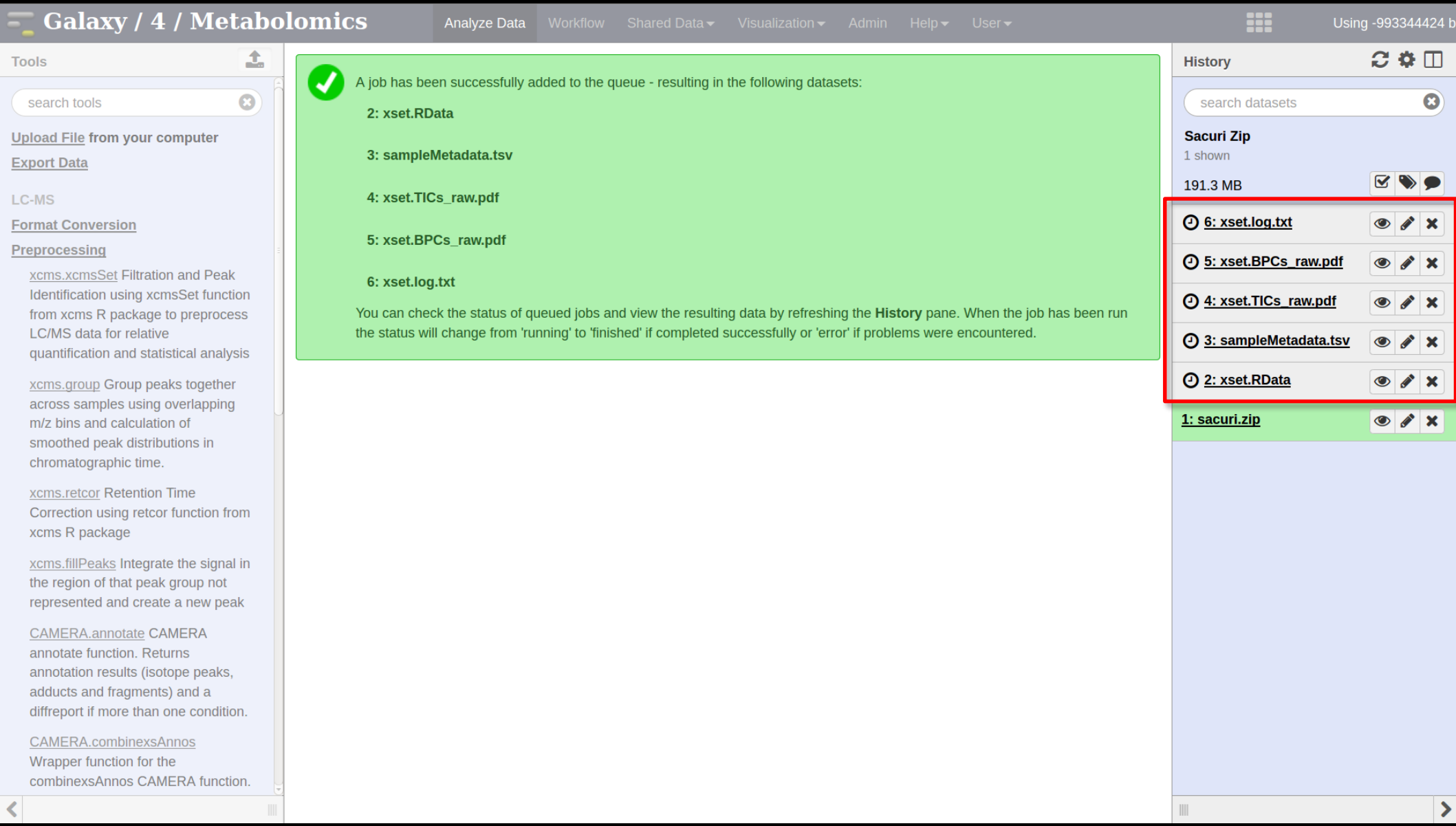
Tools can have some advanced options



Tools - form


Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
xcms.retcor Retention Time Correction using retcor function from xcms R package
xcms.fillPeaks Integrate the signal in the region of that peak group not represented and create a new peak
CAMERA.annotate CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.
CAMERA.combinexsAnnos Wrapper function for the combinexsAnnos CAMERA function.
xcms.xcmsSet version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
show
Maximum number of peaks per extracted ion chromatogram:
5
[max]
Signal to noise ratio cutoff:
10
[snthresh]
Number of steps to merge prior to filtration:
2
[steps] The peak identification algorithm combines a given number of EIBPCs prior to filtration and peak detection, as defined by the steps argument
Execute
History
search datasets
Sacuri Zip
1 shown
191.3 MB
1: sacuri.zip

Tools - form


Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
xcms.xcmsSet Filtration and Peak Identification using xcmsSet function from xcms R package to preprocess LC/MS data for relative quantification and statistical analysis
xcms.group Group peaks together across samples using overlapping m/z bins and calculation of smoothed peak distributions in chromatographic time.
xcms.retcor Retention Time Correction using retcor function from xcms R package
xcms.fillPeaks Integrate the signal in the region of that peak group not represented and create a new peak
CAMERA.annotate CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.
CAMERA.combinexsAnnos Wrapper function for the combinexsAnnos CAMERA function.
A job has been successfully added to the queue - resulting in the following datasets:
2: xset.RData
3: sampleMetadata.tsv
4: xset.TICs_raw.pdf
5: xset.BPCs_raw.pdf
6: xset.log.txt
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.
History
search datasets
Sacuri Zip
1 shown
191.3 MB
6: xset.log.txt
5: xset.BPCs_raw.pdf
4: xset.TICs_raw.pdf
3: sampleMetadata.tsv
2: xset.RData
1: sacuri.zip

Tools - Job status

- Status

Job is waiting to run

= the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs or on the requested number of processors

Tools - Job status

- Status

Job is currently running

= the job is being executed on the computing cluster

Duration time of this status depends completely on the job's attributes and the computing ressources allocated.

Some programs are executed with several processors

(using 4, 8 or 16 Gb of RAM).

And others are mono-threaded☹

Tools - Job status

- Status

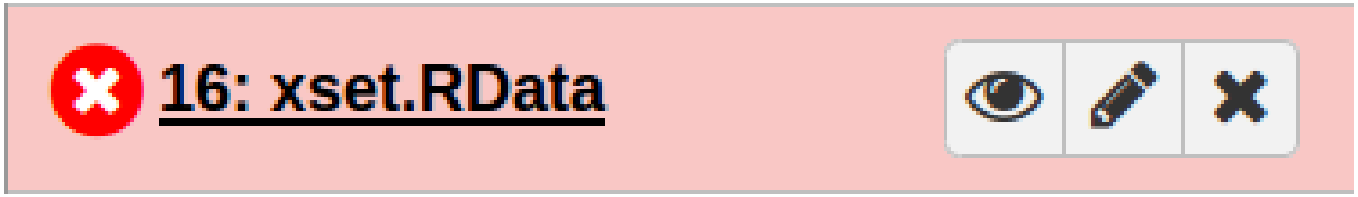
Job is finished

It’s status is OK

but warnings or errors can be hidden behind. Ah hum !

Tools - Job status

- Status



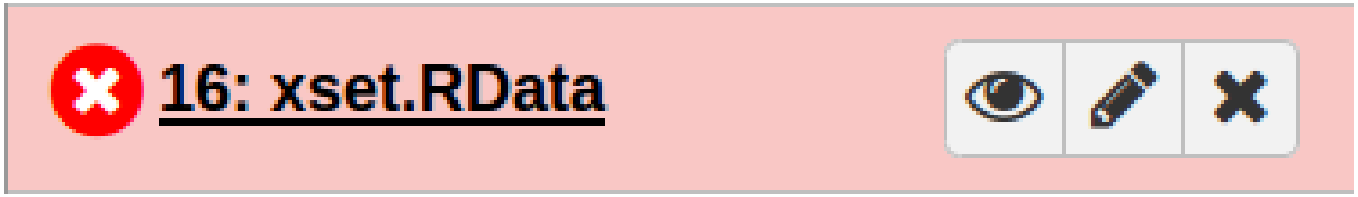
Job is finished but with an error status

= the program sends an error

The error is often explained by the program and sometimes … not.

Tools - Job status

- Status



Job is finished but with an error status

Error causes :

- ~~The user~~ :P
- Bad usage : input file, format or option
- Wrong porting of the program through Galaxy … sorry :/
- Non anticipated crash of the program

Exercise

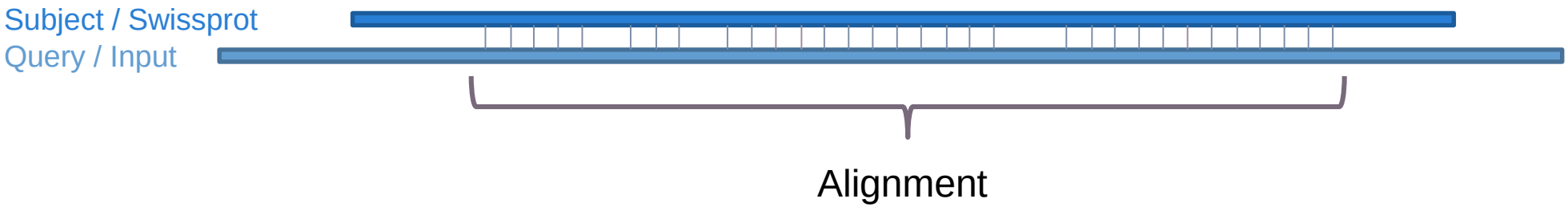
TOOLS

Tools - Exercise

- Aim of this Exercise:
  - Get blast hits with :
    - Identity > 75%
    - Alignment coverage > 75%
  - Execute and chain example of little Galaxy friendly tools together.

Tools - Exercise

- Aim of this Exercise:
  - Get blast hits with :
    - Identity > 75%
    - Alignment coverage > 75%



Coverage = alignment length / query length

Tools - Exercise

- Fetch these two files (**< 2Go**)
  - Link1:

    http://tinyurl.com/GI-Roscoff1
  - Link2:

    http://tinyurl.com/GI-Roscoff2

- Check their contents and datatypes through Galaxy.

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »
  - help : join with column 1 and column 1

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »
  - help : join with column 1 and column 1
- Step 3: use the tool « Compute an expression… »

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »
  - help : join with column 1 and column 1
- Step 3: use the tool « Compute an expression… »
  - help : ((c8-c7+1)/c14)*100

Tools - Exercise

Get blast hits with :

- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »
  - help : join with column 1 and column 1
- Step 3: use the tool « Compute an expression… »
  - help : ((c8-c7+1)/c14)*100
- Step 4: use the tool « Filter »

Tools - Exercise

Get blast hits with :

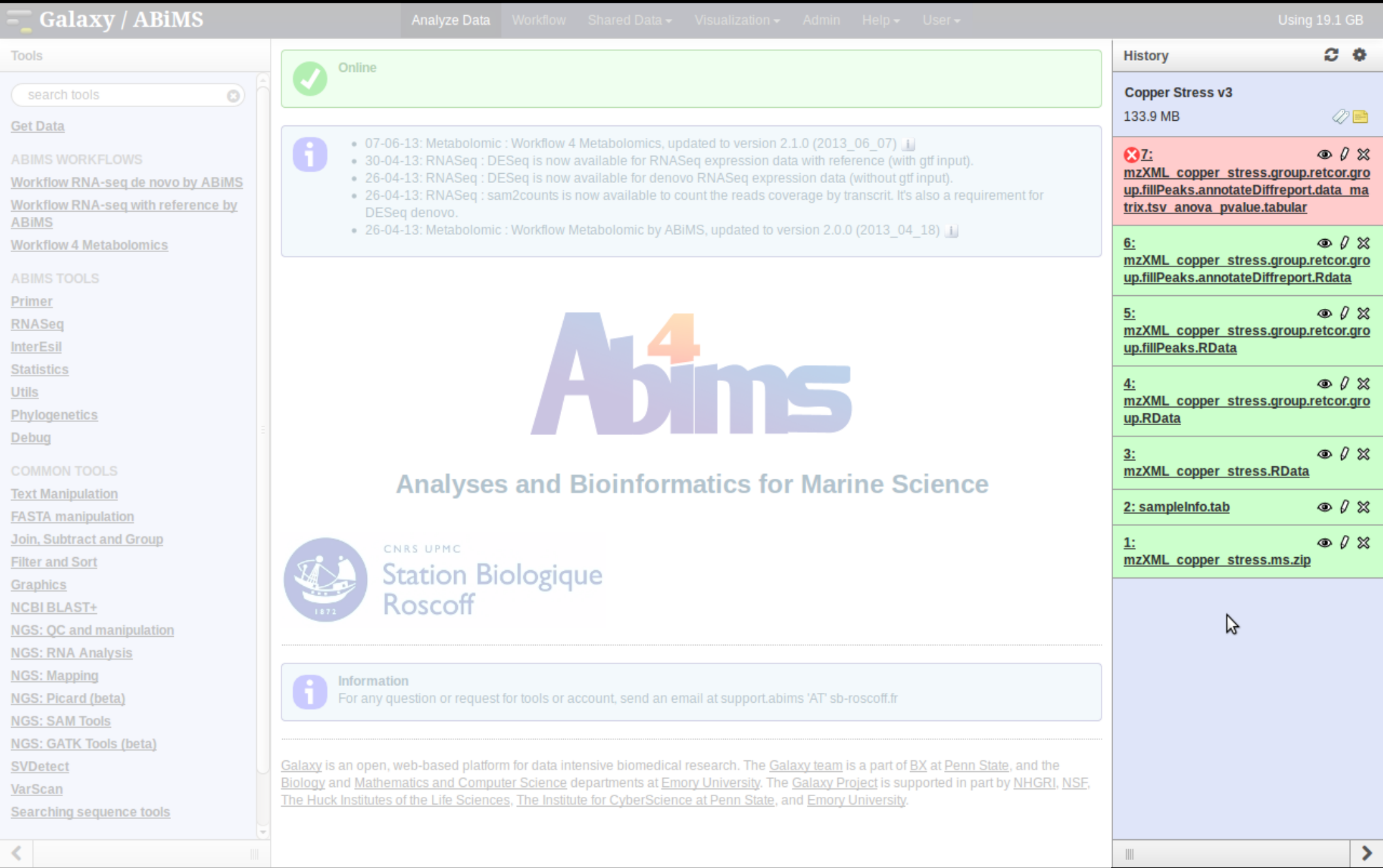
- Identity > 75%
- Alignment coverage > 75%

- Step 1: use the tool « Compute sequence length »
- Step 2: use the tool « Join Two Datasets »
  - help : join with column 1 and column 1
- Step 3: use the tool « Compute an expression… »
  - help : ((c8-c7+1)/c14)*100
- Step 4: use the tool « Filter »
  - help : c3>=75 and c15>=75

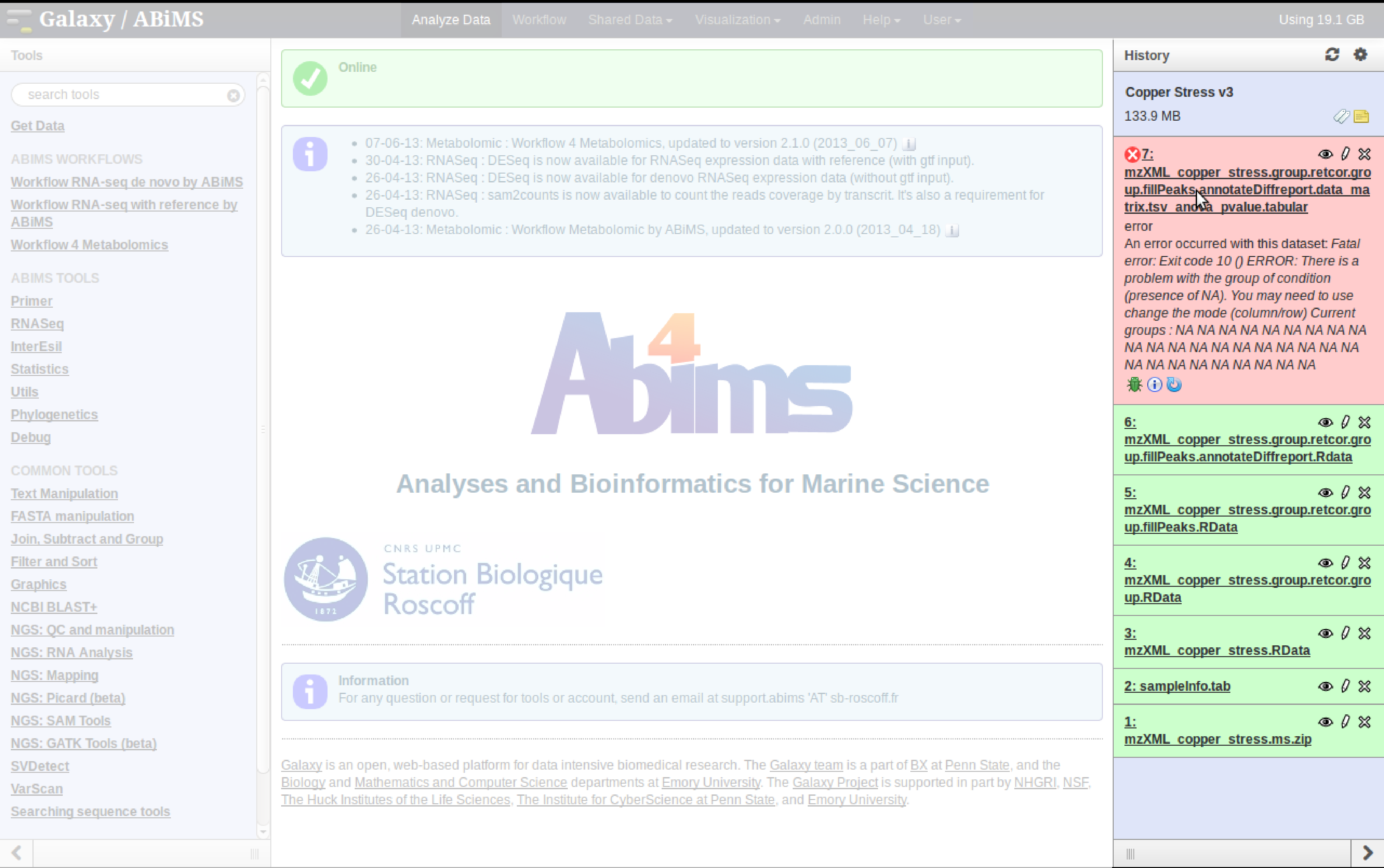
Part II

TOOLS

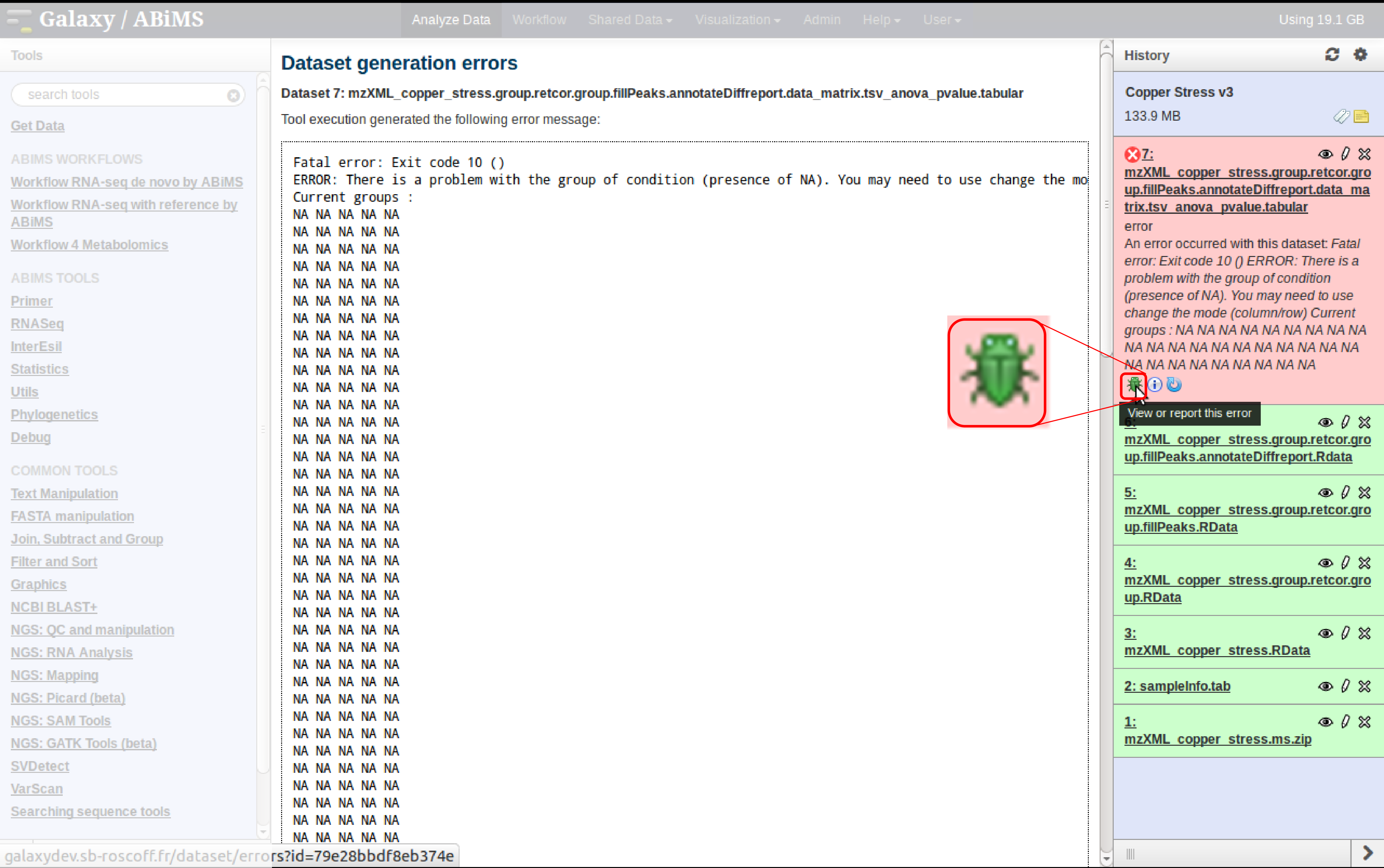
Tools – Handle errors



Tools – Handle errors

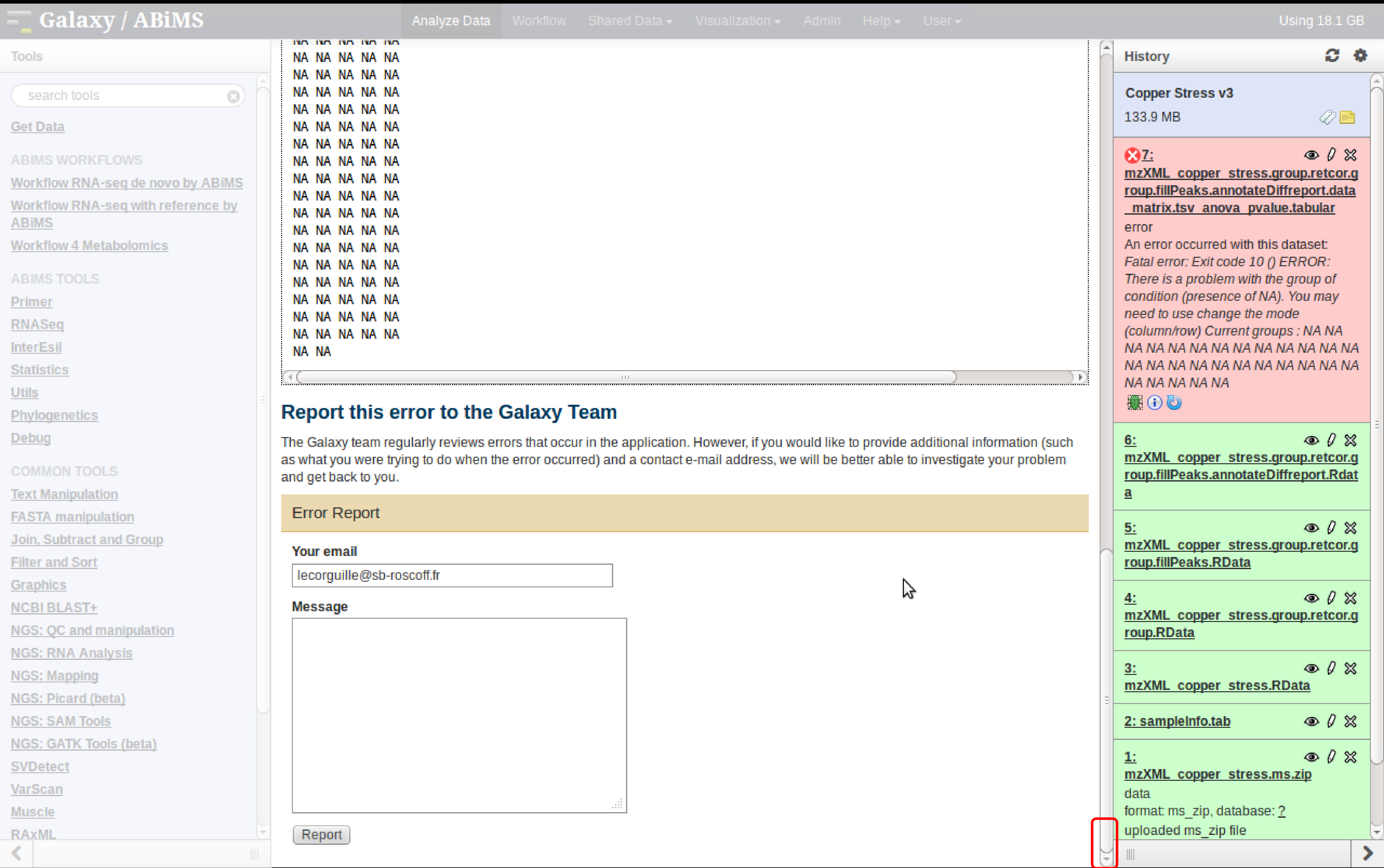


Tools – Handle errors



Tools – Handle errors

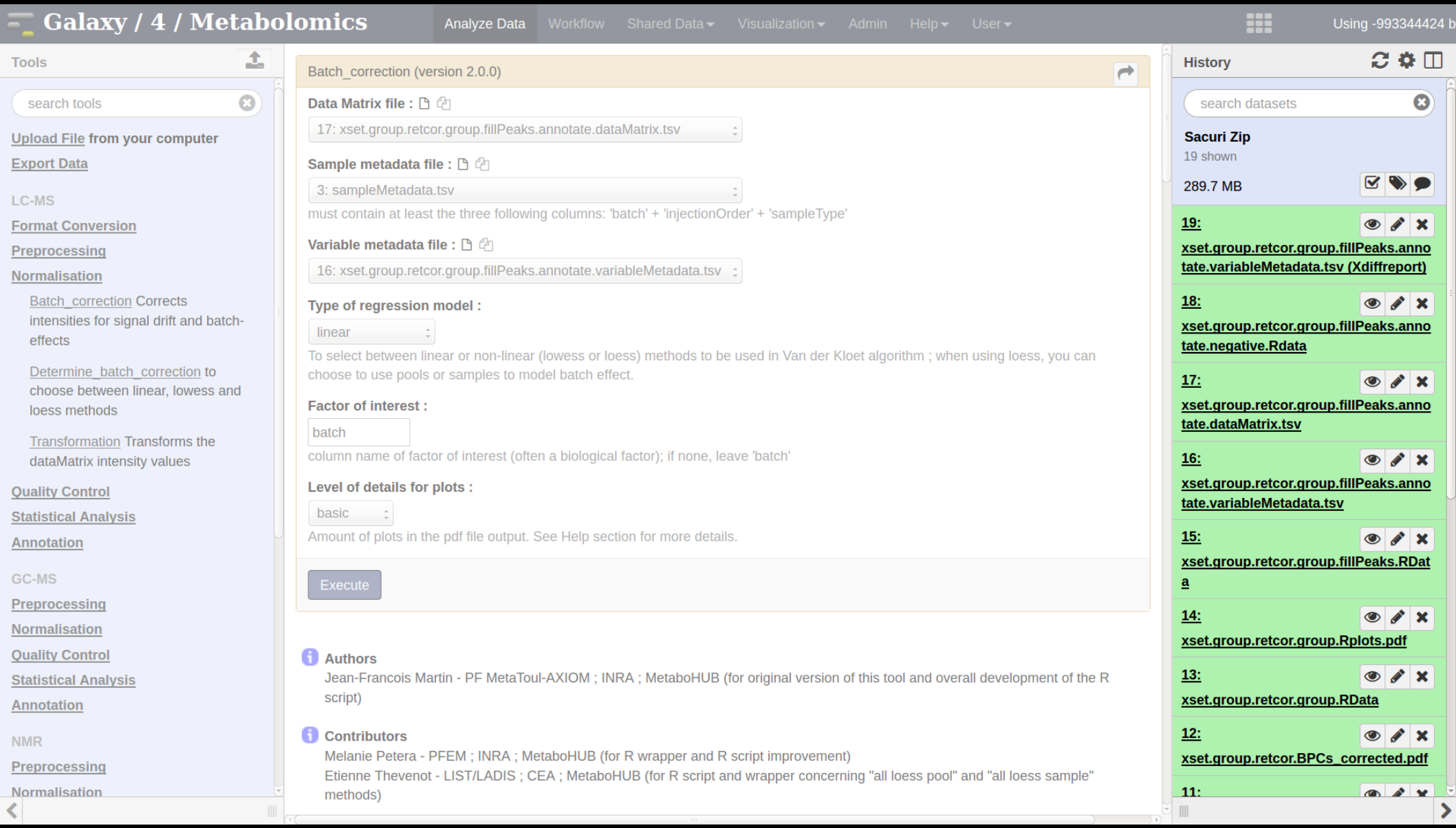
Sent to the support team



HISTORY

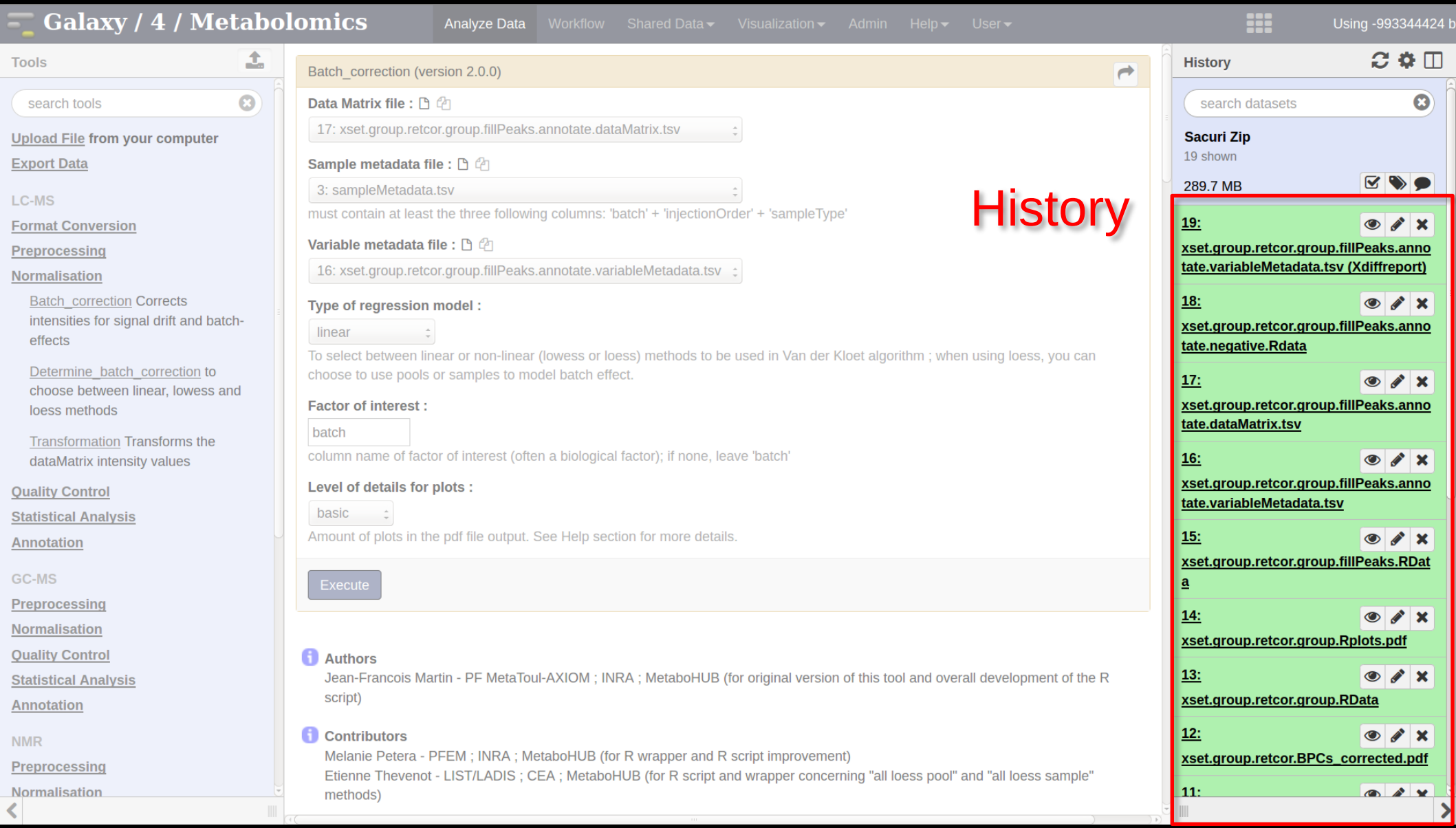
History panel

Both inputs and outputs



History panel

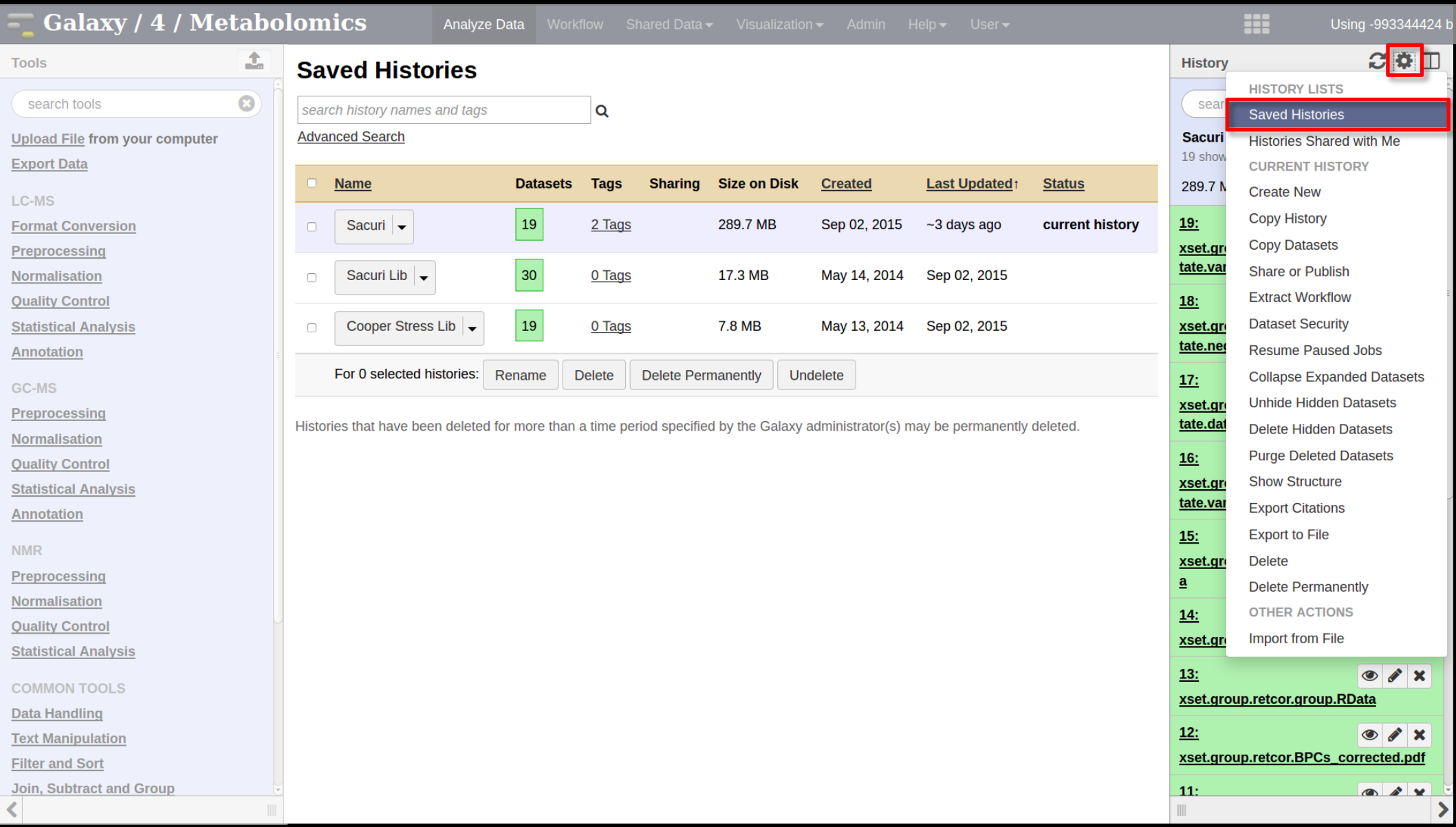
Both inputs and outputs



History panel renaming and annotation

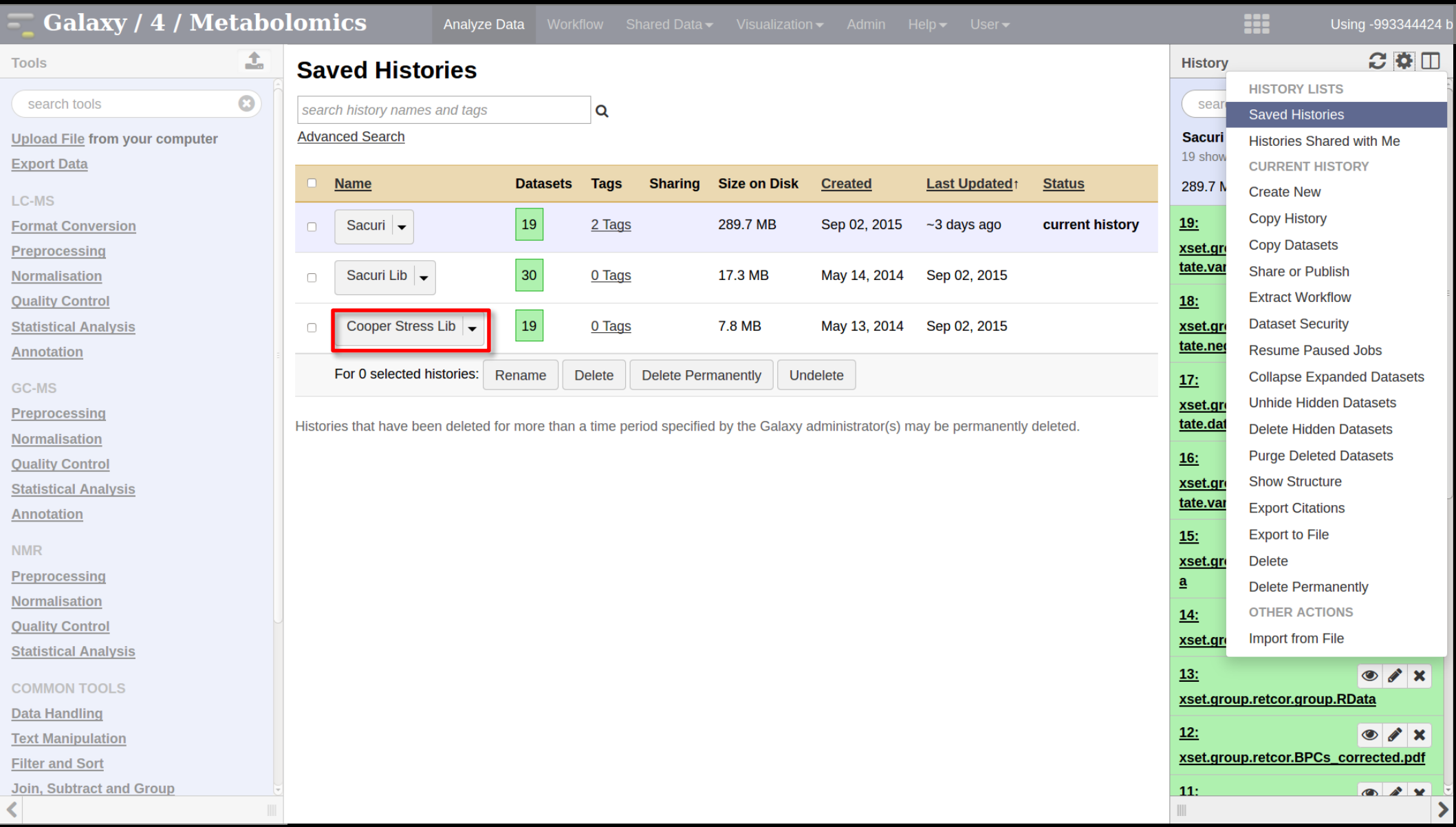
History panel

Saved histories: Rename, Delete, Delete Permanently



History panel

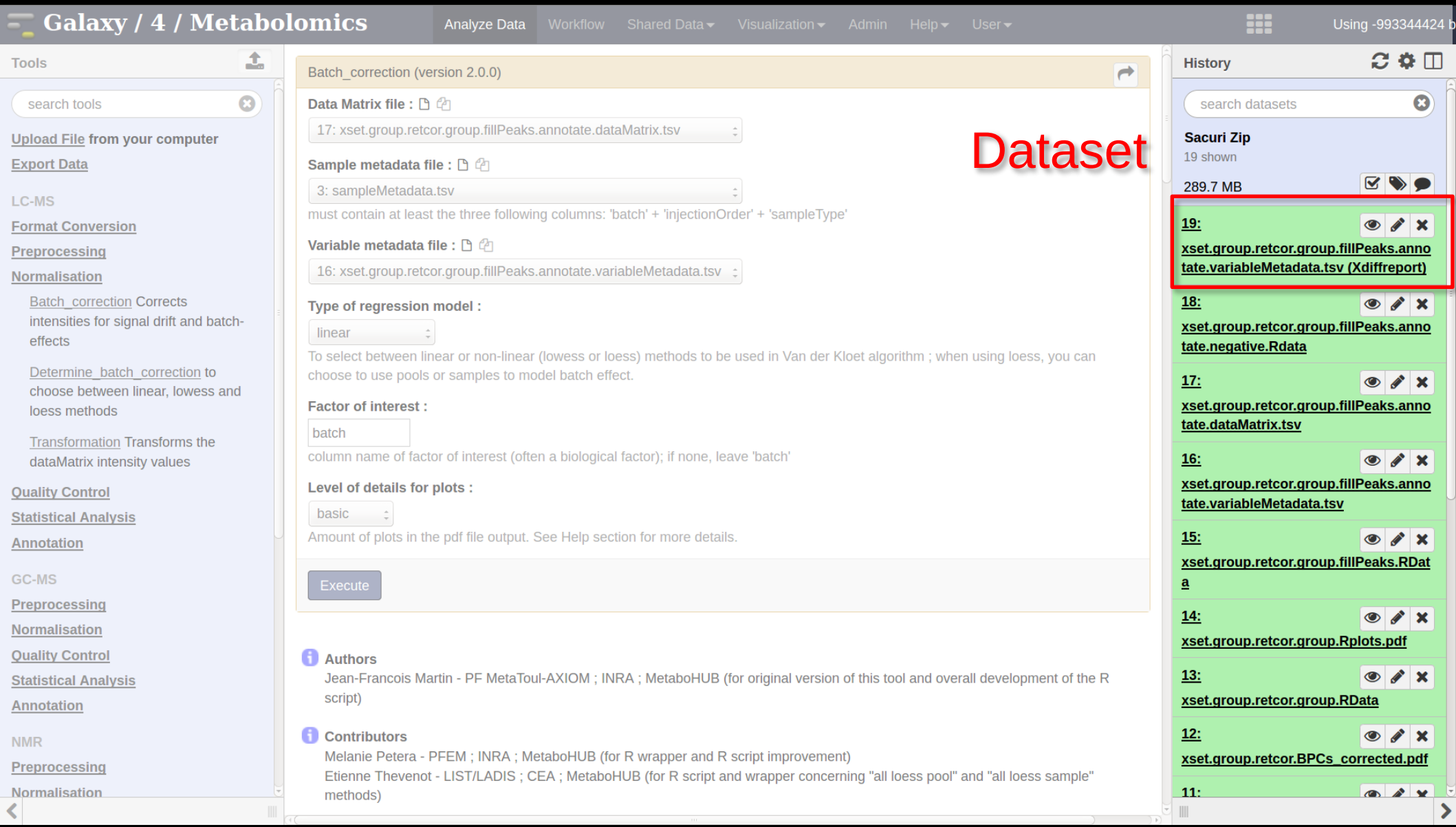
Saved histories: Switch histories



DATASET

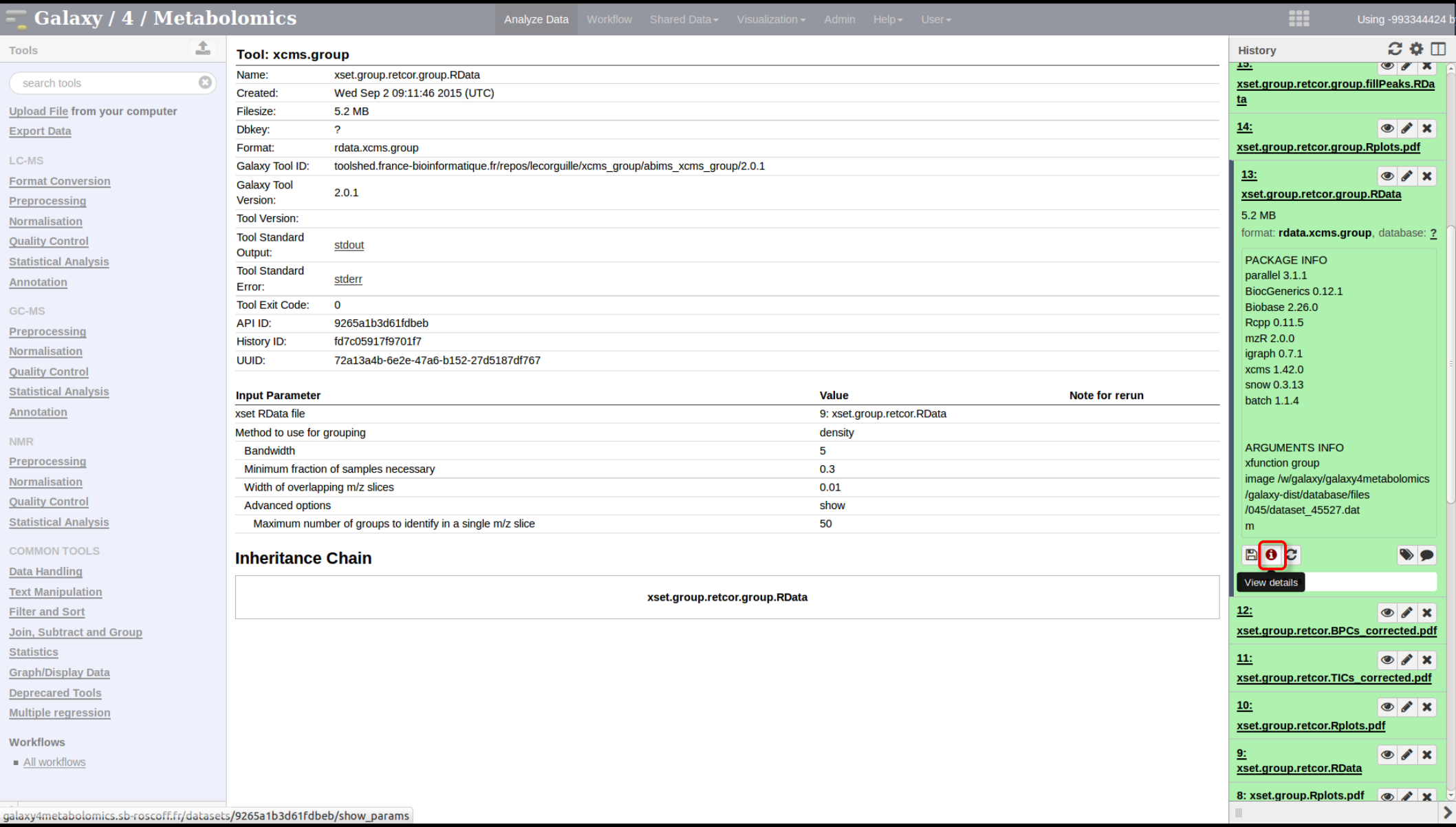
Dataset

Both inputs and outputs



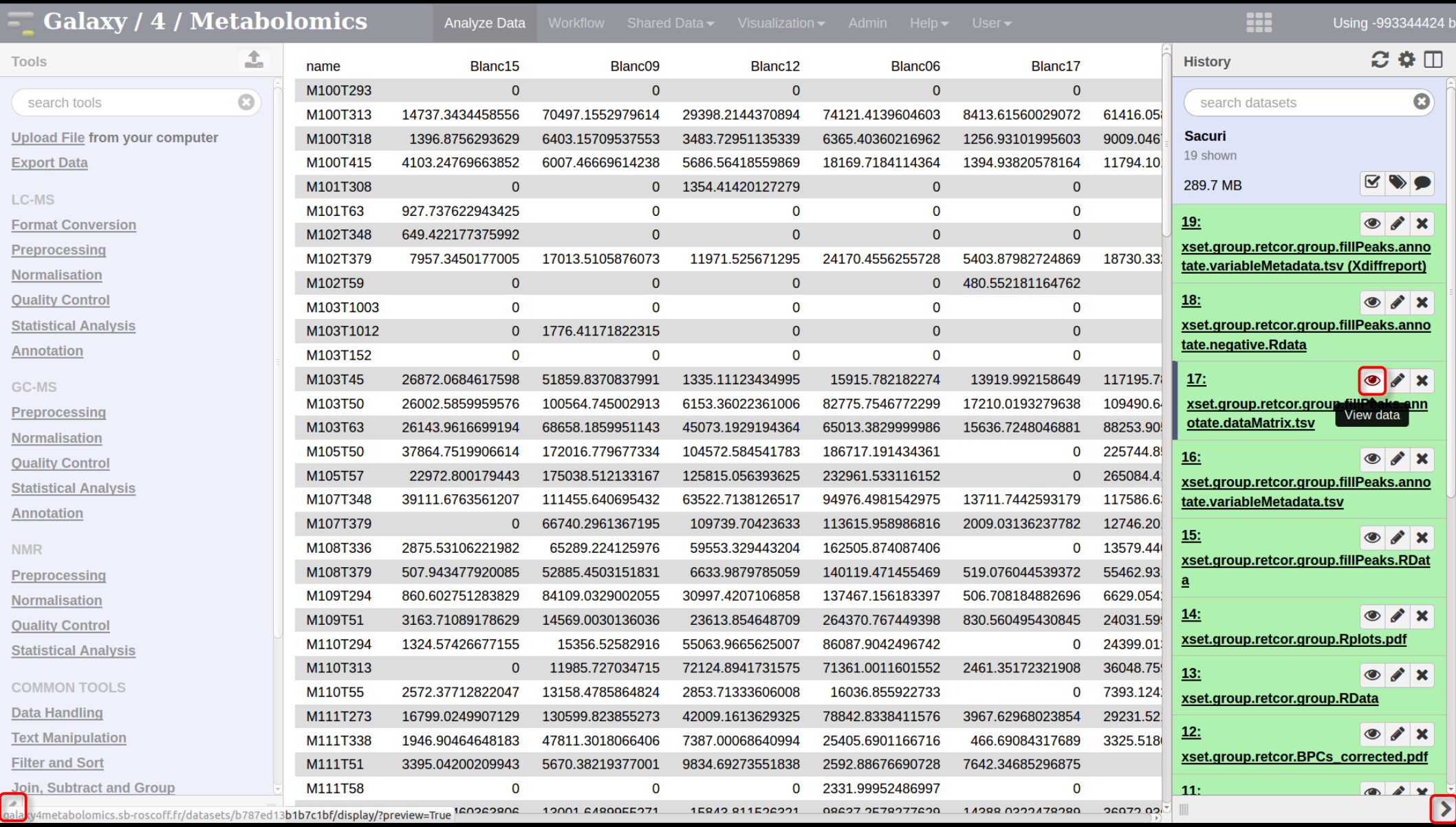
Dataset

Informations



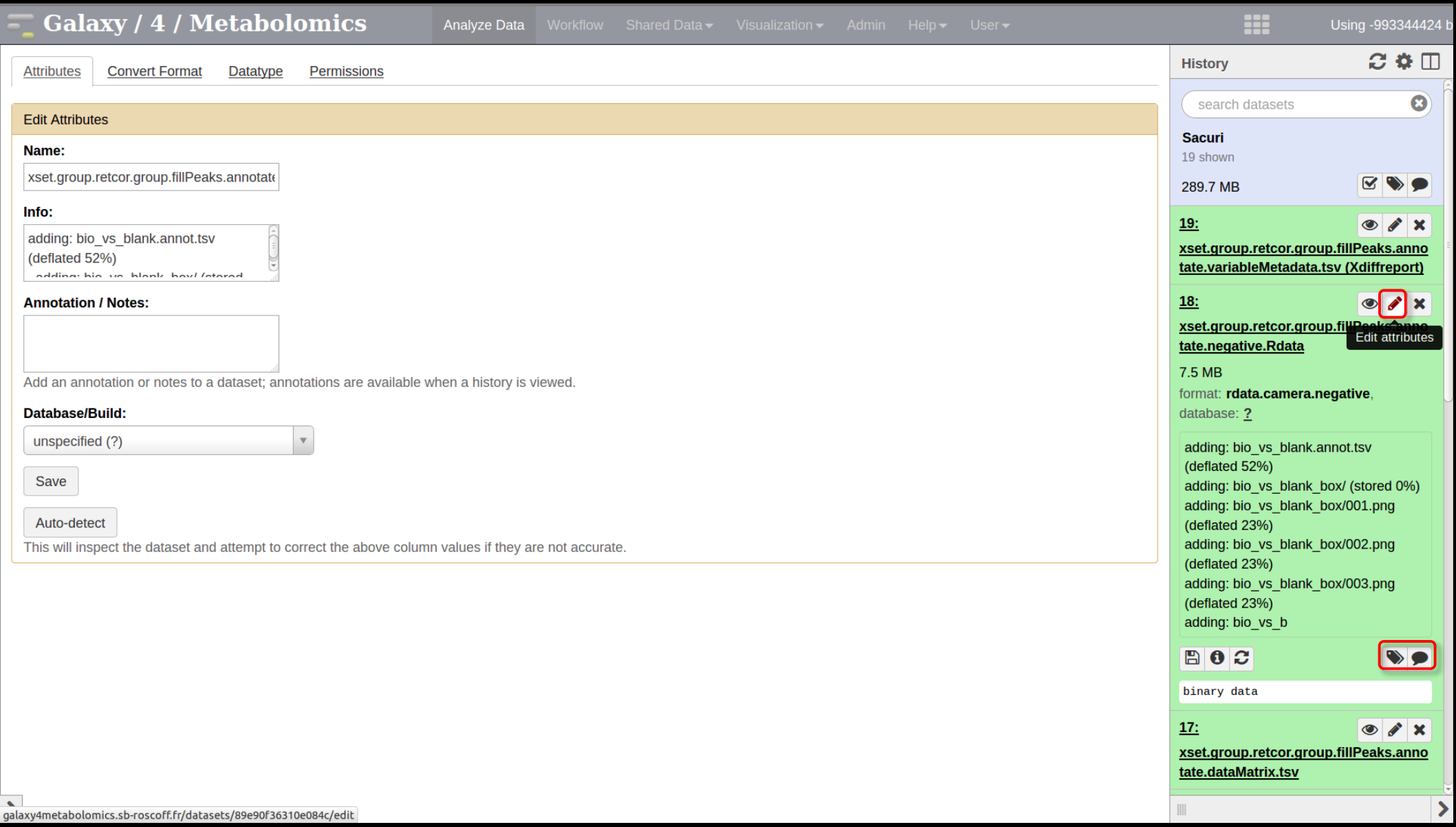
Dataset

Dataset display : text, tabular, pdf, picture, html …



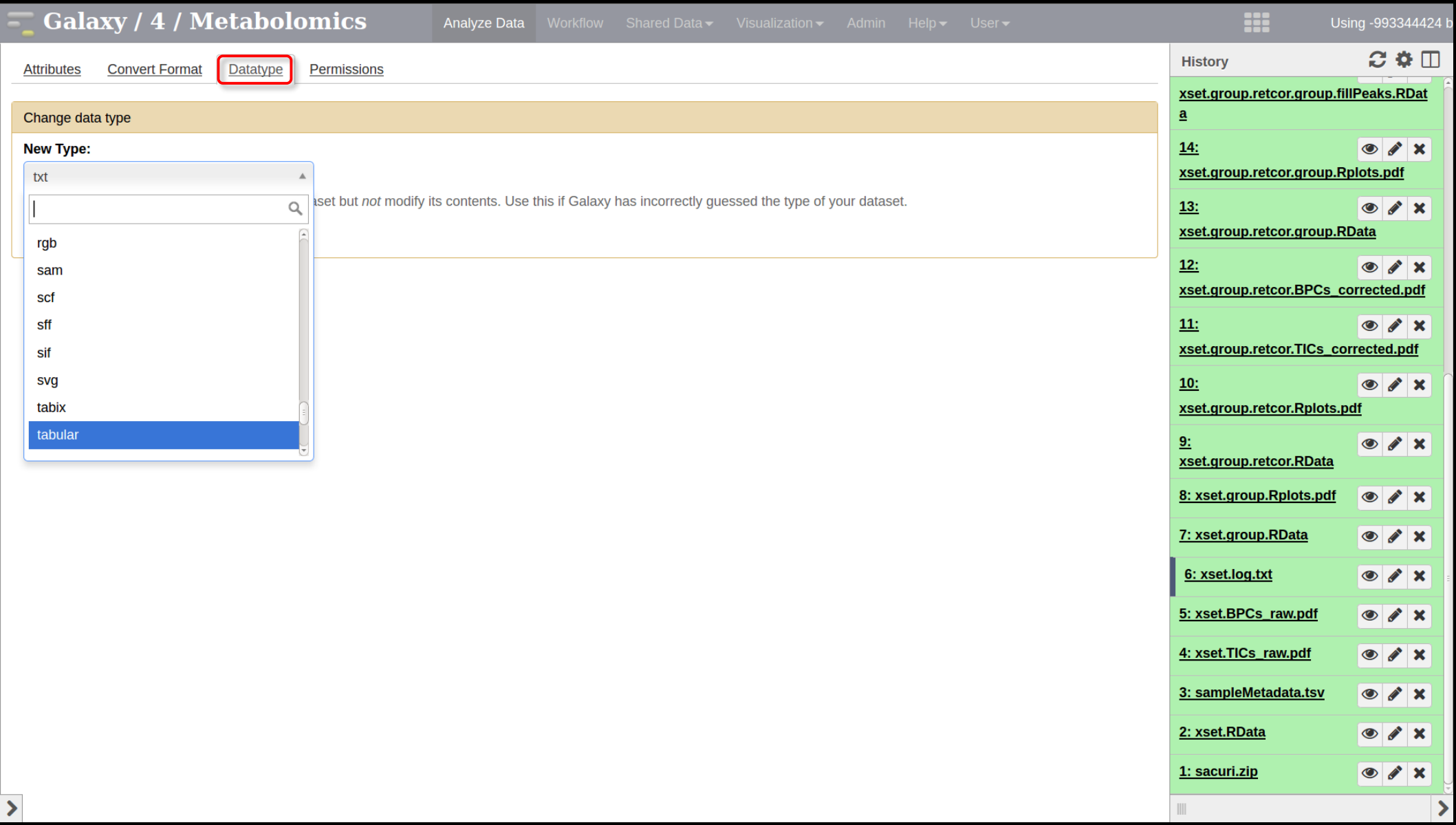
Dataset

Renaming and annotation



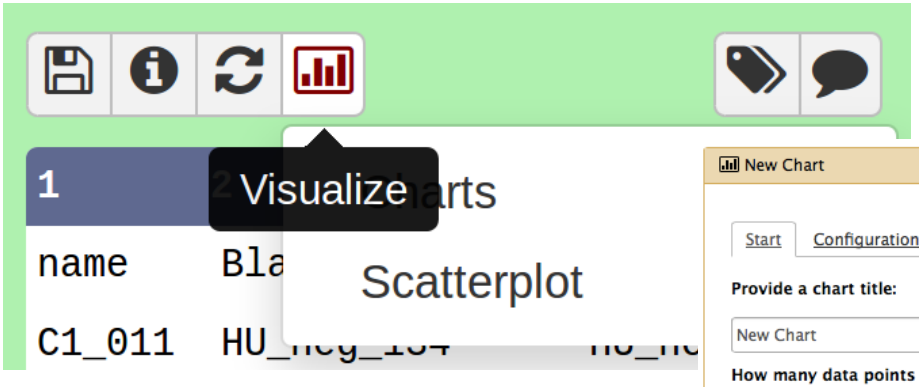
Dataset

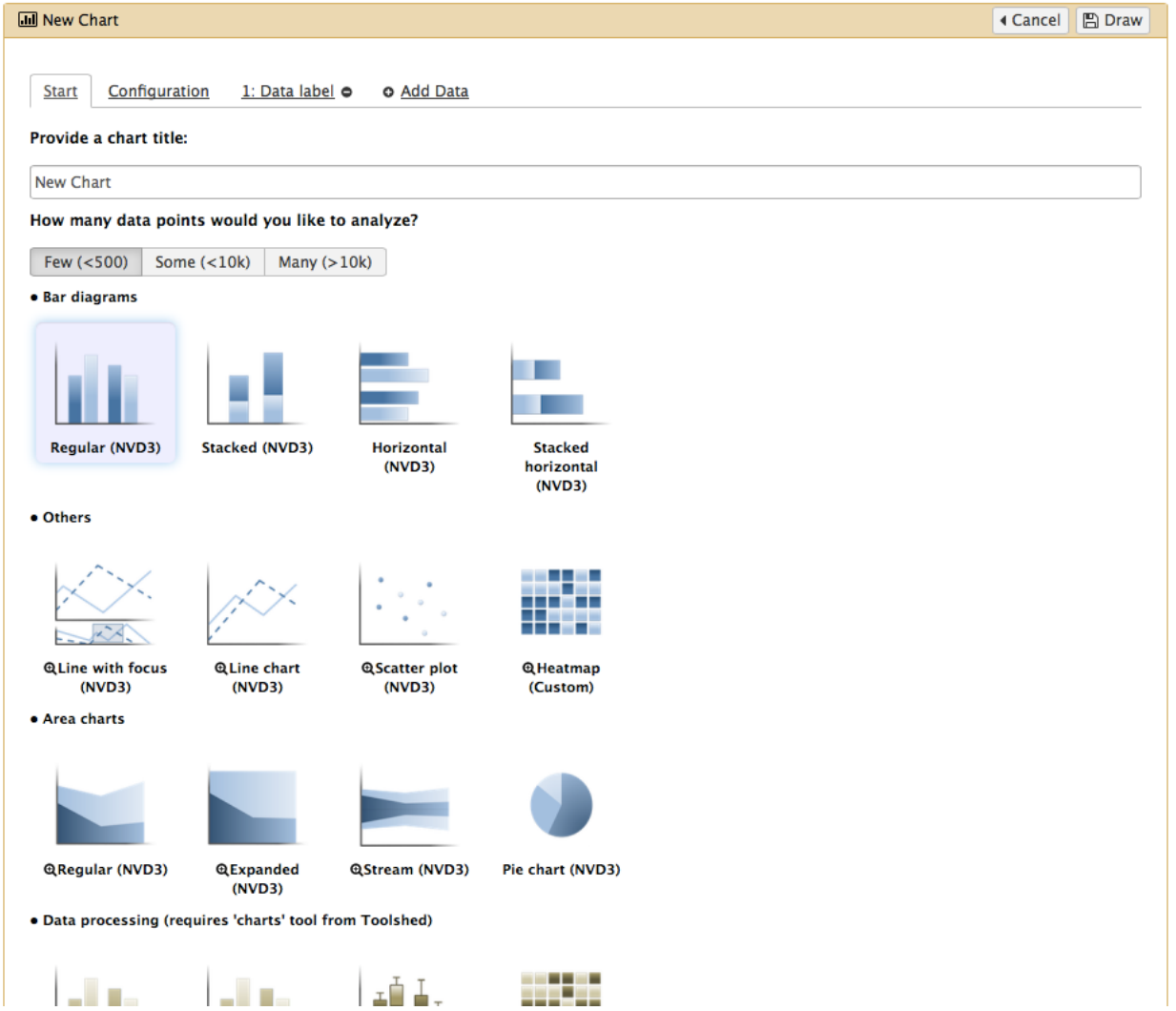
Change the Datatype of the Dataset



Dataset

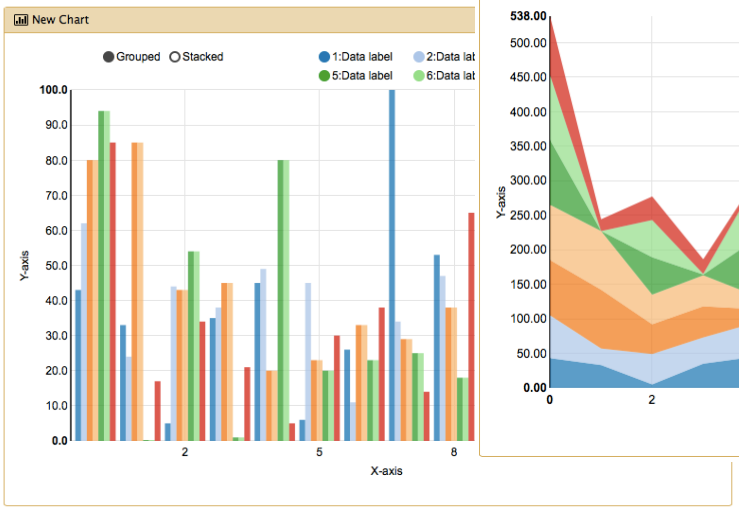
Graphics

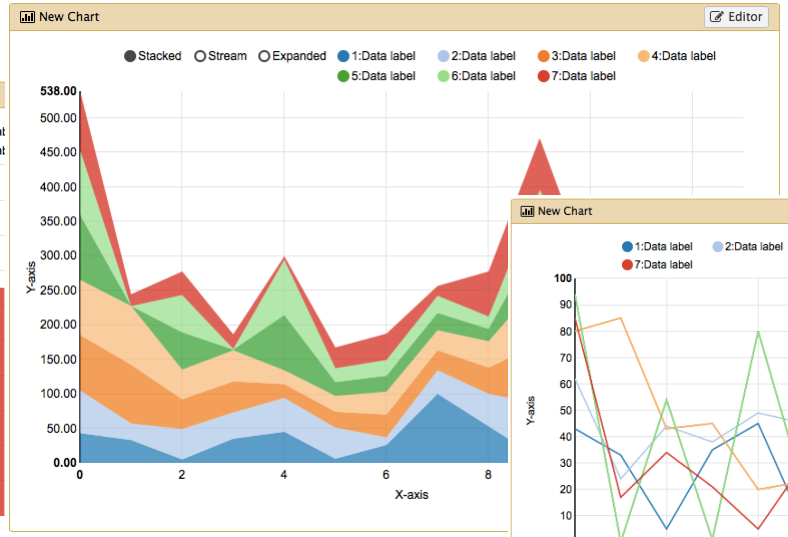


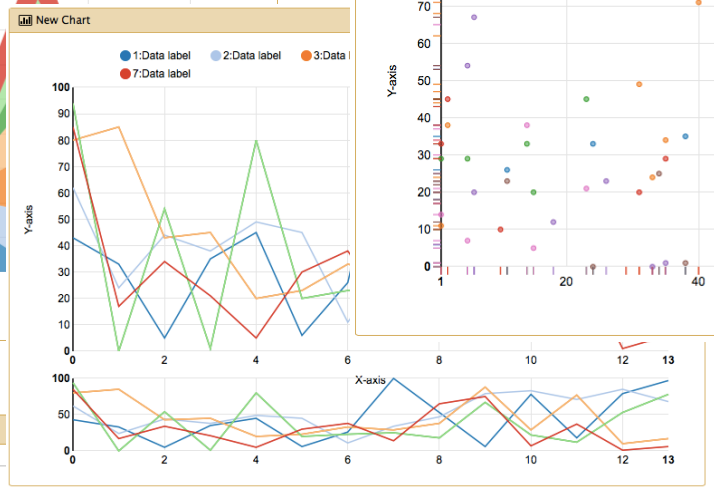


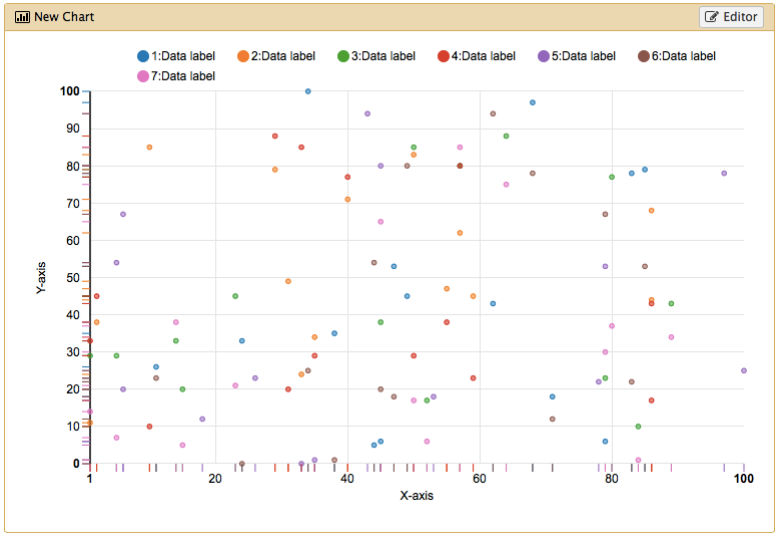
Dataset

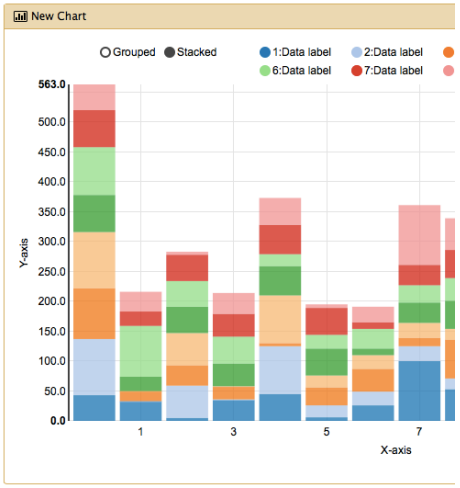
Graphics

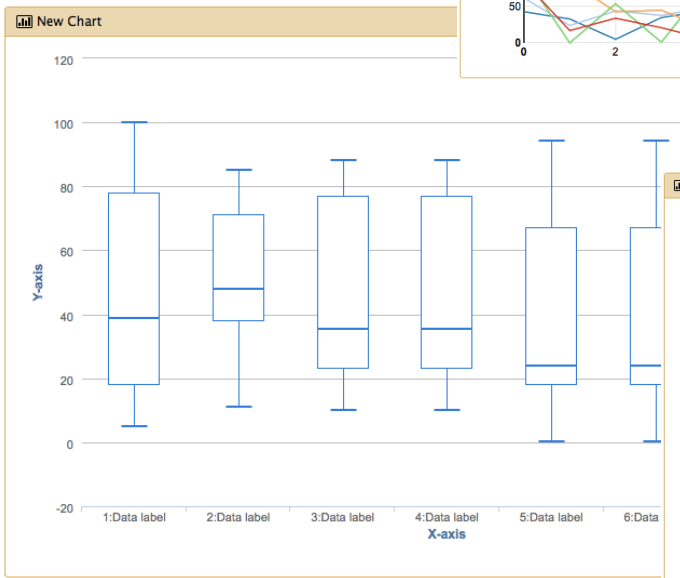


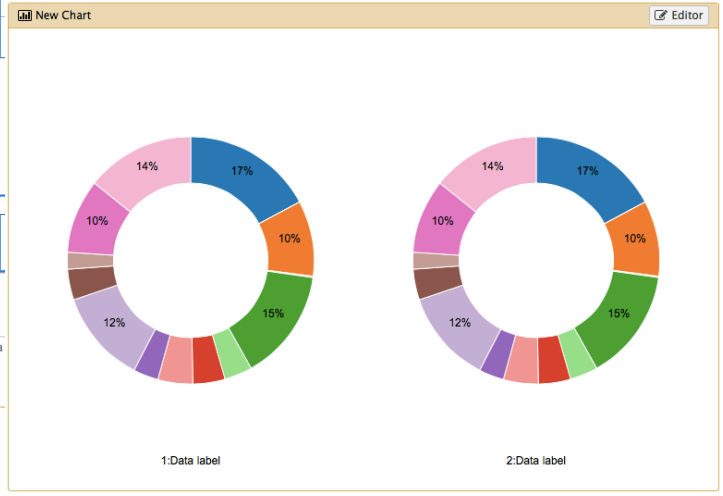






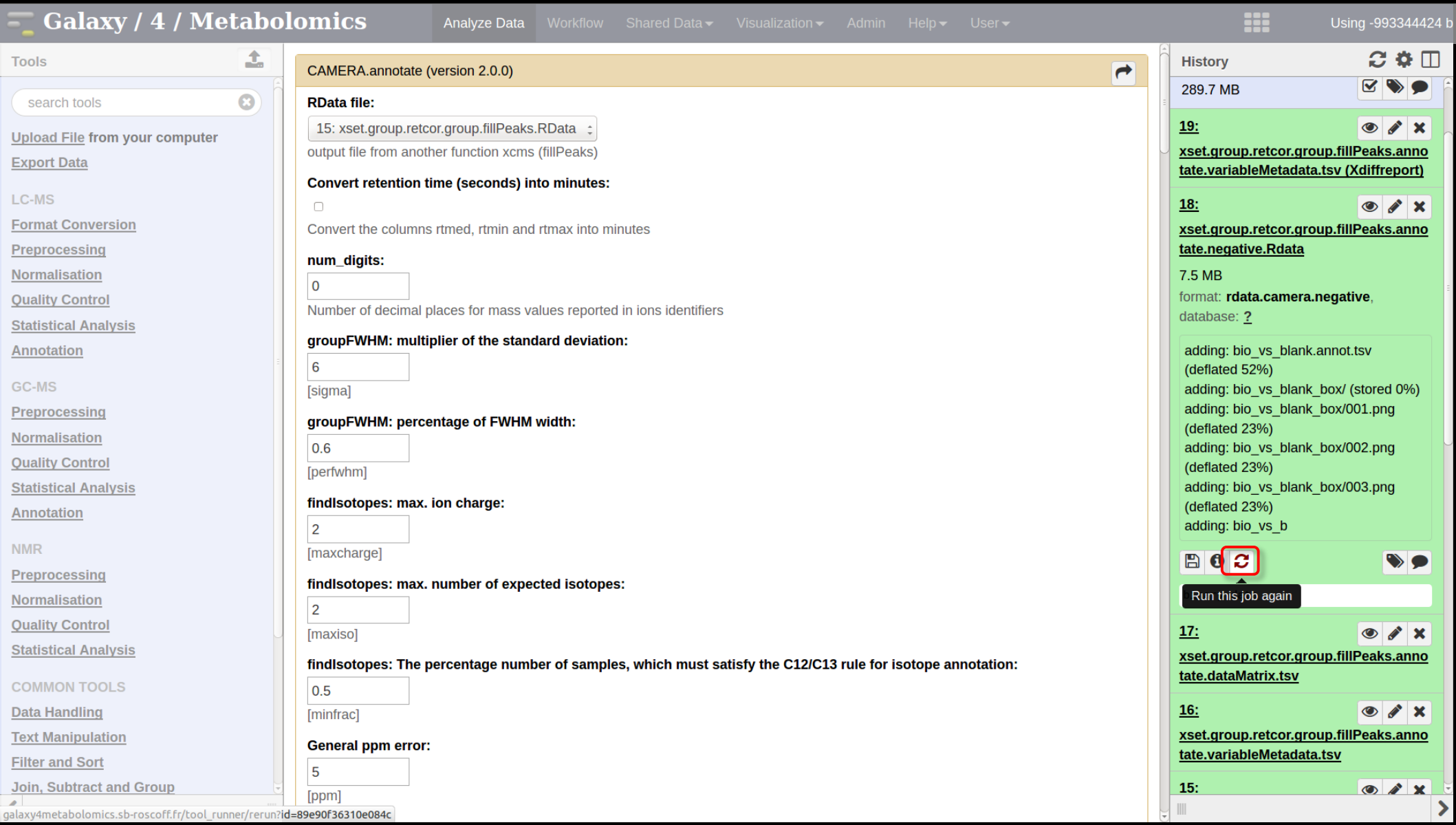






Dataset

Re-run a job

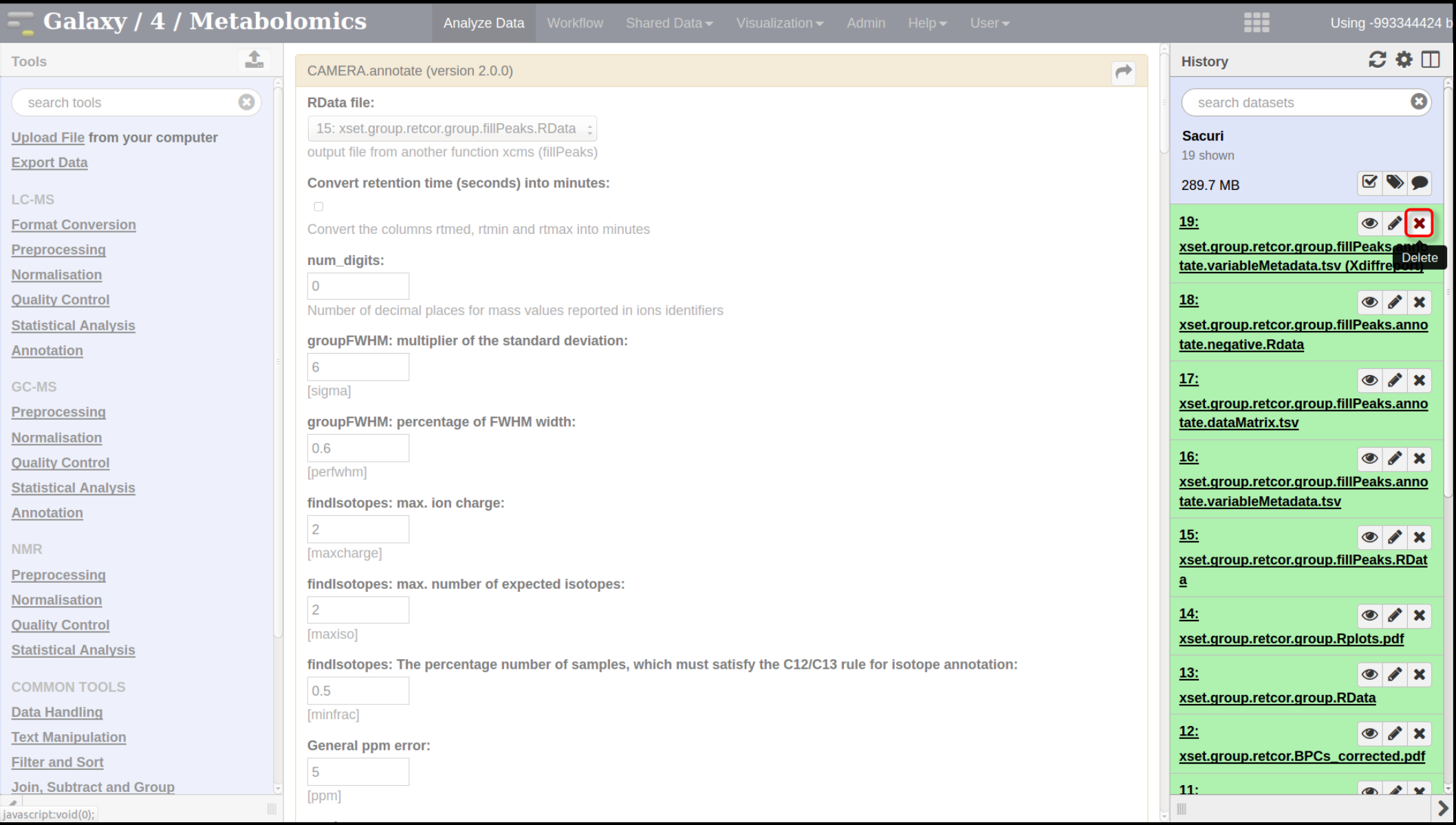


Cleanup

DATASET

Dataset / Cleanup

Delete a dataset



Dataset / Cleanup

The dataset isn't really deleted

It's in the Trash

Galaxy / 4 / Metabolomics

Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using -993344424 b

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Data Handling

Text Manipulation

Filter and Sort

Join, Subtract and Group

javascript:void(0);

CAMERA.annotate (version 2.0.0)

RData file:

15: xset.group.retcor.group.fillPeaks.RData

output file from another function xcms (fillPeaks)

Convert retention time (seconds) into minutes:

Convert the columns rtmed, rtmin and rtmax into minutes

num_digits:

0

Number of decimal places for mass values reported in ions identifiers

groupFWHM: multiplier of the standard deviation:

6

[sigma]

groupFWHM: percentage of FWHM width:

0.6

[perfwhm]

findIsotopes: max. ion charge:

2

[maxcharge]

findIsotopes: max. number of expected isotopes:

2

[maxiso]

findIsotopes: The percentage number of samples, which must satisfy the C12/C13 rule for isotope annotation:

0.5

[minfrac]

General ppm error:

5

[ppm]

History

search datasets

Sacuri

19 shown hide deleted

289.7 MB

This dataset has been deleted

Undelete it

Permanently remove it from disk

19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)

18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata

17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

15: xset.group.retcor.group.fillPeaks.RData

14: xset.group.retcor.group.Rplots.pdf

13: xset.group.retcor.group.RData

Dataset / Cleanup

"Empty Trash" : to free up disk space

Galaxy / 4 / Metabolomics

Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using -993344424 b

Tools

search tools

Upload File from your computer

Export Data

LC-MS

Format Conversion

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

GC-MS

Preprocessing

Normalisation

Quality Control

Statistical Analysis

Annotation

NMR

Preprocessing

Normalisation

Quality Control

Statistical Analysis

COMMON TOOLS

Data Handling

Text Manipulation

Filter and Sort

Join, Subtract and Group

CAMERA.annotate (version 2.0.0)

RData file:

15: xset.group.retcor.group.fillPeaks.RData

output file from another function xcms (fillPeaks)

Convert retention time (seconds) into minutes:

Convert the columns rtmed, rtmin and rtmax into minutes

num_digits:

0

Number of decimal places for mass values reported in ions identifiers

groupFWHM: multiplier of the standard deviation:

6

[sigma]

groupFWHM: percentage of FWHM width:

0.6

[perfwhm]

findIsotopes: max. ion charge:

2

[maxcharge]

findIsotopes: max. number of expected isotopes:

2

[maxiso]

findIsotopes: The percentage number of samples, which must satisfy the C12/C13 rule for isotope annotation:

0.5

[minfrac]

General ppm error:

5

[ppm]

History

HISTORY LISTS

Saved Histories

Histories Shared with Me

CURRENT HISTORY

Create New

Copy History

Copy Datasets

Share or Publish

Extract Workflow

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Purge Deleted Datasets

Show Structure

Export Citations

Export to File

Delete

Delete Permanently

OTHER ACTIONS

Import from File

14: xset.group.retcor.group.Rplots.pdf

13: xset.group.retcor.group.RData

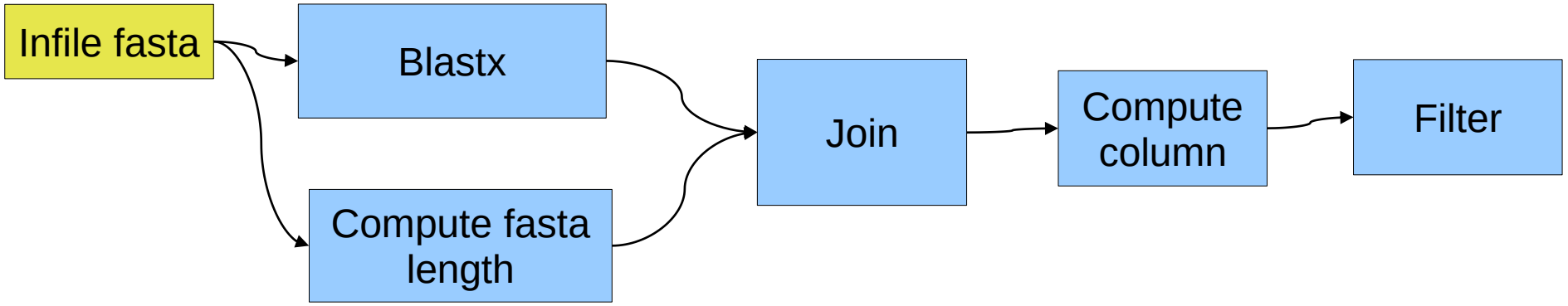
WORKFLOW

Workflow

- A workflow is a sequence of tool operations and parameters
- Can match the experiment protocol
- A workflow is built to be replayed (more or less strict)

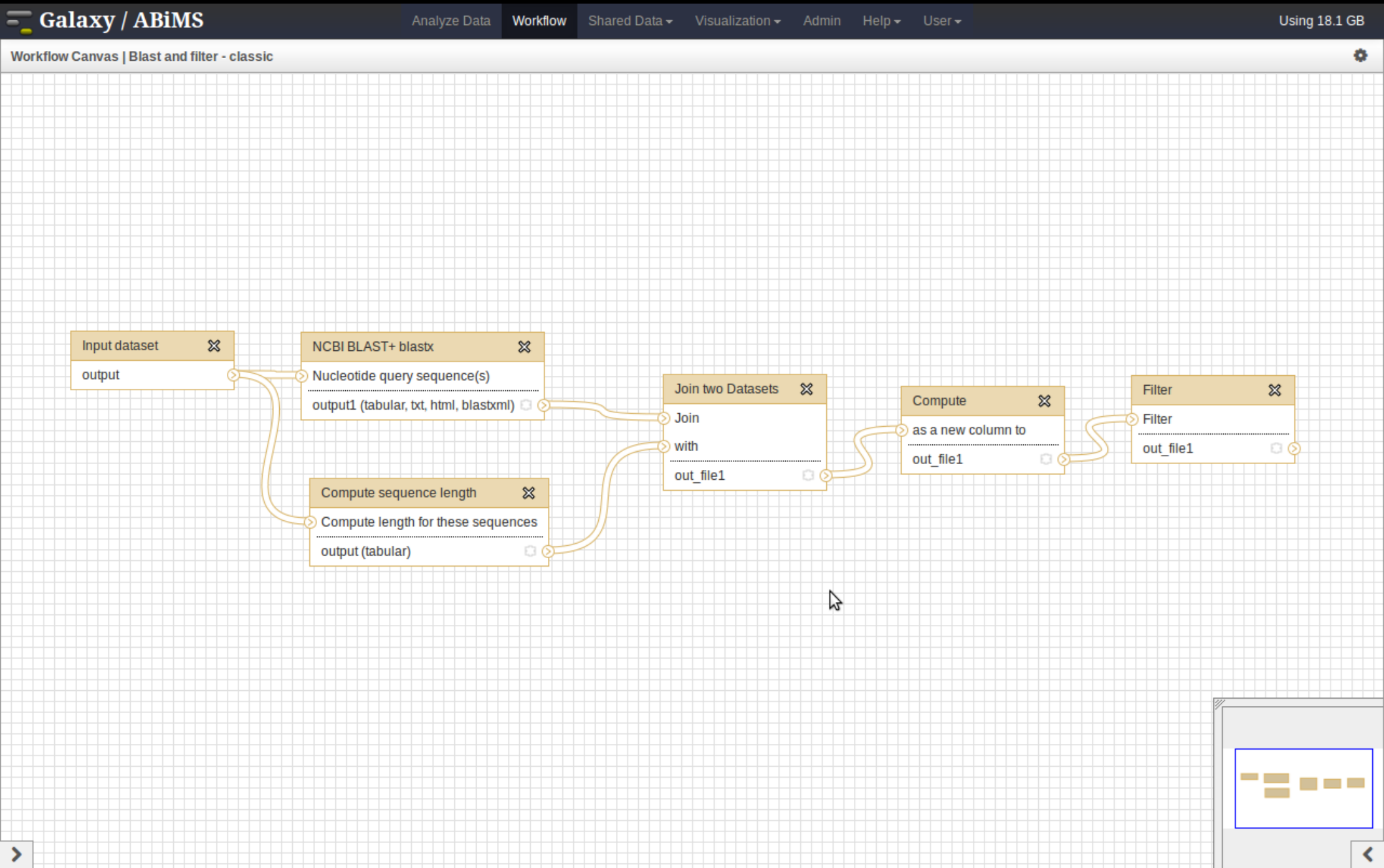
Workflow

- Our workflow



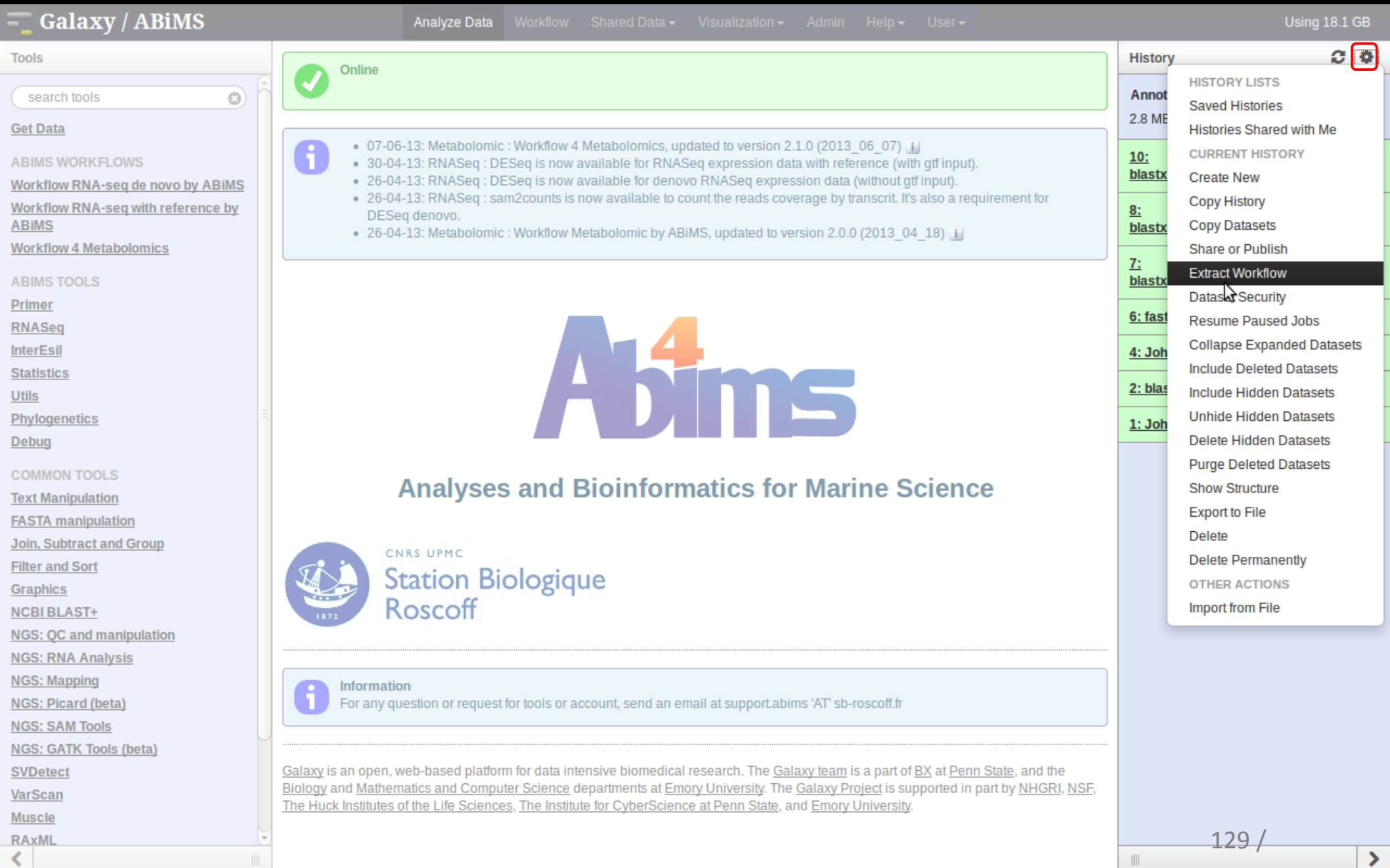
Workflow

Our workflow with Galaxy



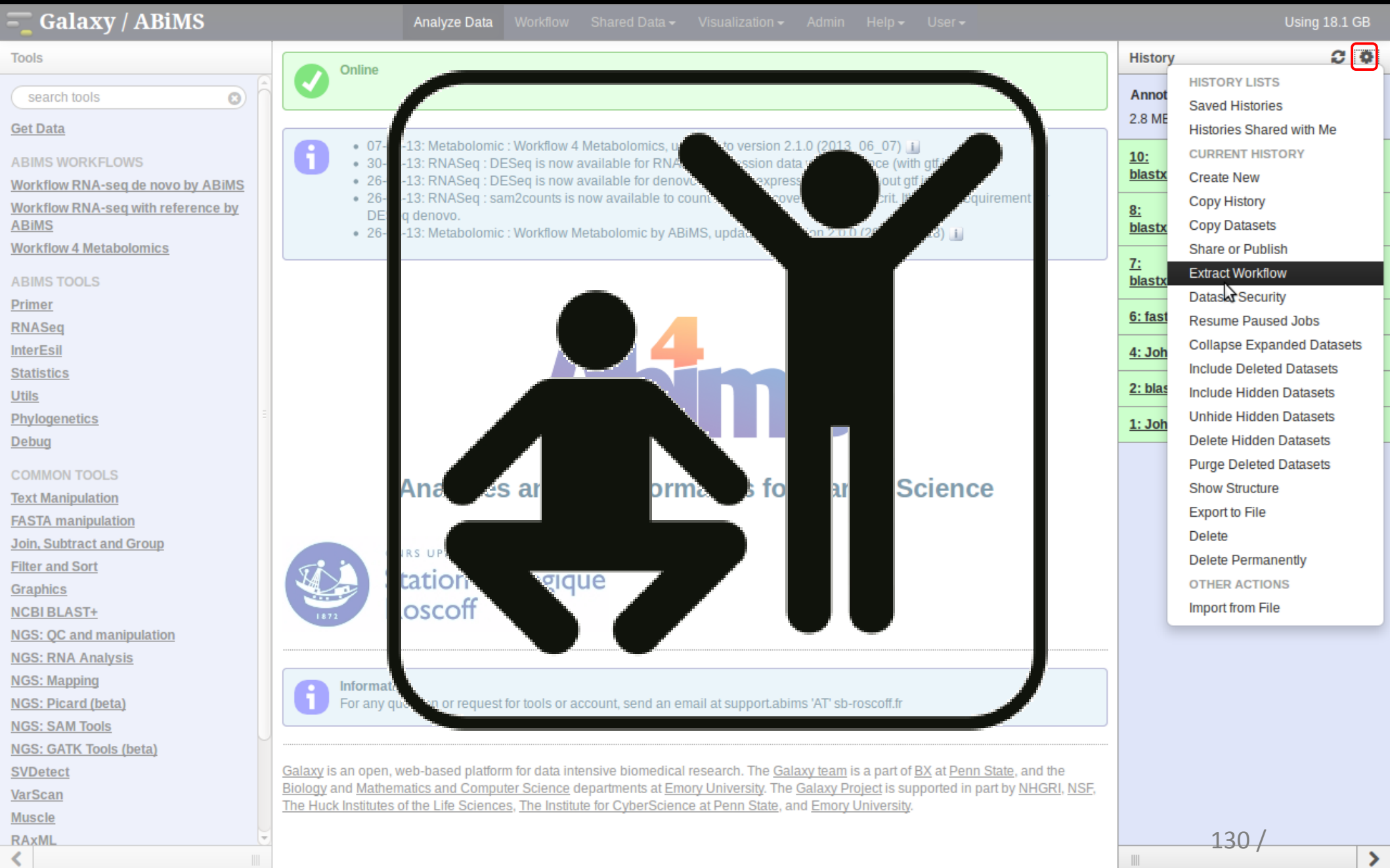
Workflow

From a history



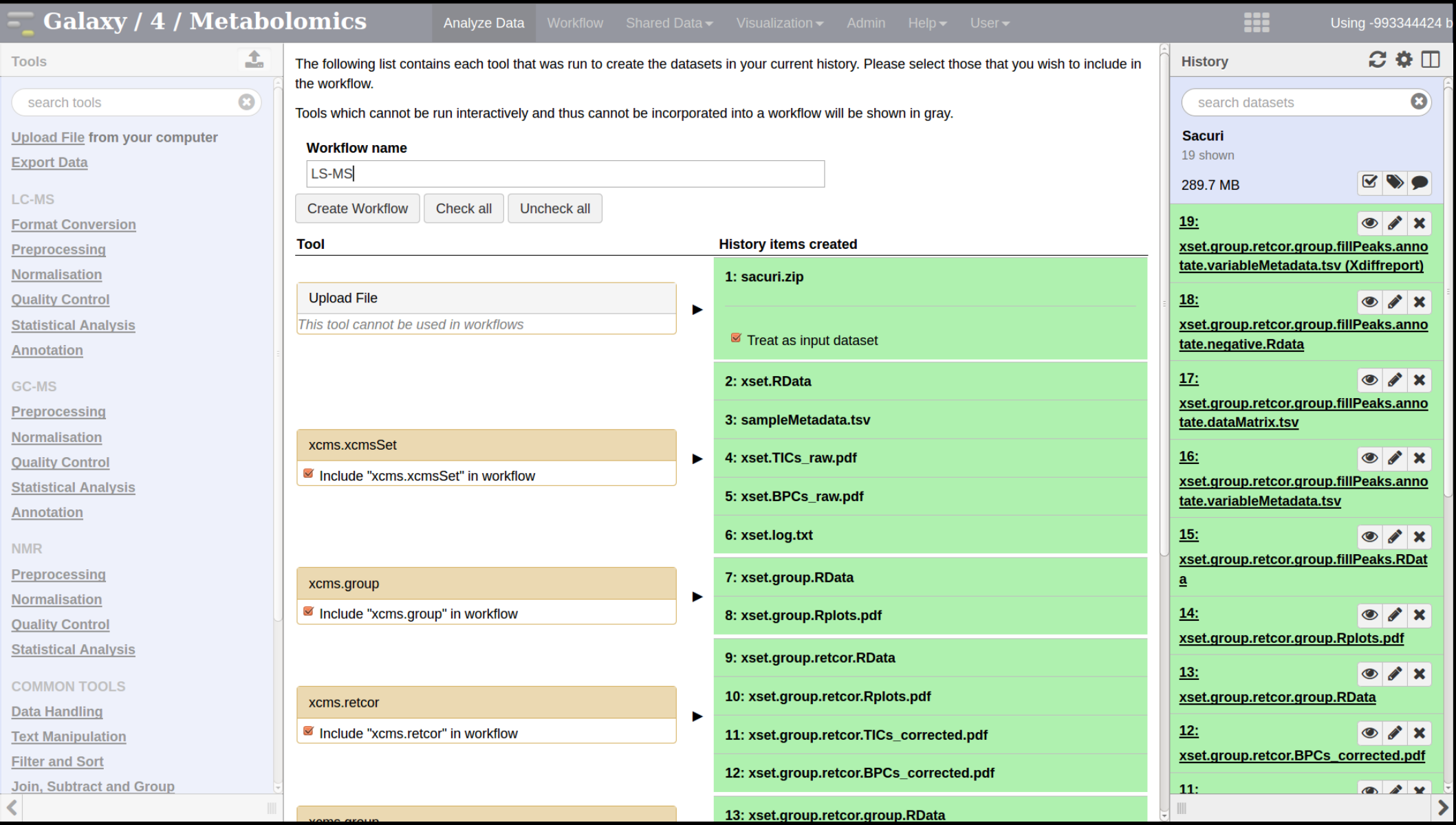
Workflow

From a history



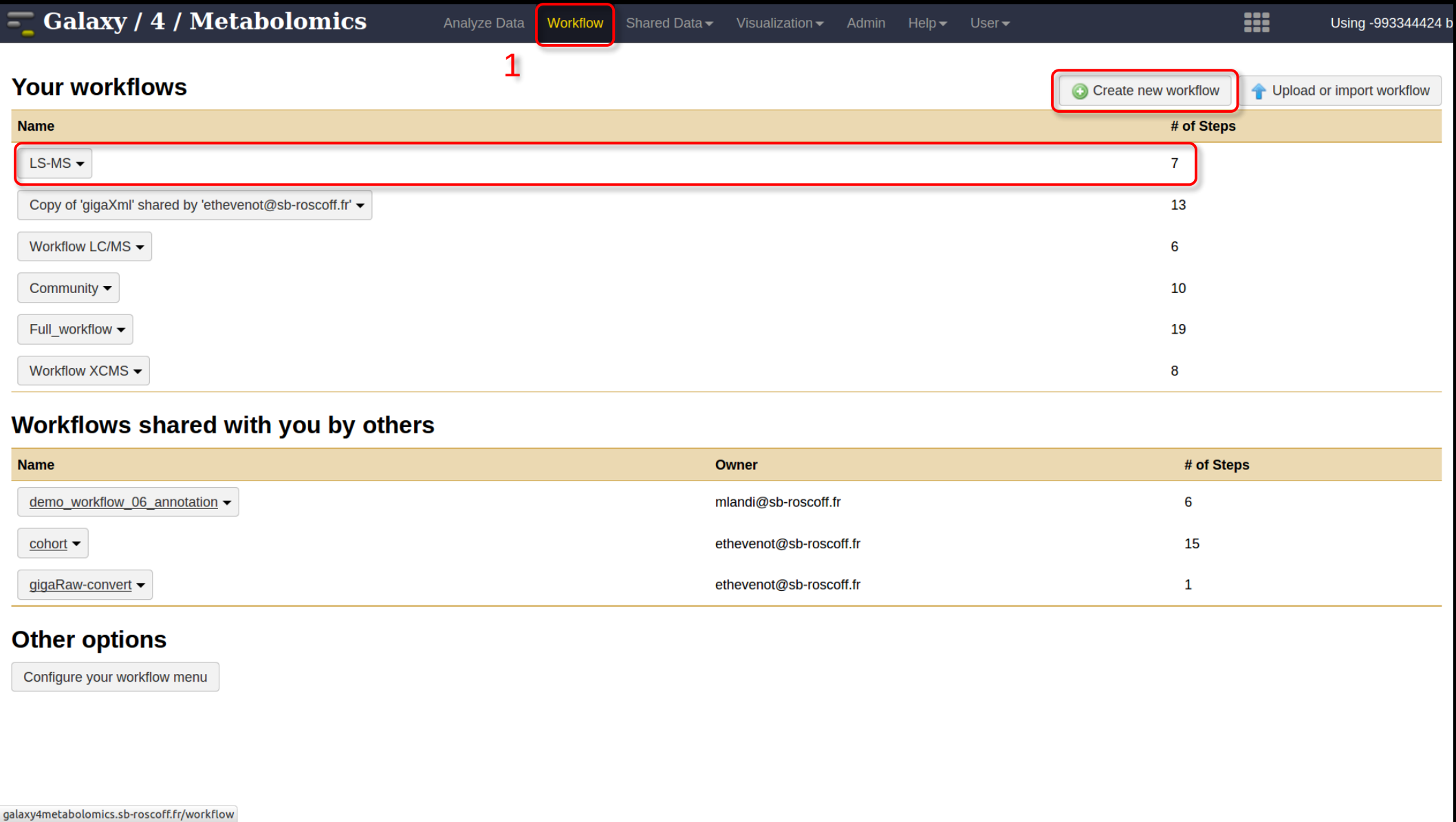
Workflow

From a history



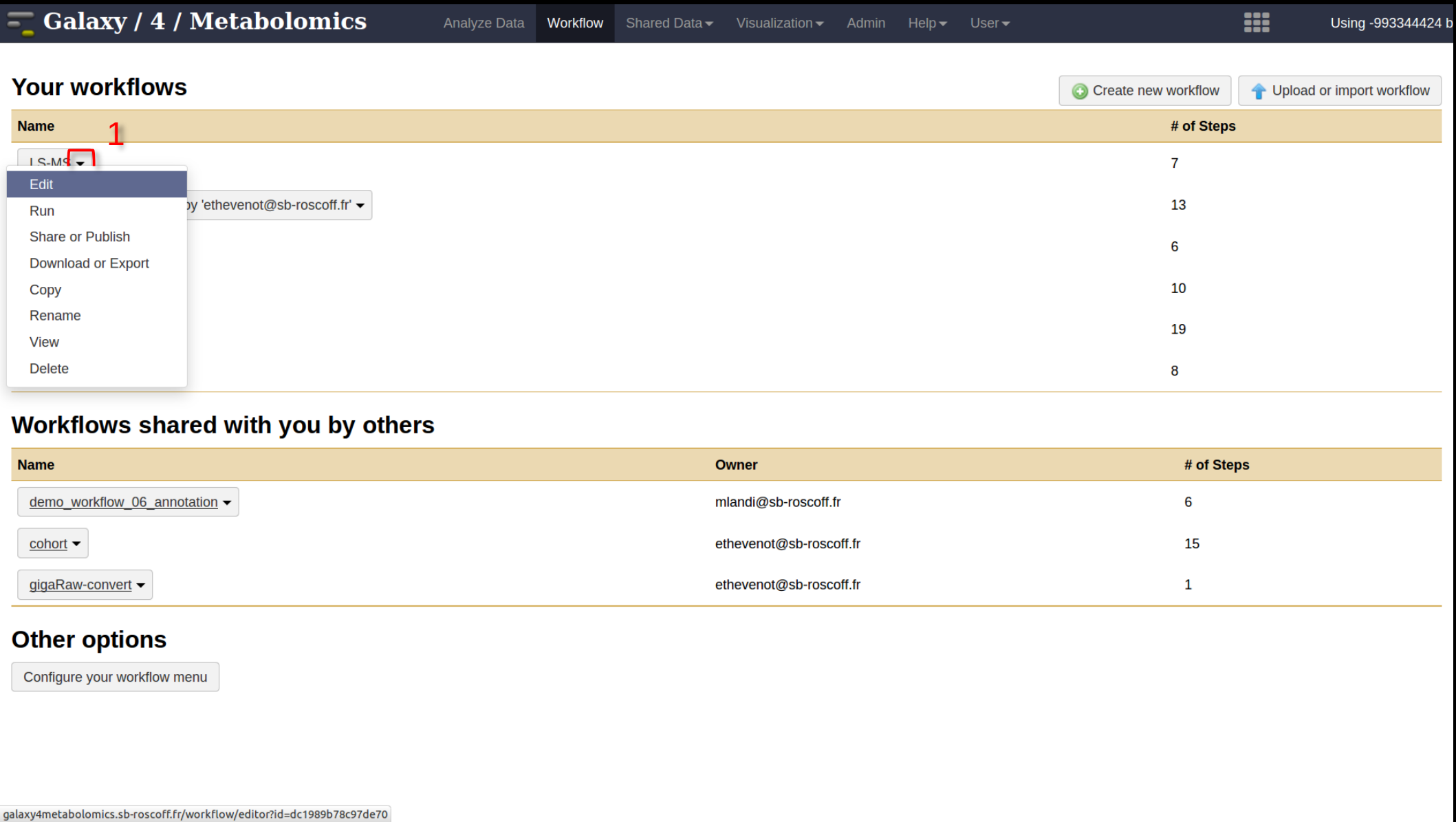
Workflow

Workflow manager



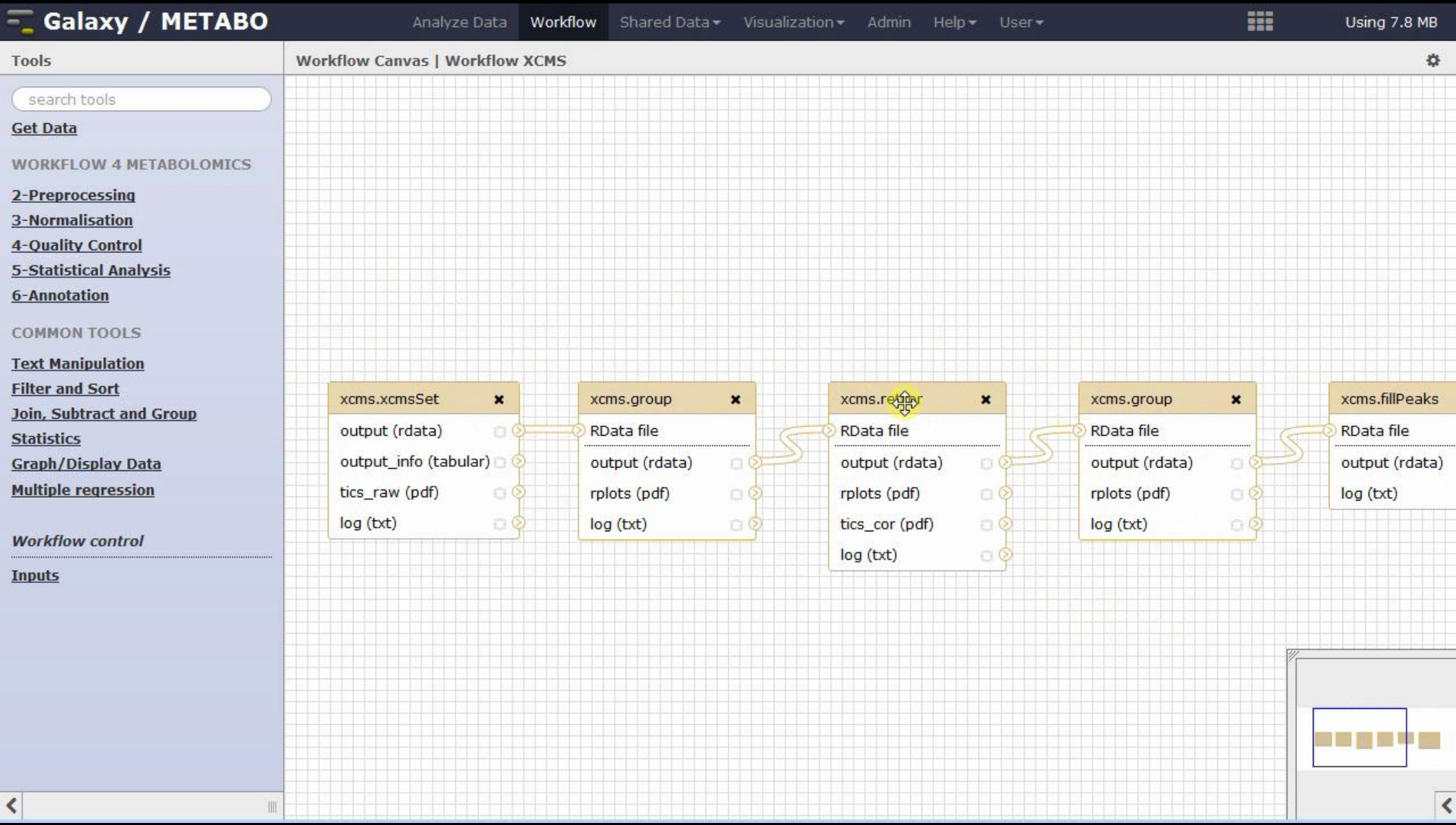
Workflow

Edit a workflow



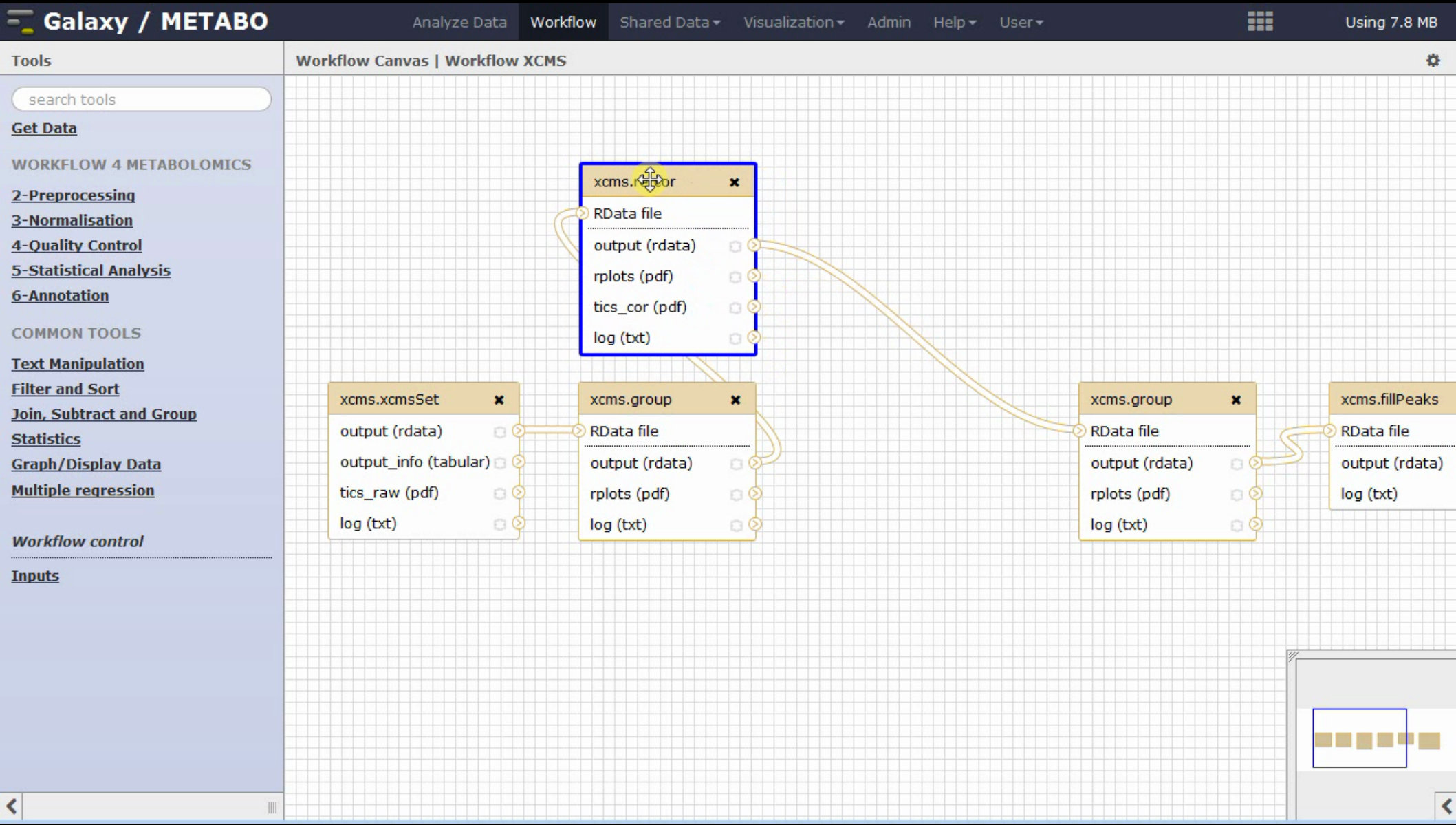
Workflow

Edit a workflow : drag and drop



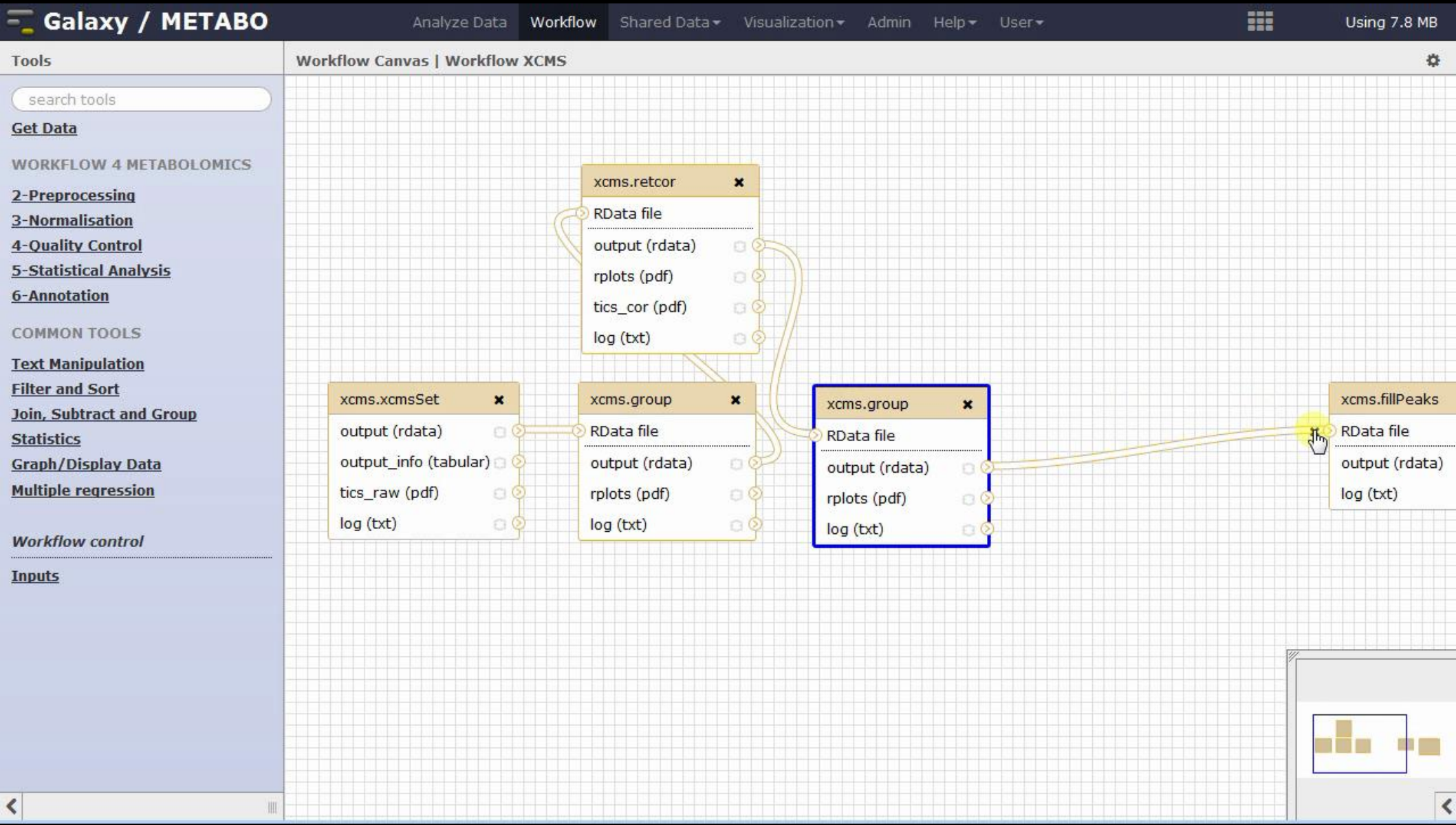
Workflow

Edit a workflow : drag and drop



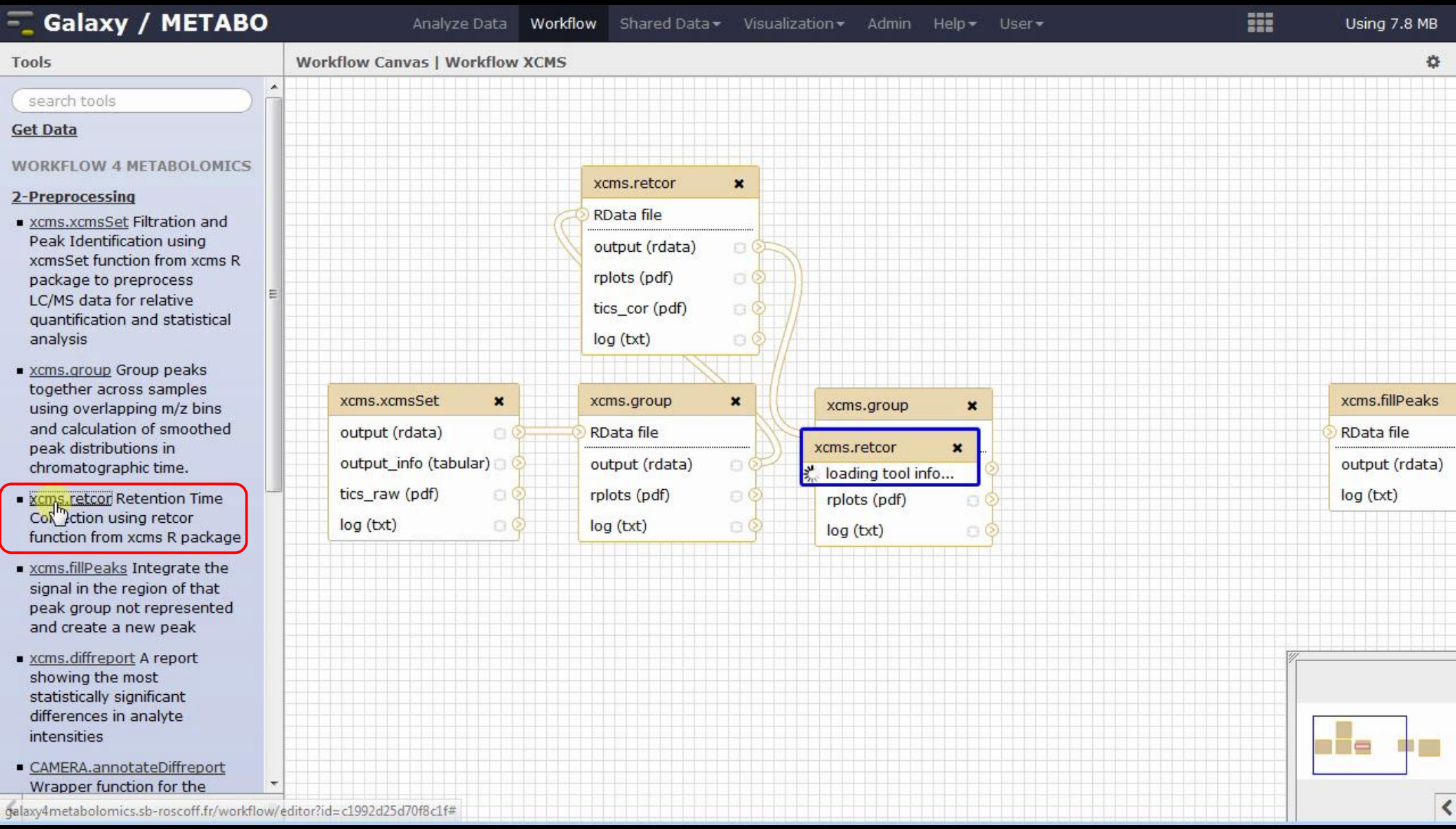
Workflow

Edit a workflow : delete a noodle



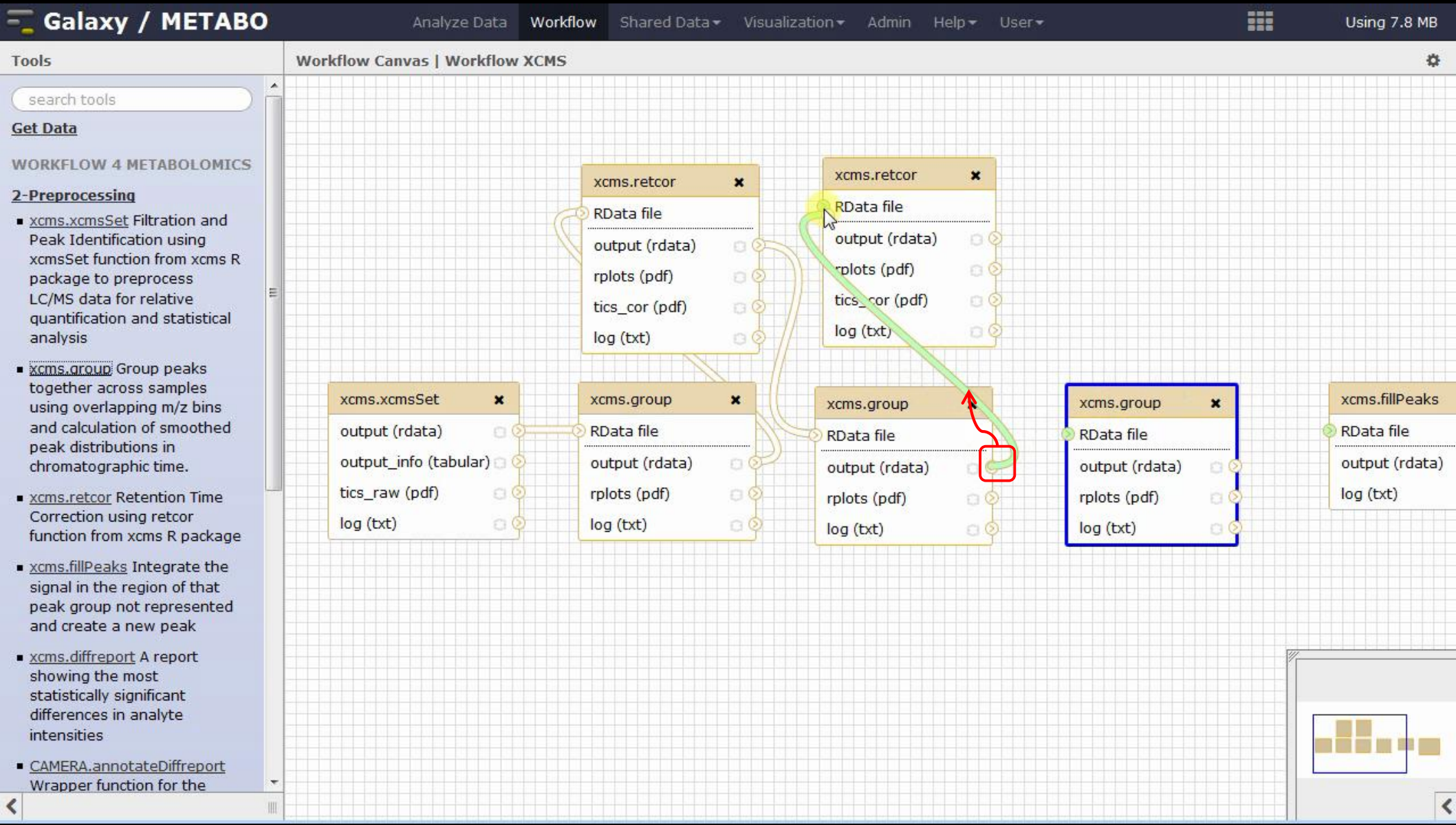
Workflow

Edit a workflow : add a tool



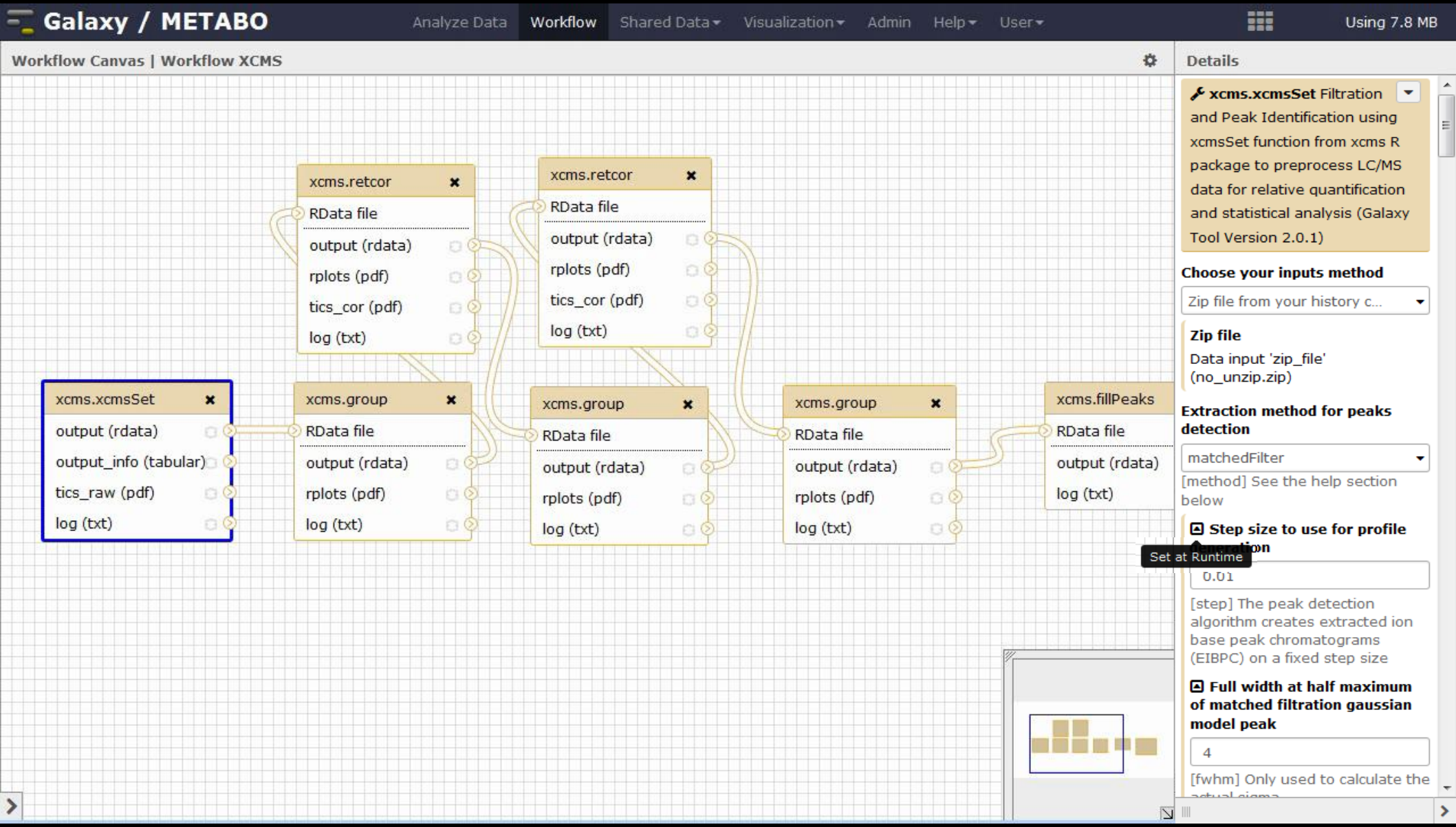
Workflow

Edit a workflow : add a noodle



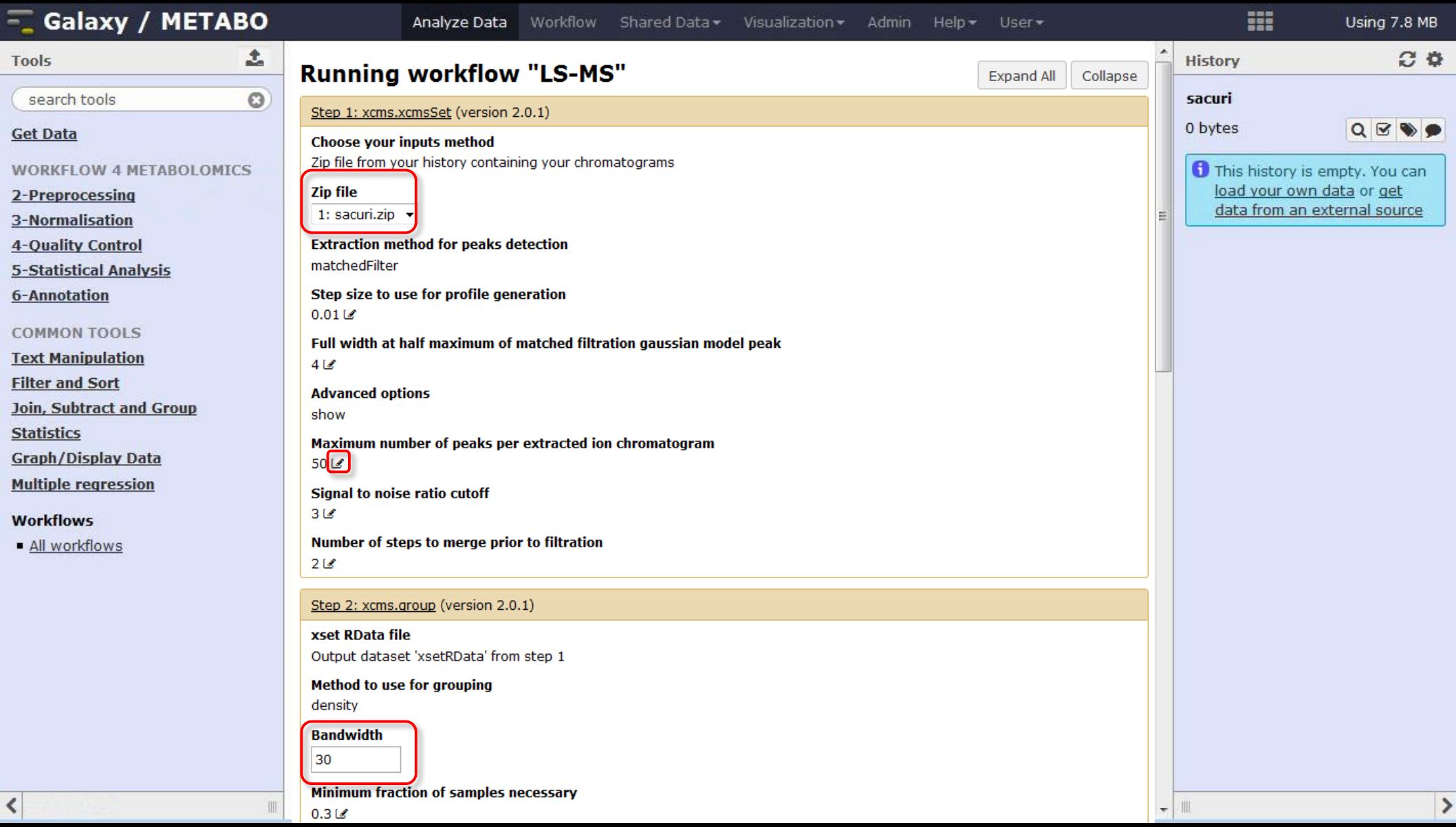
Workflow

Edit a workflow : set or release a parameter



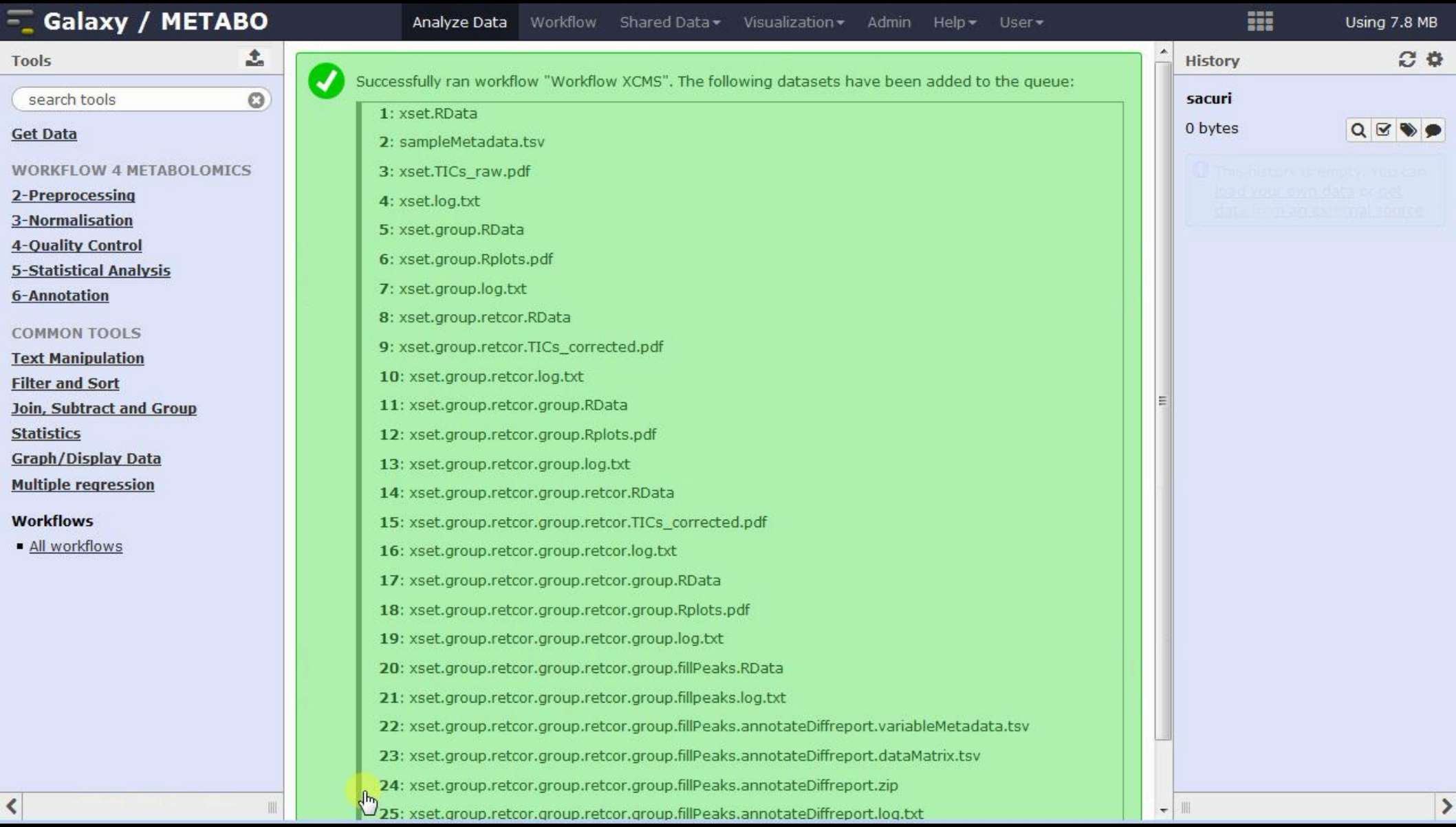
Workflow

Run a workflow



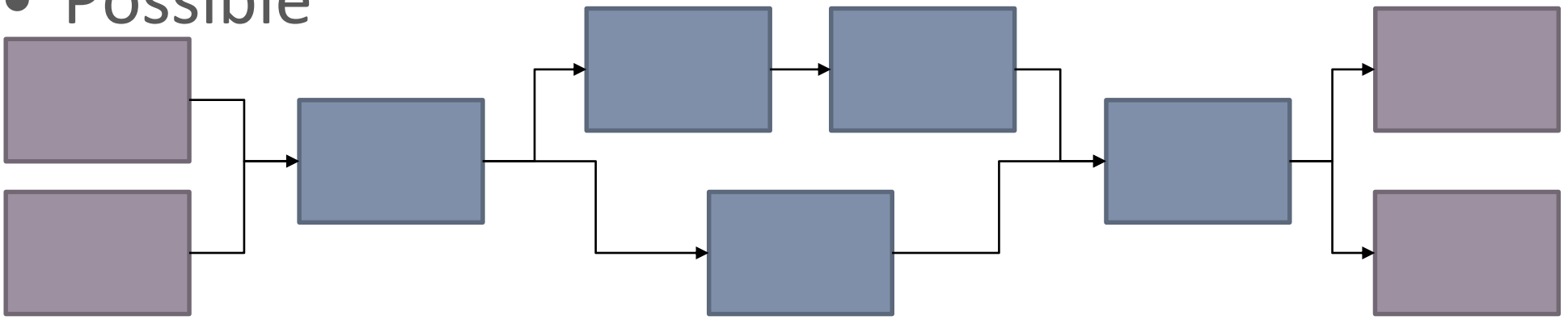
Workflow

Run a workflow : HOP!

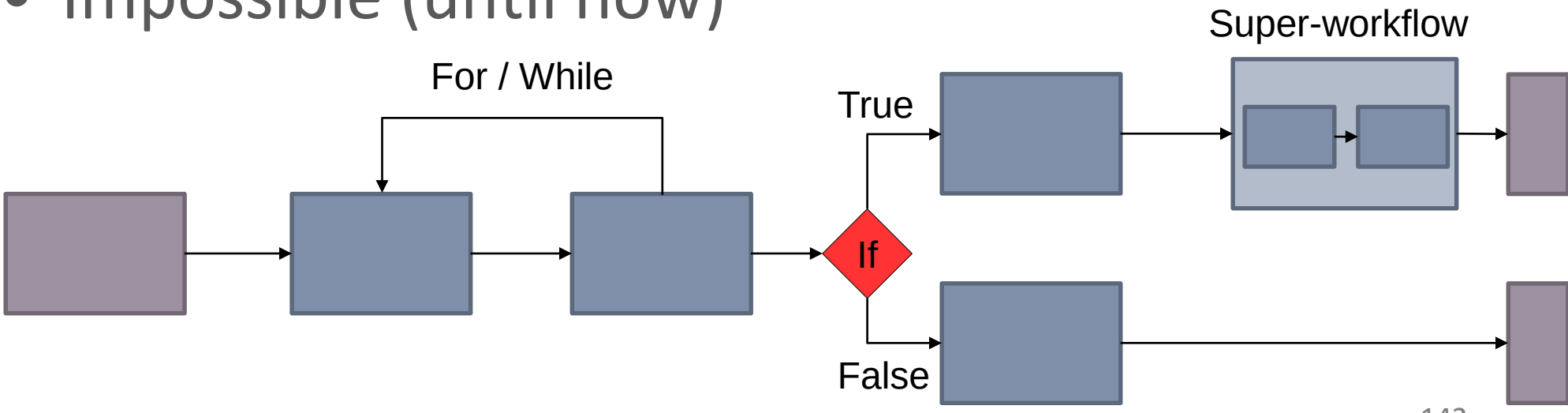


Workflow

- Possible



- Impossible (until now)



SHARE

Share

biologist ↔ biologist

- Sharing histories or datasets
  - With or without linked workflow

Share

bioanalyst ↔ biologist

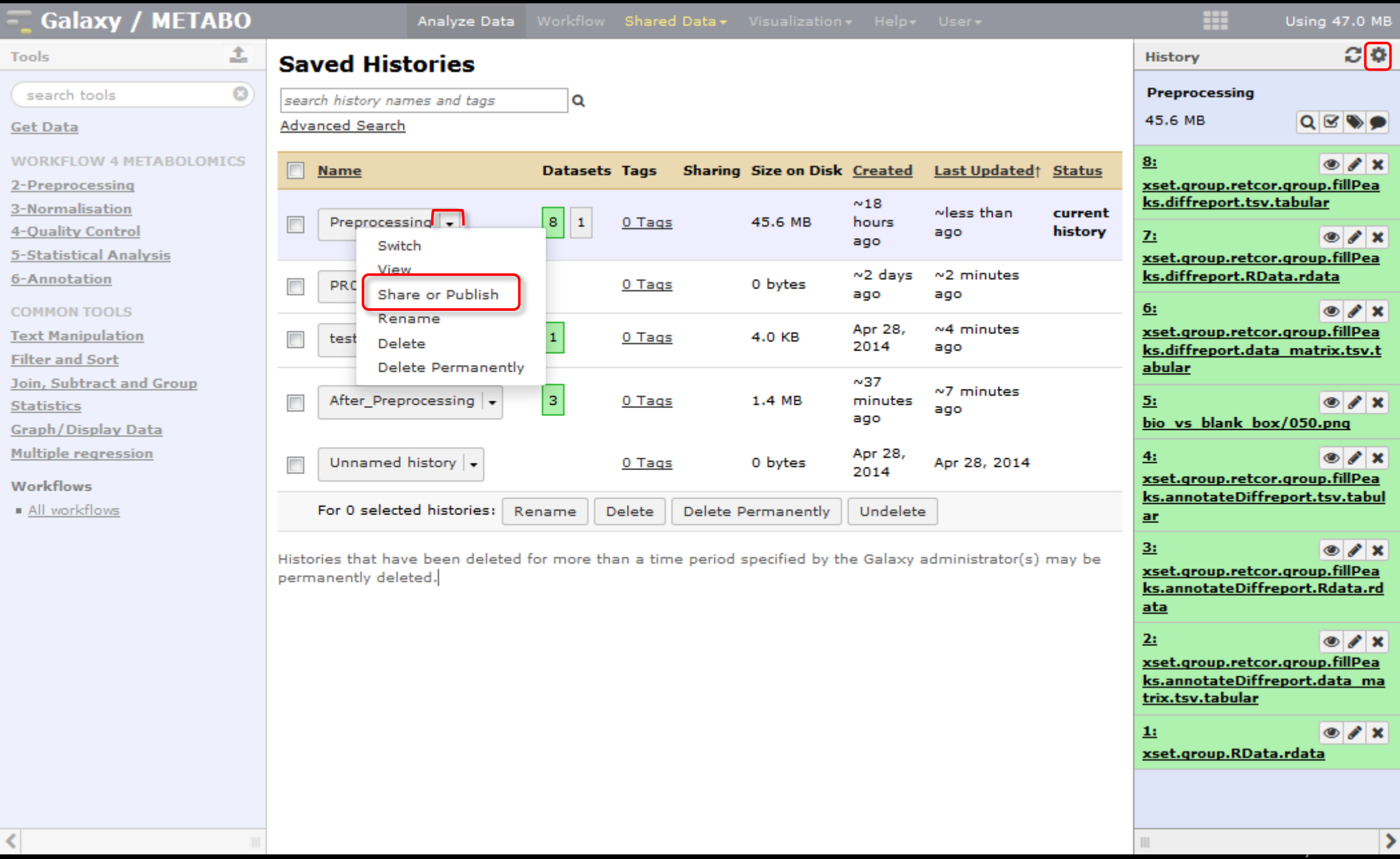
- Sharing workflows
  - Pre-configured parameters
  - With or without release parameters (set at runtime)
    - According to the user-end knowledge

bioinformatician ↔ bioinformatician

- Sharing tools ,scripts and wrappers
  - Toolshed

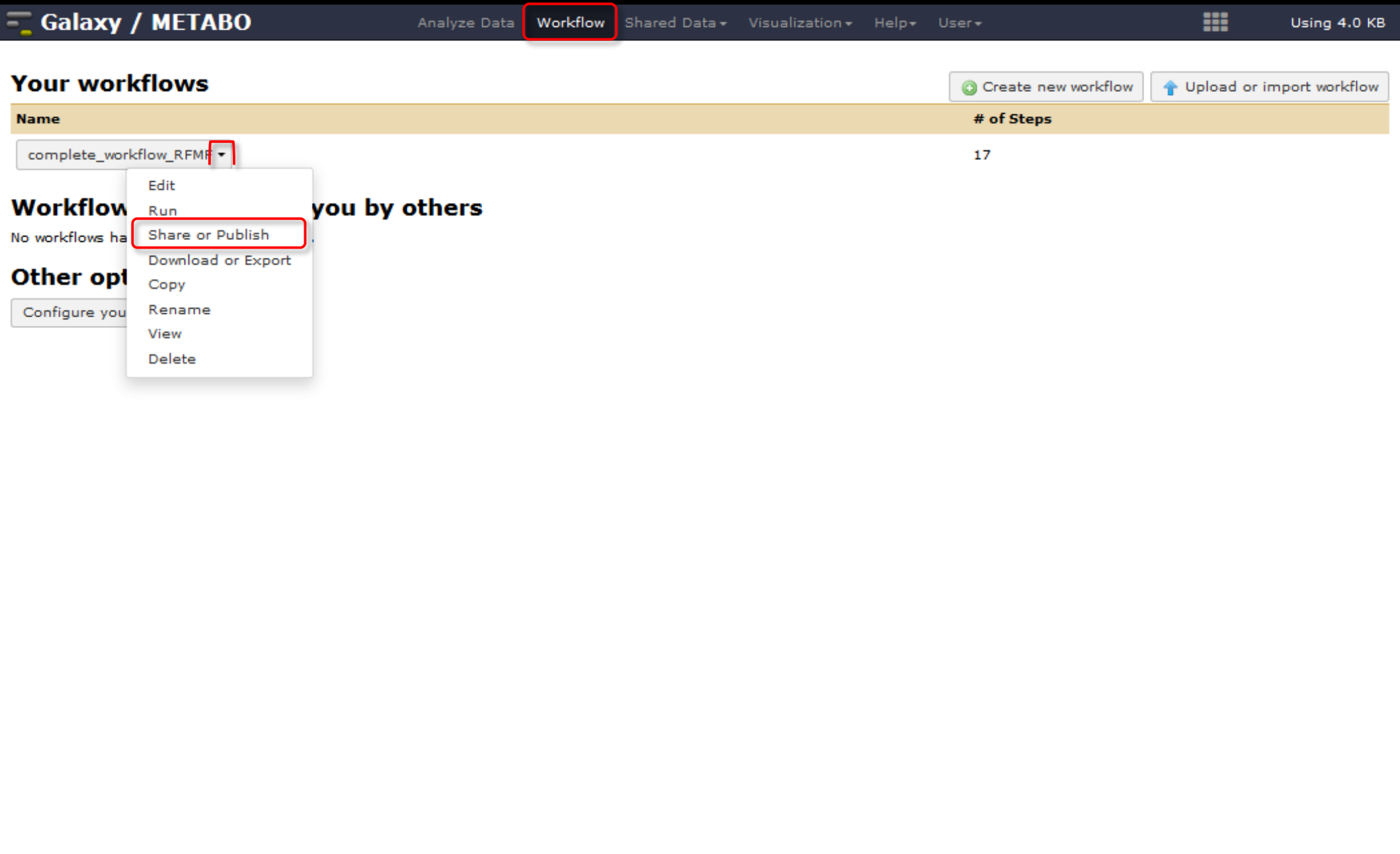
Share

Datasets



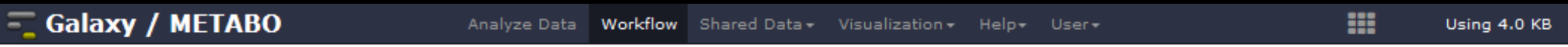
Share

Workflow



Share

Mode



Share or Publish Workflow 'complete_workflow_RFMF'

Make Workflow Accessible via Link and Publish It

This workflow is currently restricted so that only you and the users listed below can access it. You can:

Make Workflow Accessible via Link → Restricted community

Generates a web link that you can share with other people so that they can view and import the workflow.

Make Workflow Accessible and Publish → All the Galaxy server users

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's Published Workflows section, where it is publicly listed and searchable.

Share Workflow with Individual Users

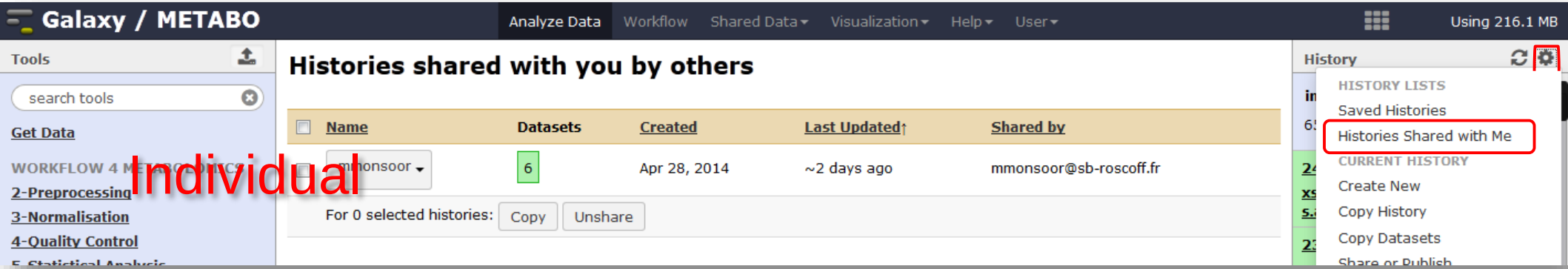
You have not shared this workflow with any users.

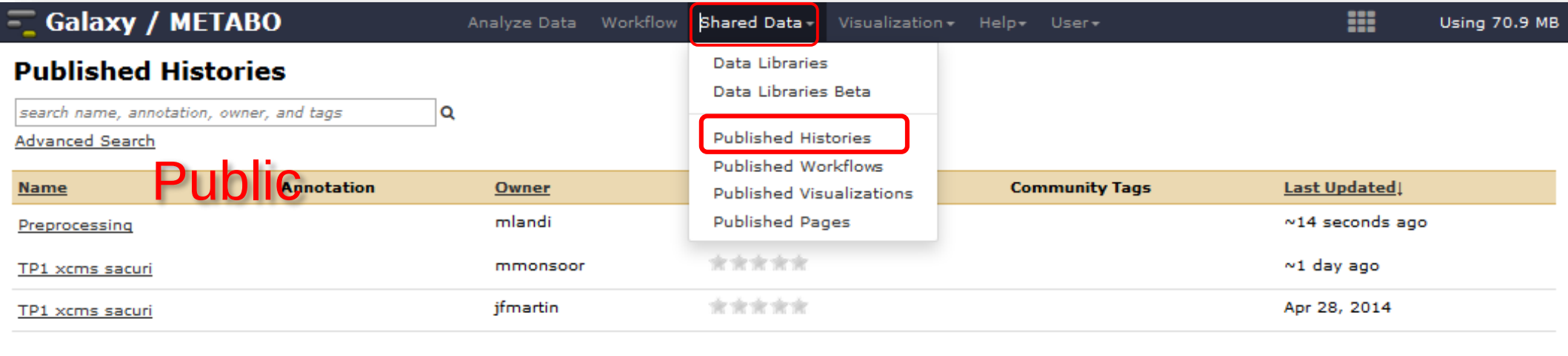
Share with a user → Designated community (login@sb-roscoff.fr)

Back to Workflows List

Share

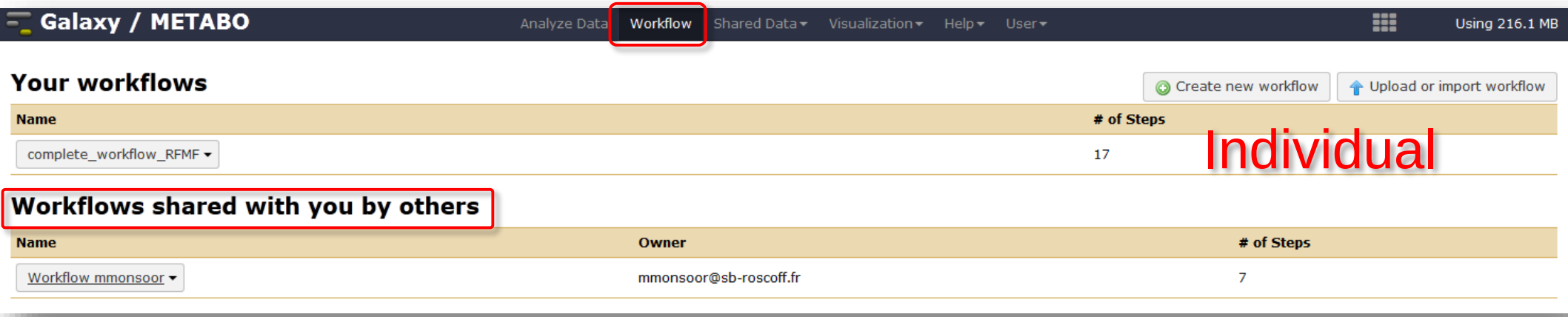
- Get shared histories

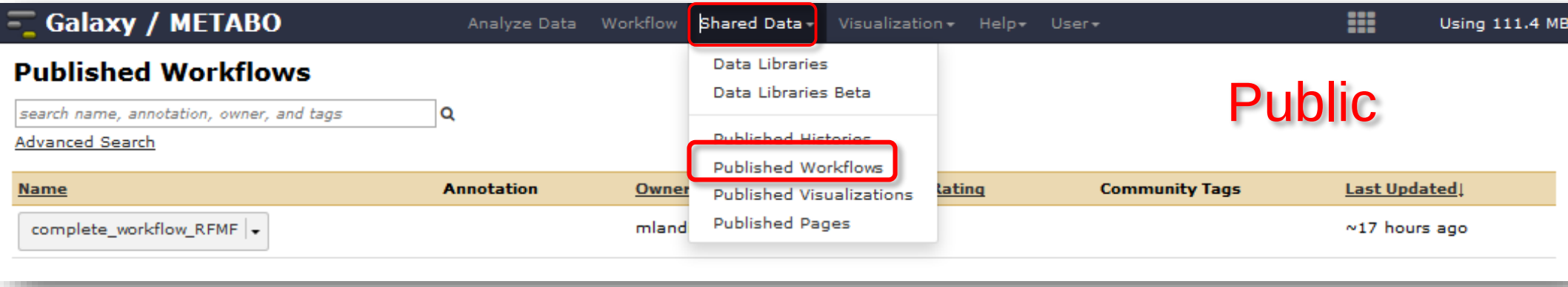




Share

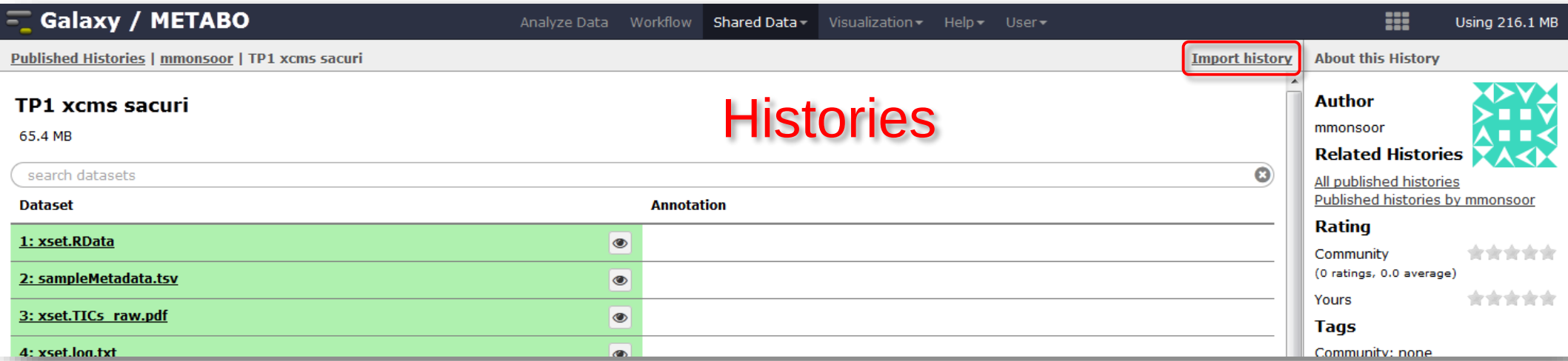
- Get shared workflows

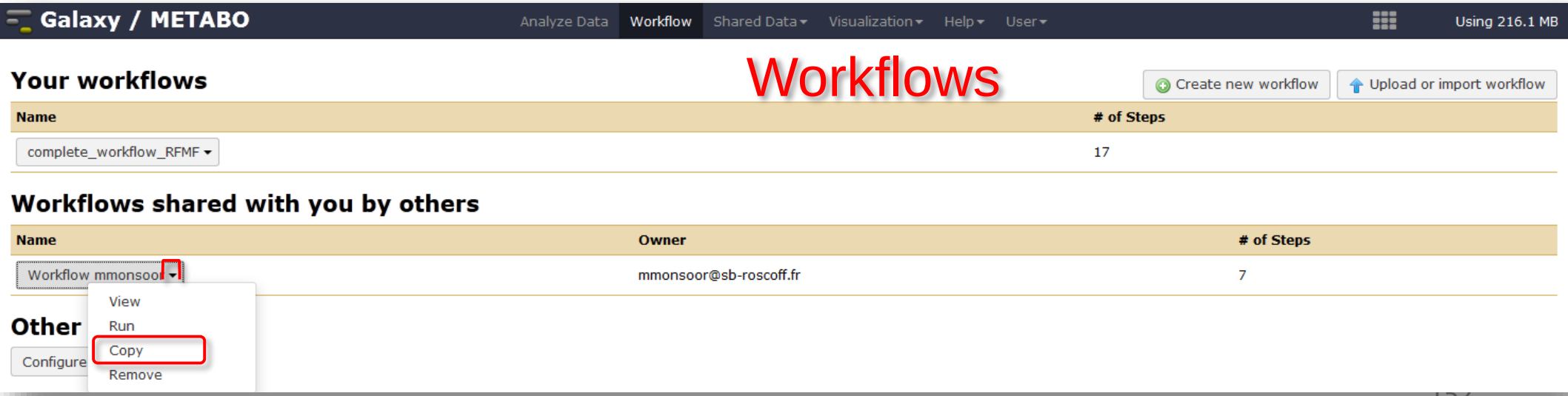




Share

- Import shared





Share

Level 5

- Share of tools and descriptions in the ToolShed

Level 4

- Launch autonomously tools
- Use advanced parameters
- Use the Galaxy API
- Provide workflow for collegues Level 1-3

Level 3

- Launch autonomously tools
- Use workflow more or less presetted

Level 2

- Use presetted workflow

Level 1

- Share his data to collegues Level 2-5

END

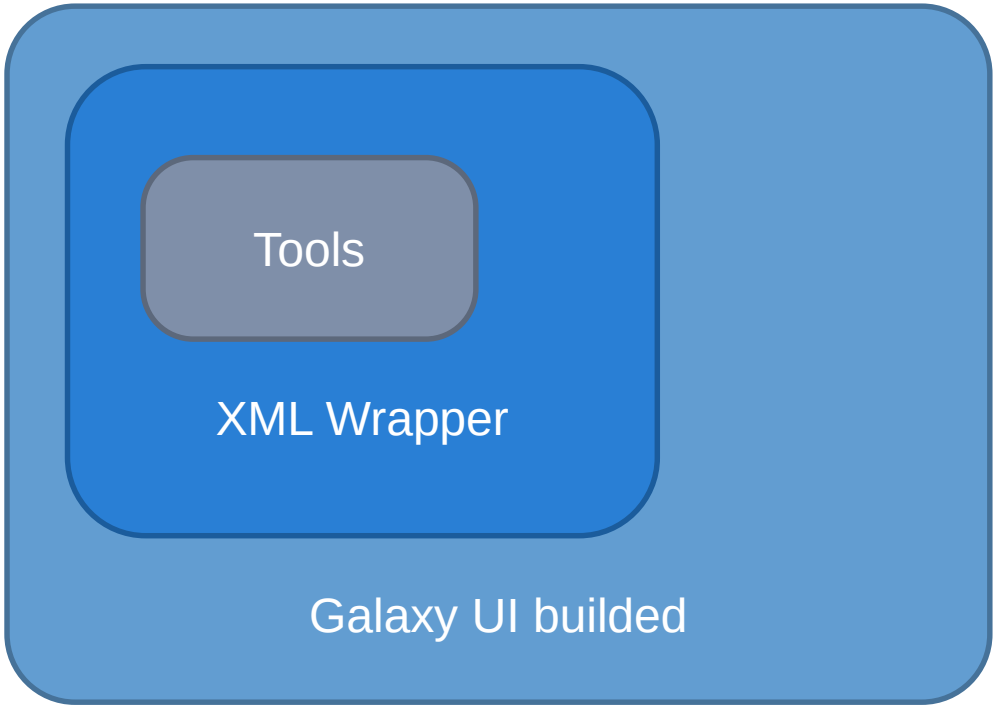
BONUS

How are tools born?

BONUS

Bonus

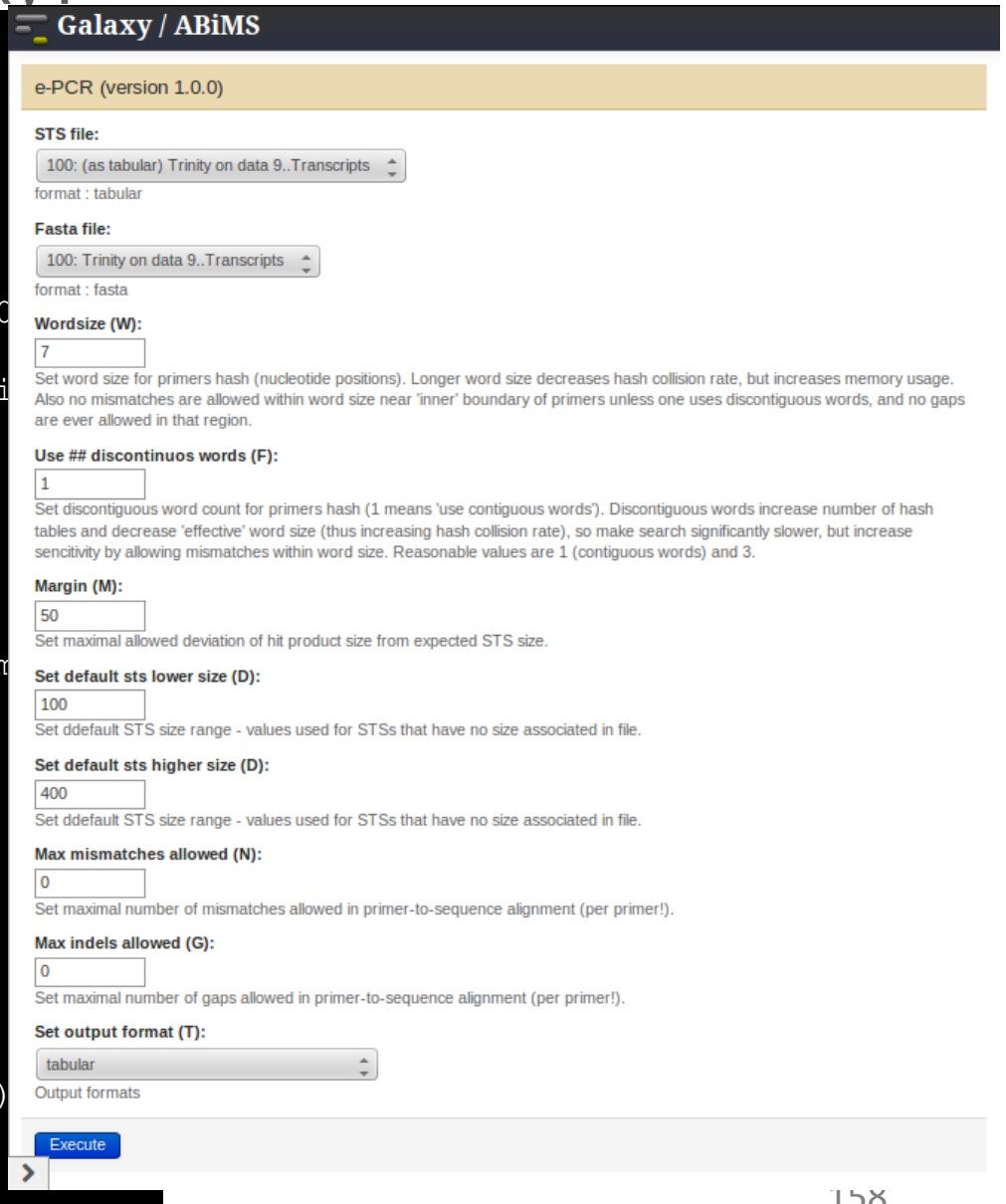
- How to import a tool in Galaxy?



Bonus

- How to import a tool in Galaxy?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
        -m ##    Margin (default 50)
        -w ##    Wordsize  (default 7)
        -n ##    Max mismatches allowed (default 0
        -g ##    Max indels allowed (default 0)
        -f ##    Use ## discontiguos words, slow i
                        ##>1
        -o ##    Set output file
        -t ##    Set output format:
                 1 - classic, range (pos1..pos2)
                 2 - classic, midpoint
                 3 - tabular
                 4 - tabular with alignment in com
                        (slow)
        -d##-##  Set default size range
                        (default 100-350)
        -p +-    Turn hits postprocess on/off
        -v ##    Verbosity flags
        -a a|f   Use presize alignmens (only if
                        gaps>0), slow
                  a - Allways or f - as Fallback
        -x +-    Use 5'-end lowercase masking of
                        primers (default -)
        -u +-    Uppercase all primers (default -)

[...]
```



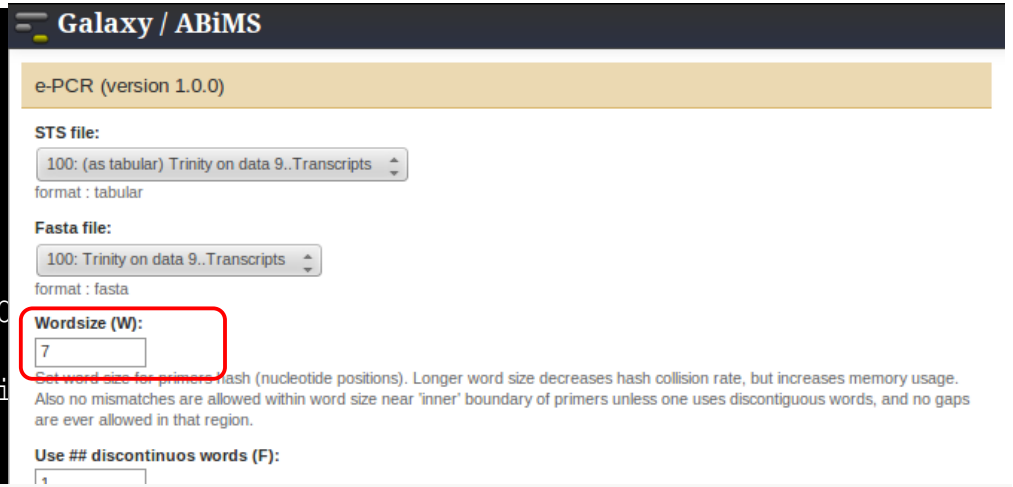
Bonus

- How to import a tool in Galaxy?

```
<tool id="abims_epcr" name="e-PCR">
    <!-- author : lecorguille@sb-roscoff.fr -->
    <!-- date : 11-05-12 -->
    <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>

    <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile $infile_fasta > $output</command>

    <inputs>
        <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
        <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
        <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontiguous words, and no gaps are ever allowed in that region." />
        <param name="wordcnt" type="integer" label="Use ## discontinuos words (F)" value="1" help="Set discontiguous word count for primers hash (1 means 'use contiguous words'). Discontiguous words increase number of hash tables and decrease 'effective' word size (thus increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word size. Reasonable values are 1 (contiguous words) and 3." />
        <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from expected STS size." />
        <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values used for STSs that have no size associated in file." />
        <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range - values used for STSs that have no size associated in file." />
        <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed in primer-to-sequence alignment (per primer!)." />
        <param name="max_gap" type="integer" label="Max indels allowed (G)" value="0" help="Set maximal number of gaps allowed in primer-to-sequence alignment (per primer!)." />
        <param name="output_format" type="select" help="Output formats">
            <label>Set output format (T)</label>
            <option value="1">classic, range (pos1..pos2)</option>
            <option value="2">classic, midpoint</option>
            <option value="3" selected="true">tabular</option>
            <option value="4">tabular with alignment in comments (slow)</option>
        </param>
    </inputs>

    <outputs>
        <data name="output" format="tabular" />
    </outputs>
```

- How to import a tool in Galaxy?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
        -m ##    Margin (default 50)
        -w ##    Wordsize  (default 7)
        -n ##    Max mismatches allowed (default 0
        -g ##    Max indels allowed (default 0)
        -f ##    Use ## discontiguos words, slow i
                          ##>1
        -o ##    Set output file
```




```
<tool id="abims_epcr" name="e-PCR">
    <!-- author : lecorguille@sb-roscoff.fr -->
    <!-- date : 11-05-12 -->
    <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if
    any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>
    <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile
    $infile_fasta > $output</command>
    <inputs>
        <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
        <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
        <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions).
        Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near
        'inner' boundary of primers unless one uses discontiguous words, and no gaps are ever allowed in that region." />
        <param name="wordcnt" type="integer" label="Use ## discontinuos words (F)" value="1" help="Set discontiguous word count for primers
        hash (1 means 'use contiguous words'). Discontiguous words increase number of hash tables and decrease 'effective' word size (thus
        increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word
        size. Reasonable values are 1 (contiguous words) and 3." />
        <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from
        expected STS size." />
        <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values
        used for STSs that have no size associated in file." />
        <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range -
        values used for STSs that have no size associated in file." />
        <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed
```