UiO : Universitetet i Oslo

Lifeportal

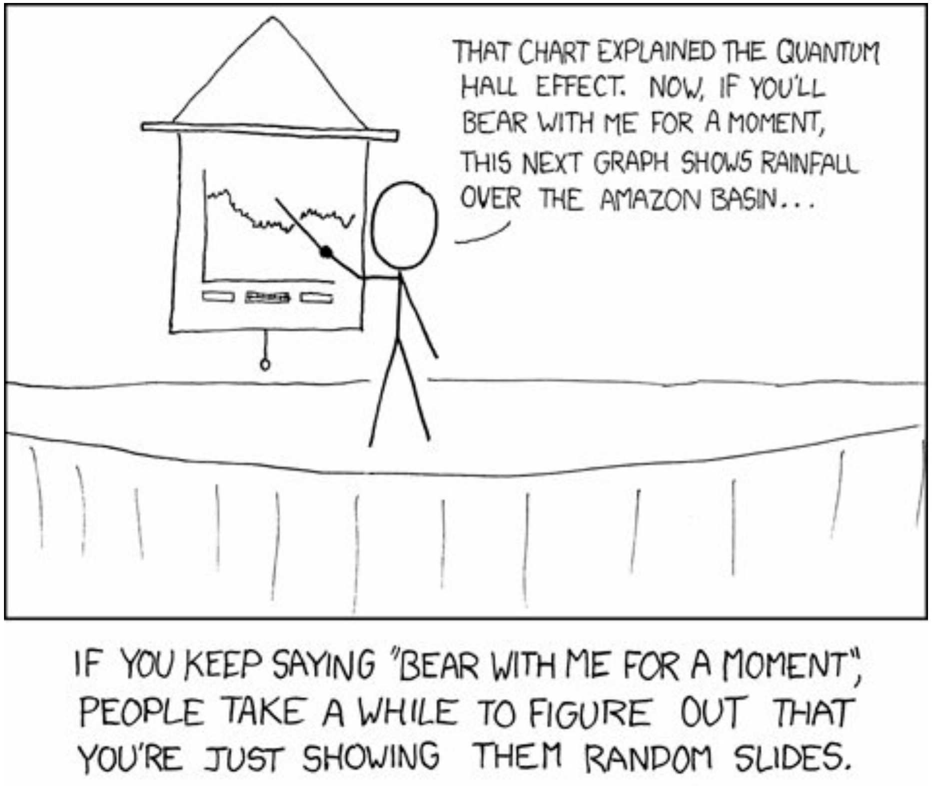
Easy, flexible and powerful service for life scientists

Katerina Michalickova, Nikolay Vazov, Thomas Röblitz

USIT / ITF / FT & FI

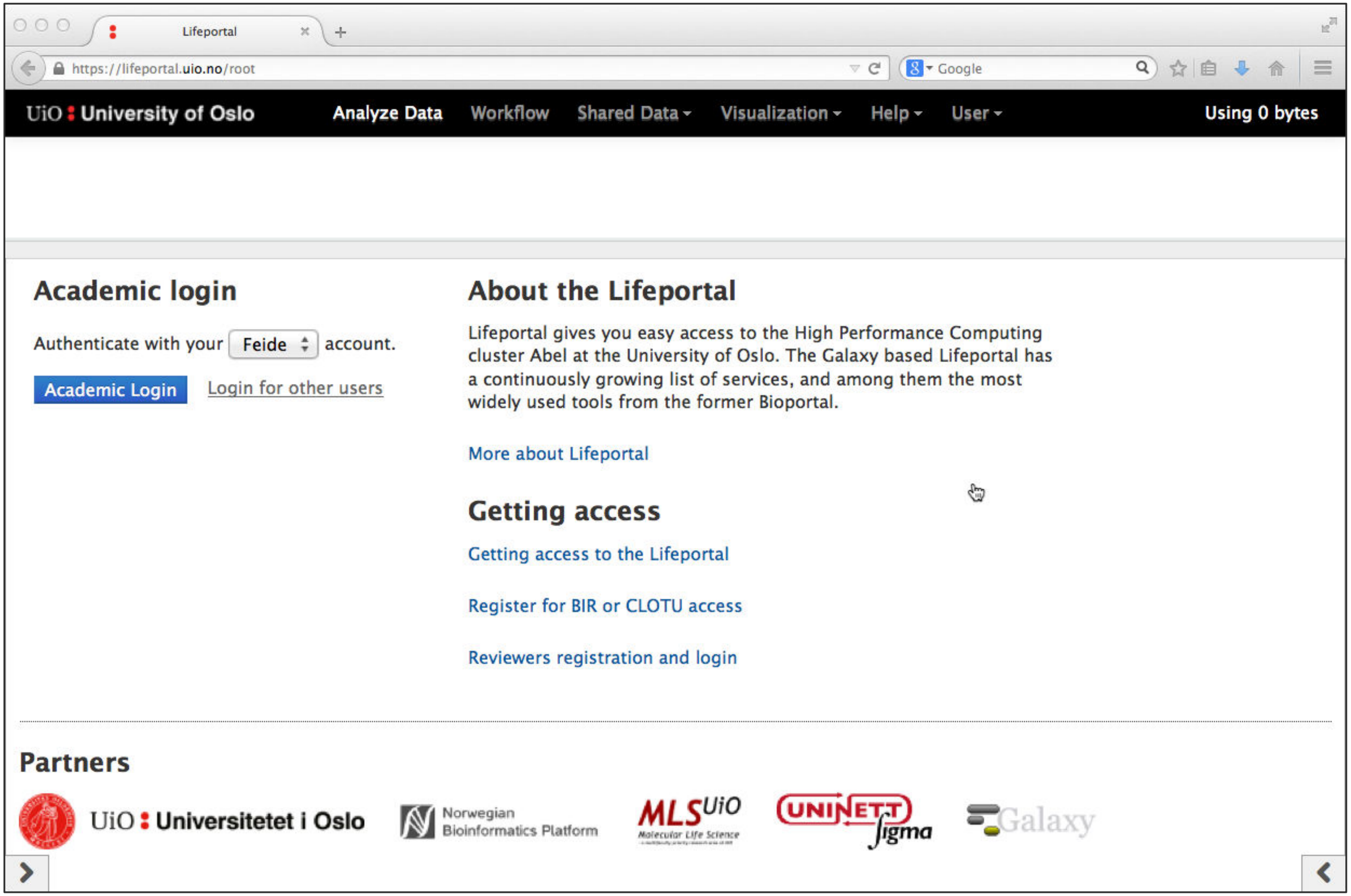
Topics

- Intro
- Sales
- Demo
- Implementation
- Roadmap



© xkcd.com

Intro



Academic login

Authenticate with your Feide account.

About the Lifeportal

Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.

More about Lifeportal

Getting access

Getting access to the Lifeportal

Register for BIR or CLOTU access

Reviewers registration and login

Partners

Objectives

Introduce beginners to Lifeportal, and encourage existing users to make it even better by challenging us !

Sales – Key benefits

© www.golime.co

- Easy
  - web frontend to many life science tools
  - integrated sharing and reproducibility
- Flexible
  - customizations
- Powerful
  - HPC cluster Abel as backend (10000+ cores)
- Service
  - user support

Sales – How much does it cost you ?

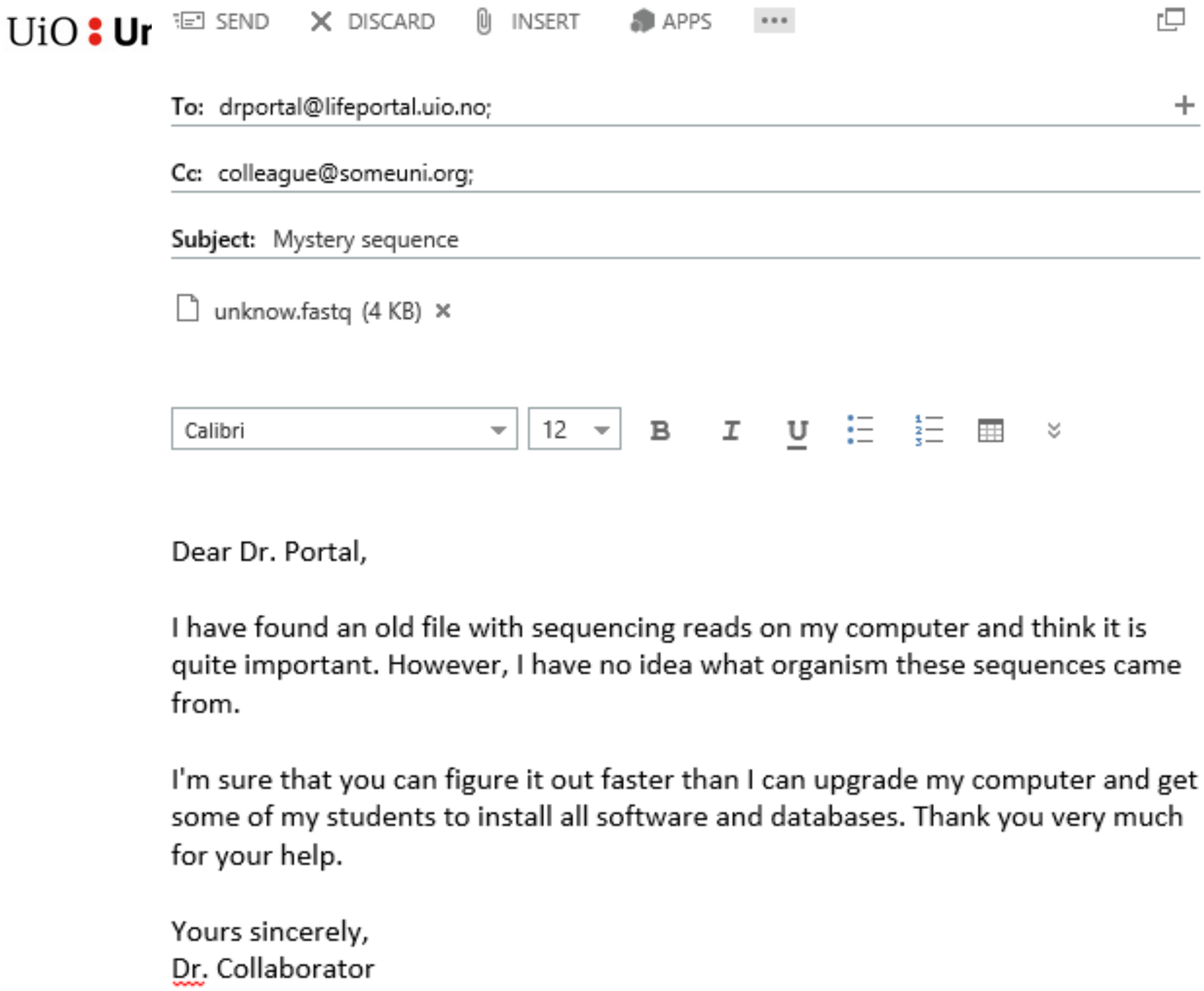
- per core hour
- per GB storage
- per bytes transferred
- per tool installed
- per support request
- per hour spend on system maintenance

It’s free !!!

Demo

- Run a simple job
- Run another job
- Create a workflow
- Sharing
- Pages

Demo Abel handshake



To: drportal@lifeportal.uio.no;

Cc: colleague@someuni.org;

Subject: Mystery sequence

unknow.fastq (4 KB)

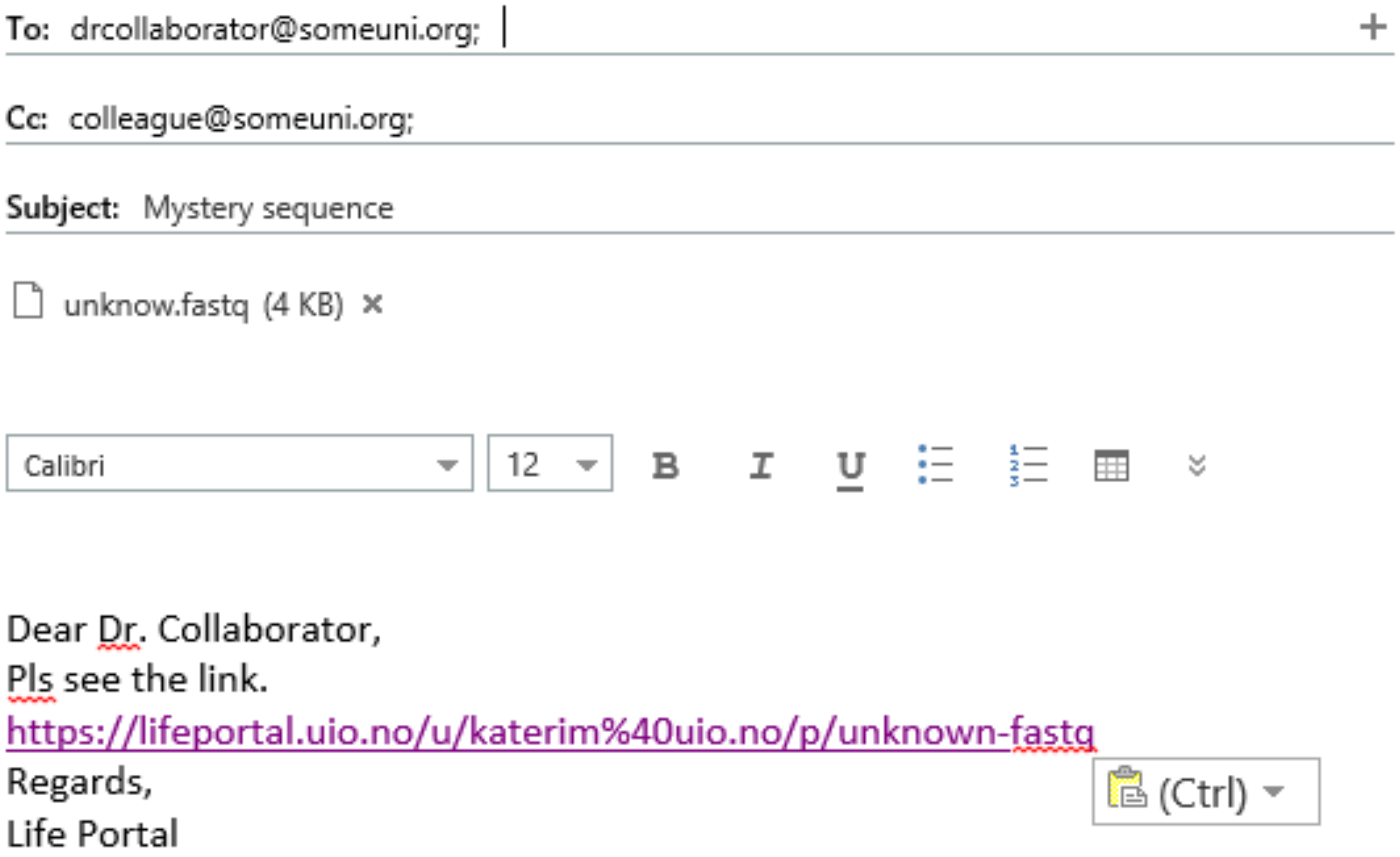
Dear Dr. Portal,

I have found an old file with sequencing reads on my computer and think it is quite important. However, I have no idea what organism these sequences came from.

I'm sure that you can figure it out faster than I can upgrade my computer and get some of my students to install all software and databases. Thank you very much for your help.

Yours sincerely,
Dr. Collaborator

Demo



To: drcollaborator@someuni.org;

Cc: colleague@someuni.org;

Subject: Mystery sequence

unknow.fastq (4 KB)

Dear Dr. Collaborator,
Pls see the link.
https://lifeportal.uio.no/u/katerim%40uio.no/p/unknown-fastq
Regards,
Life Portal

Some useful Lifeportal features

- (chose) your user status ?
    - «simple» user
    - project manager

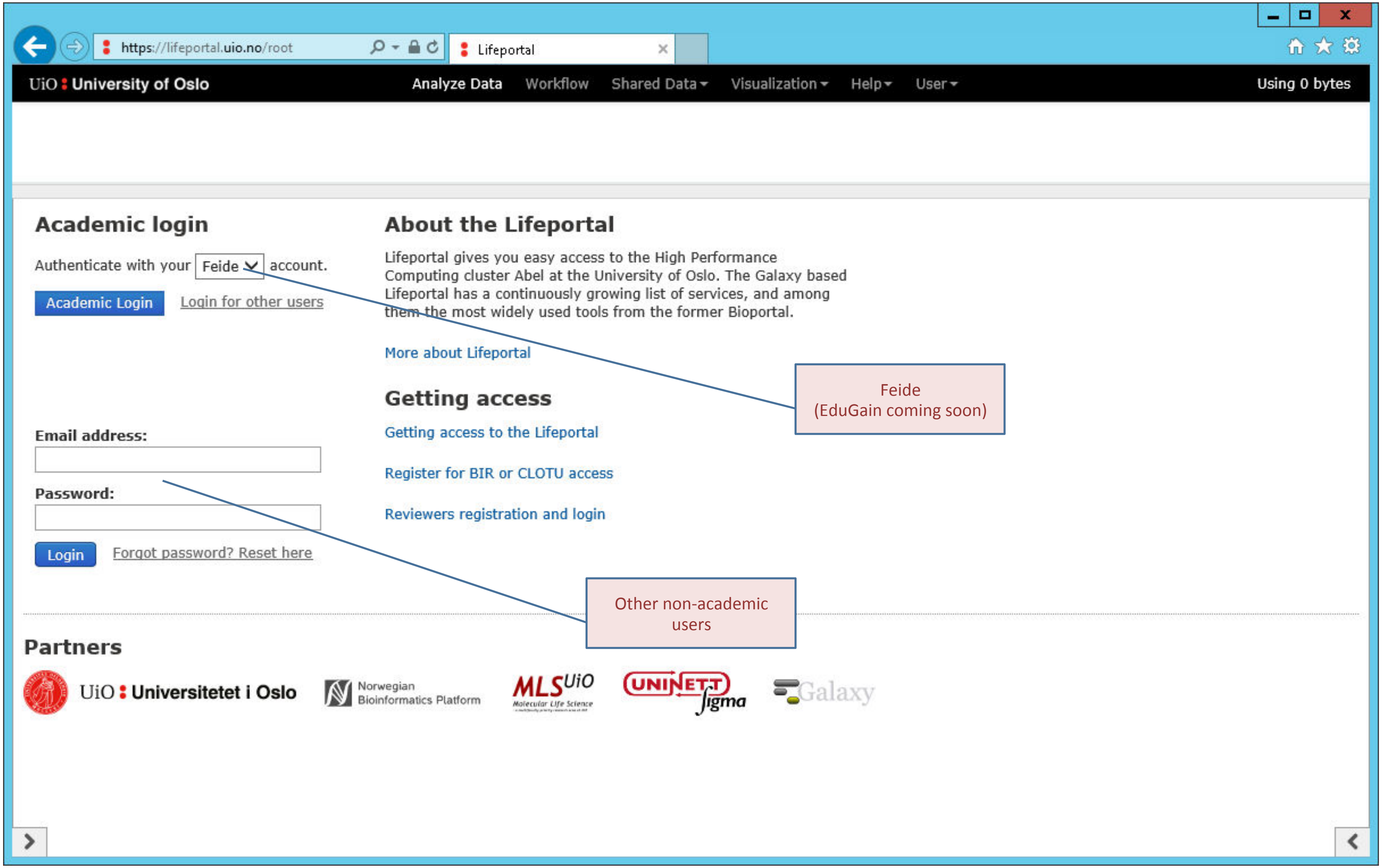
- Project management

- Jobs running on a cluster

Lifeportal features : users

- User status :

  1. FEIDE users
  2. EduGAIN (soon) – 900 institutions
  3. Users (external collaborators) added by users



Academic login

Authenticate with your Feide account.

Email address:

Password:

About the Lifeportal

Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.

More about Lifeportal

Getting access

Getting access to the Lifeportal

Register for BIR or CLOTU access

Reviewers registration and login

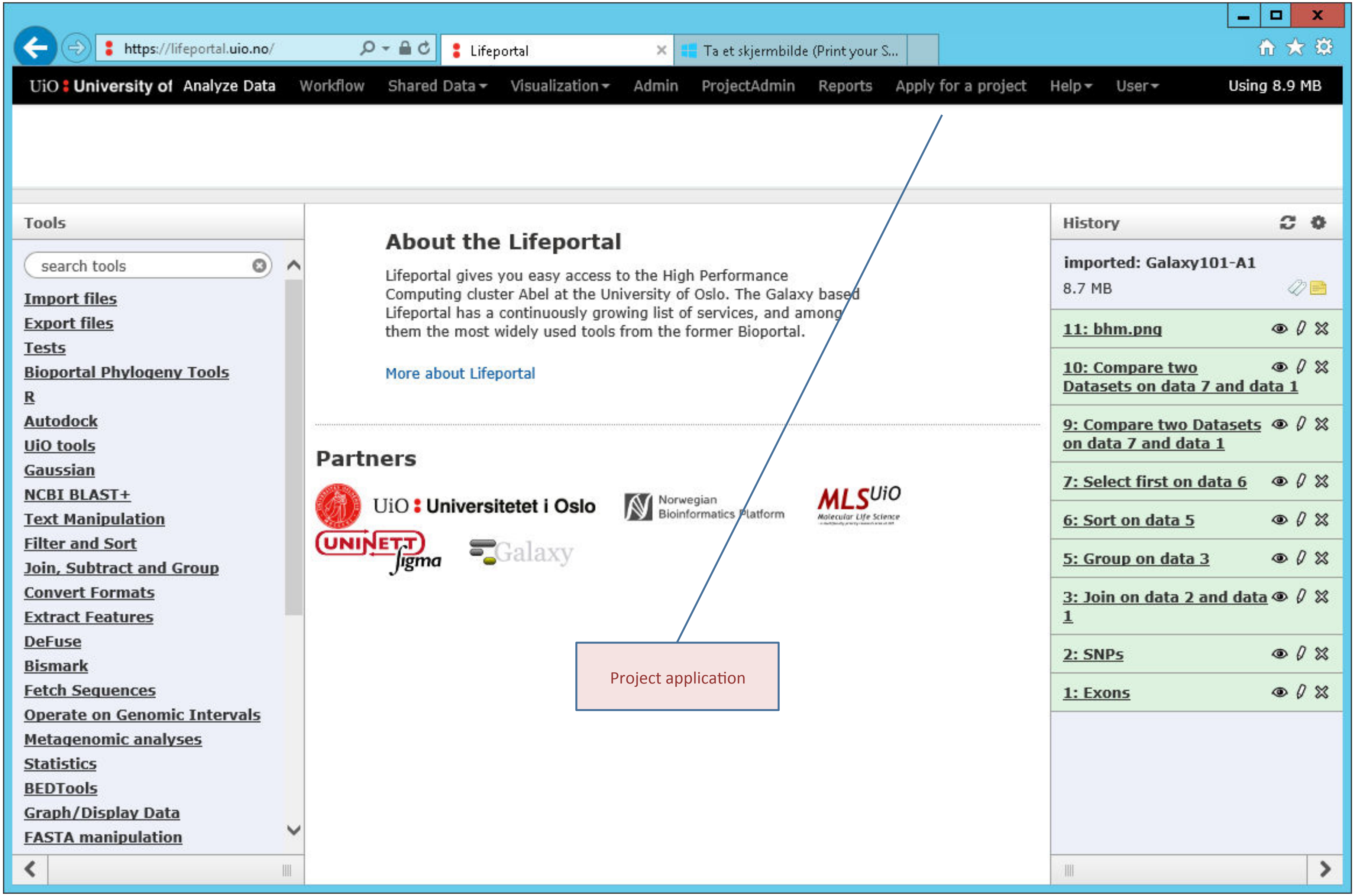
Partners

Feide
(EduGain coming soon)

Other non-academic users

Lifeportal features : project management

1. Lifeportal allows the users to apply for projects and become ***project managers***
2. Project managers manage :
   - Project members : add, delete, assign other (new) users to projects
   - Project resources : period of activity and CPU hrs



Import files
Export files
Tests
Bioportal Phylogeny Tools
R
Autodock
UiO tools
Gaussian
NCBI BLAST+
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
DeFuse
Bismark
Fetch Sequences
Operate on Genomic Intervals
Metagenomic analyses
Statistics
BEDTools
Graph/Display Data
FASTA manipulation

About the Lifeportal

Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.

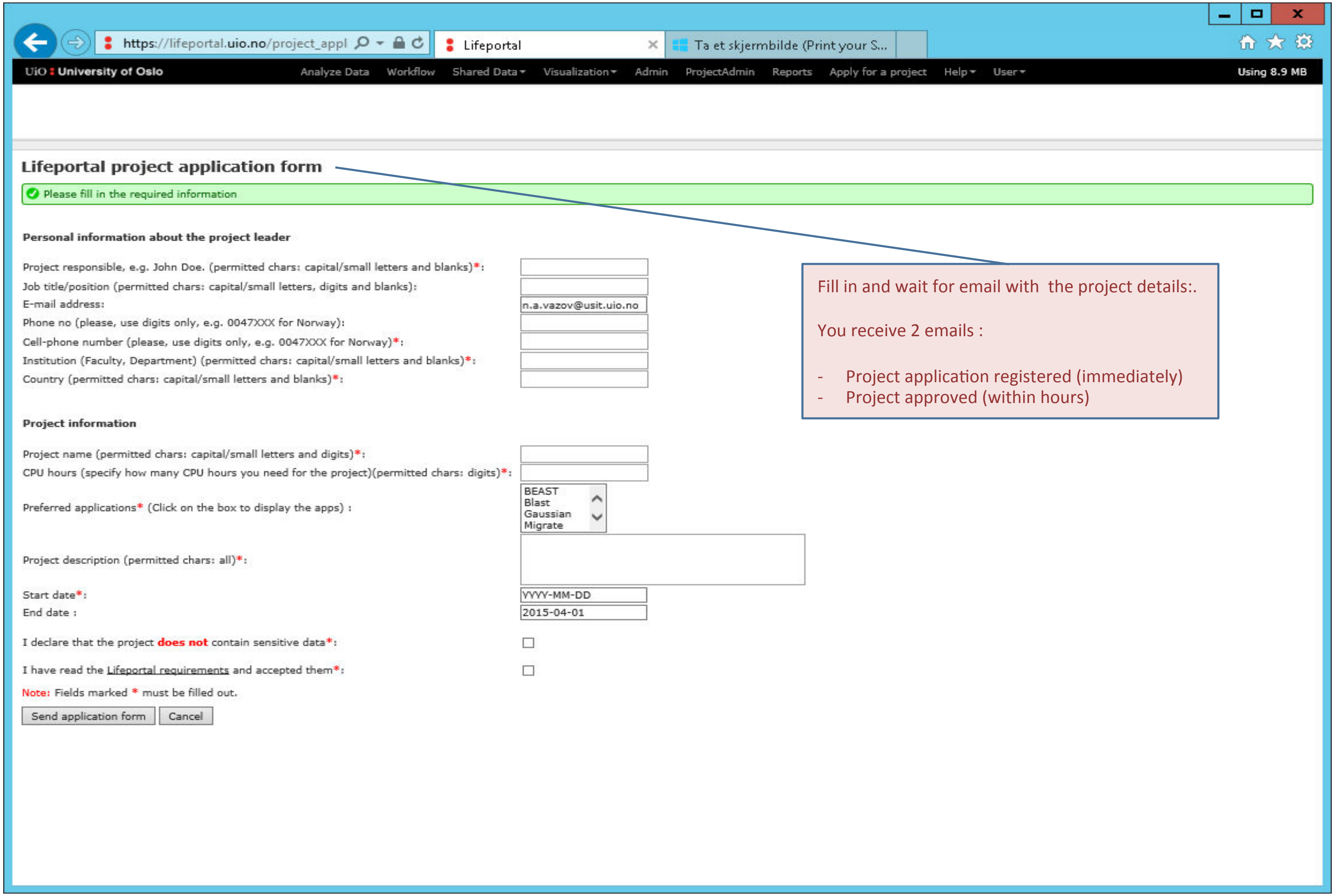
More about Lifeportal

Partners

imported: Galaxy101-A1
8.7 MB

11: bhm.png
10: Compare two Datasets on data 7 and data 1
9: Compare two Datasets on data 7 and data 1
7: Select first on data 6
6: Sort on data 5
5: Group on data 3
3: Join on data 2 and data 1
2: SNPs
1: Exons

Project application



Lifeportal project application form

Please fill in the required information

Personal information about the project leader

Project responsible, e.g. John Doe. (permitted chars: capital/small letters and blanks)*:
Job title/position (permitted chars: capital/small letters, digits and blanks):
E-mail address: n.a.vazov@usit.uio.no
Phone no (please, use digits only, e.g. 0047XXX for Norway):
Cell-phone number (please, use digits only, e.g. 0047XXX for Norway)*:
Institution (Faculty, Department) (permitted chars: capital/small letters and blanks)*:
Country (permitted chars: capital/small letters and blanks)*:

Project information

Project name (permitted chars: capital/small letters and digits)*:
CPU hours (specify how many CPU hours you need for the project)(permitted chars: digits)*:
Preferred applications* (Click on the box to display the apps) : BEAST
Blast
Gaussian
Migrate
Project description (permitted chars: all)*:
Start date*: YYYY-MM-DD
End date : 2015-04-01

I declare that the project **does not** contain sensitive data*: ☐

I have read the Lifeportal requirements and accepted them*: ☐

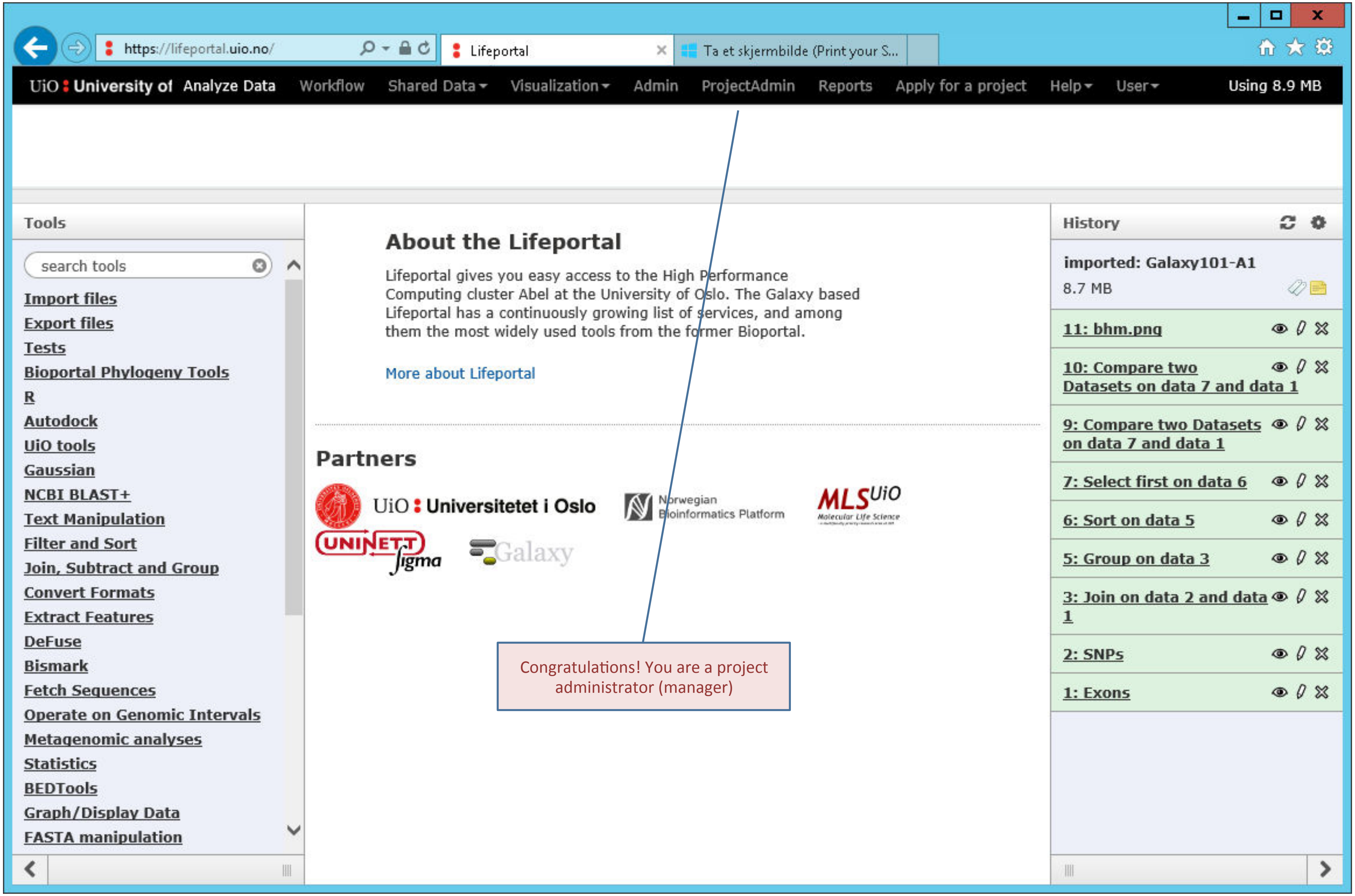
Note: Fields marked * must be filled out.

Fill in and wait for email with the project details:.

You receive 2 emails :

- Project application registered (immediately)
- Project approved (within hours)

UiO : Universitetet i Oslo

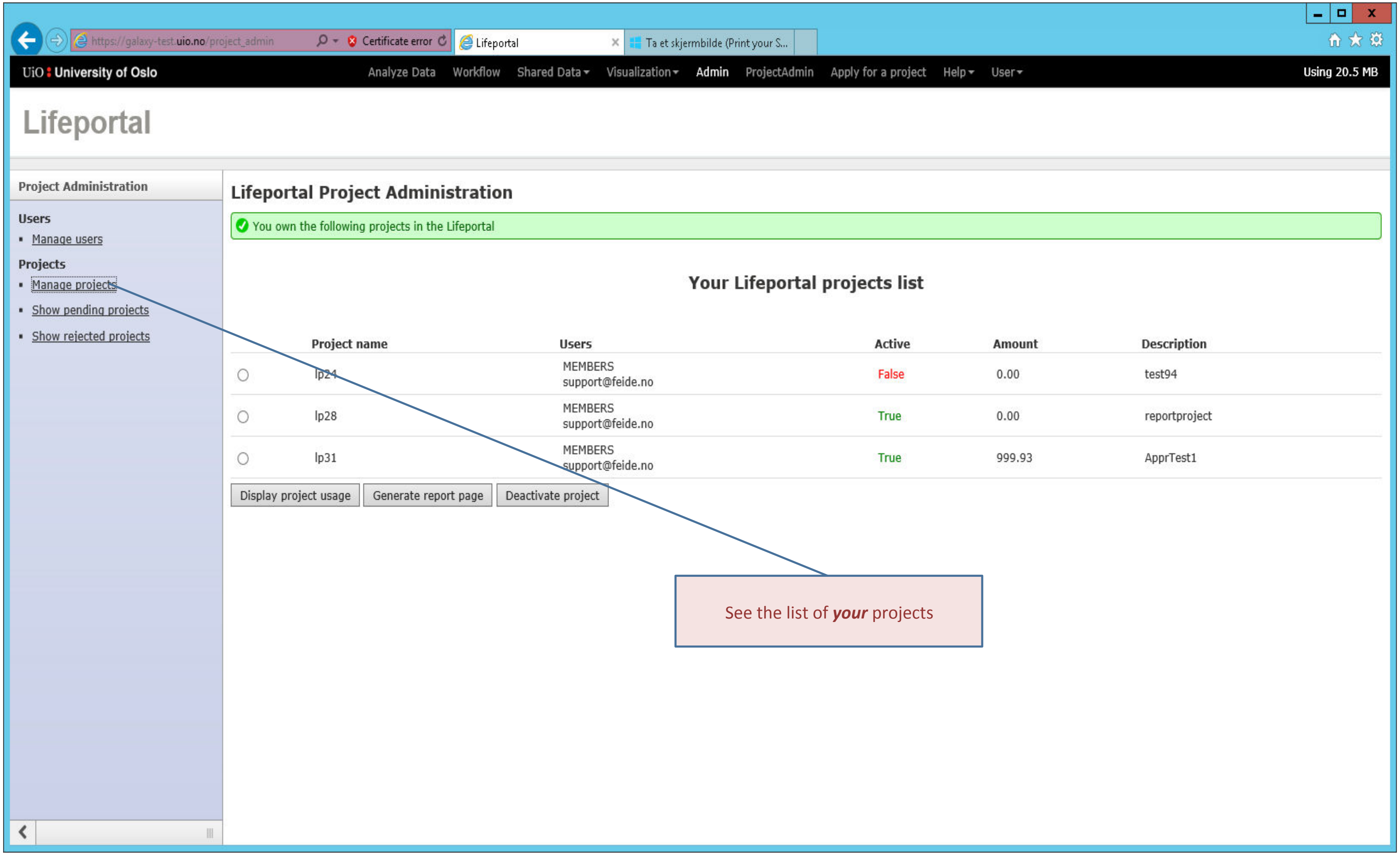


About the Lifeportal

Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.

More about Lifeportal

Congratulations! You are a project administrator (manager)



Lifeportal Project Administration

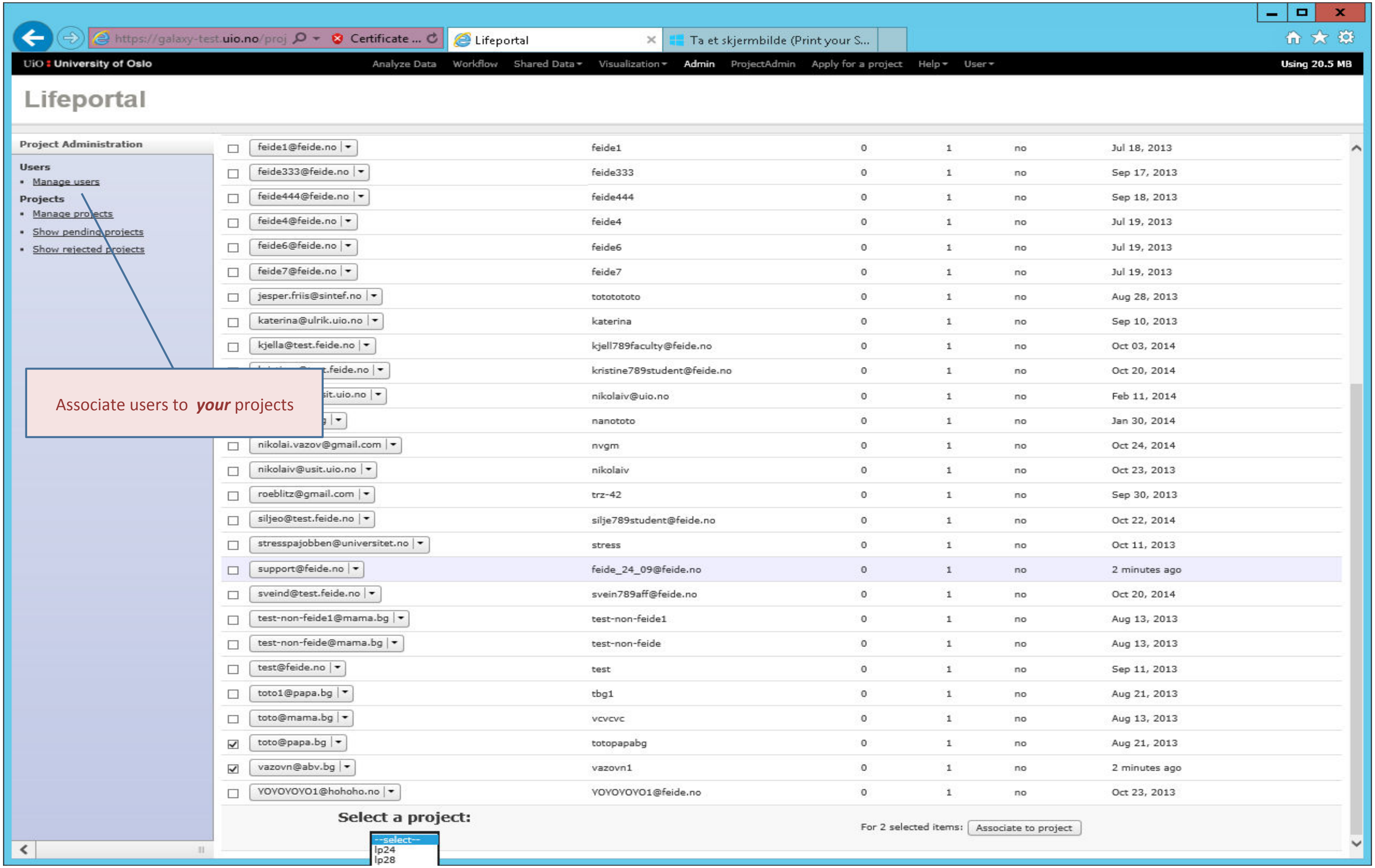
✓ You own the following projects in the Lifeportal

Your Lifeportal projects list

	Project name	Users	Active	Amount	Description
○	lp24	MEMBERS support@feide.no	False	0.00	test94
○	lp28	MEMBERS support@feide.no	True	0.00	reportproject
○	lp31	MEMBERS support@feide.no	True	999.93	ApprTest1

See the list of ***your*** projects

UiO : Universitetet i Oslo



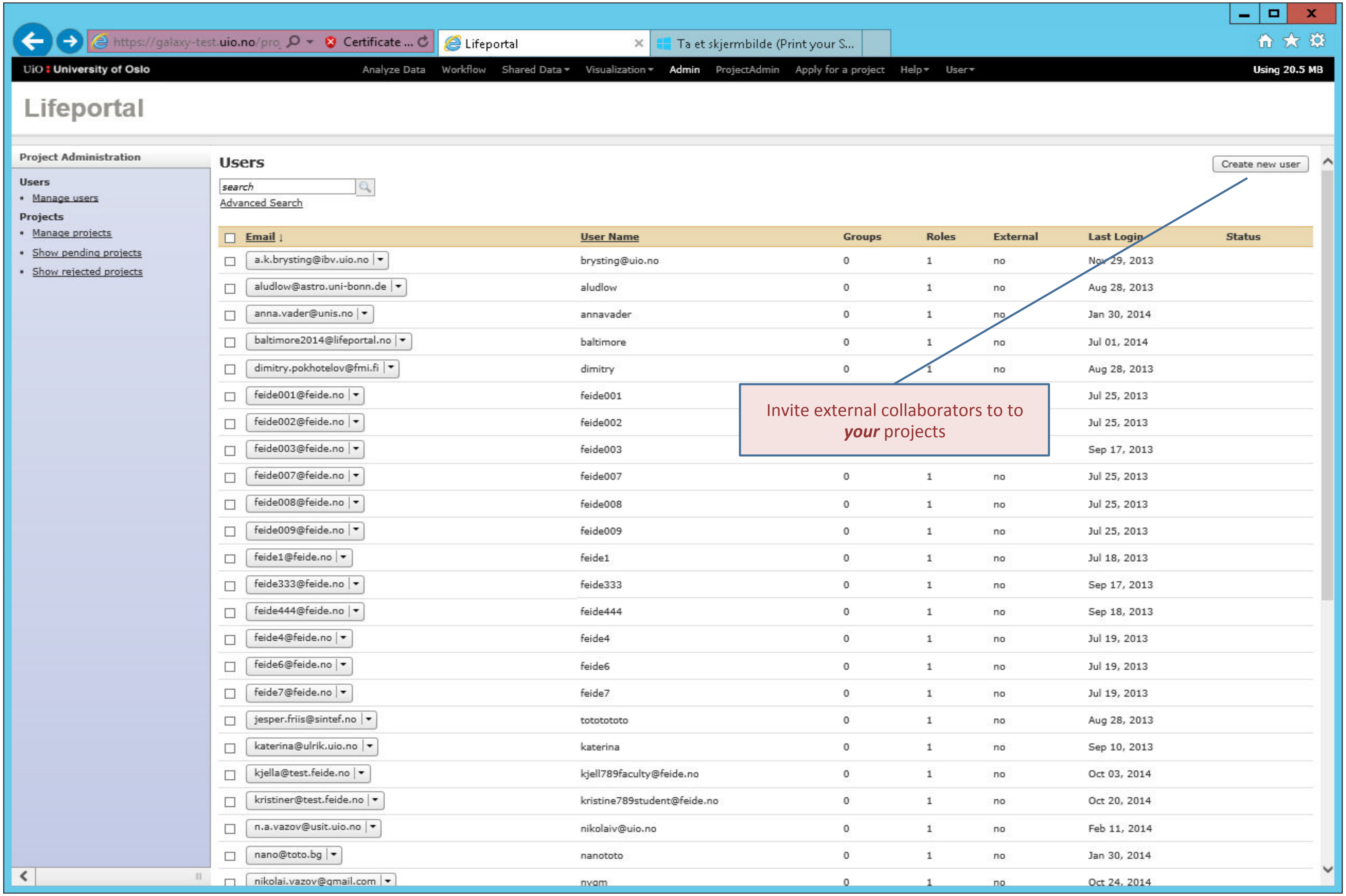
Associate users to ***your*** projects

☐	feide1@feide.no	feide1	0	1	no	Jul 18, 2013
☐	feide333@feide.no	feide333	0	1	no	Sep 17, 2013
☐	feide444@feide.no	feide444	0	1	no	Sep 18, 2013
☐	feide4@feide.no	feide4	0	1	no	Jul 19, 2013
☐	feide6@feide.no	feide6	0	1	no	Jul 19, 2013
☐	feide7@feide.no	feide7	0	1	no	Jul 19, 2013
☐	jesper.friis@sintef.no	tototototo	0	1	no	Aug 28, 2013
☐	katerina@ulrik.uio.no	katerina	0	1	no	Sep 10, 2013
☐	kjella@test.feide.no	kjell789faculty@feide.no	0	1	no	Oct 03, 2014
	t.feide.no	kristine789student@feide.no	0	1	no	Oct 20, 2014
	sit.uio.no	nikolaiv@uio.no	0	1	no	Feb 11, 2014
	g	nanototo	0	1	no	Jan 30, 2014
☐	nikolai.vazov@gmail.com	nvgm	0	1	no	Oct 24, 2014
☐	nikolaiv@usit.uio.no	nikolaiv	0	1	no	Oct 23, 2013
☐	roeblitz@gmail.com	trz-42	0	1	no	Sep 30, 2013
☐	siljeo@test.feide.no	silje789student@feide.no	0	1	no	Oct 22, 2014
☐	stresspajobben@universitet.no	stress	0	1	no	Oct 11, 2013
☐	support@feide.no	feide_24_09@feide.no	0	1	no	2 minutes ago
☐	sveind@test.feide.no	svein789aff@feide.no	0	1	no	Oct 20, 2014
☐	test-non-feide1@mama.bg	test-non-feide1	0	1	no	Aug 13, 2013
☐	test-non-feide@mama.bg	test-non-feide	0	1	no	Aug 13, 2013
☐	test@feide.no	test	0	1	no	Sep 11, 2013
☐	toto1@papa.bg	tbg1	0	1	no	Aug 21, 2013
☐	toto@mama.bg	vcvcvc	0	1	no	Aug 13, 2013
☑	toto@papa.bg	totopapabg	0	1	no	Aug 21, 2013
☑	vazovn@abv.bg	vazovn1	0	1	no	2 minutes ago
☐	YOYOYOYO1@hohoho.no	YOYOYOYO1@feide.no	0	1	no	Oct 23, 2013

Select a project:

--select--
lp24
lp28

For 2 selected items: Associate to project

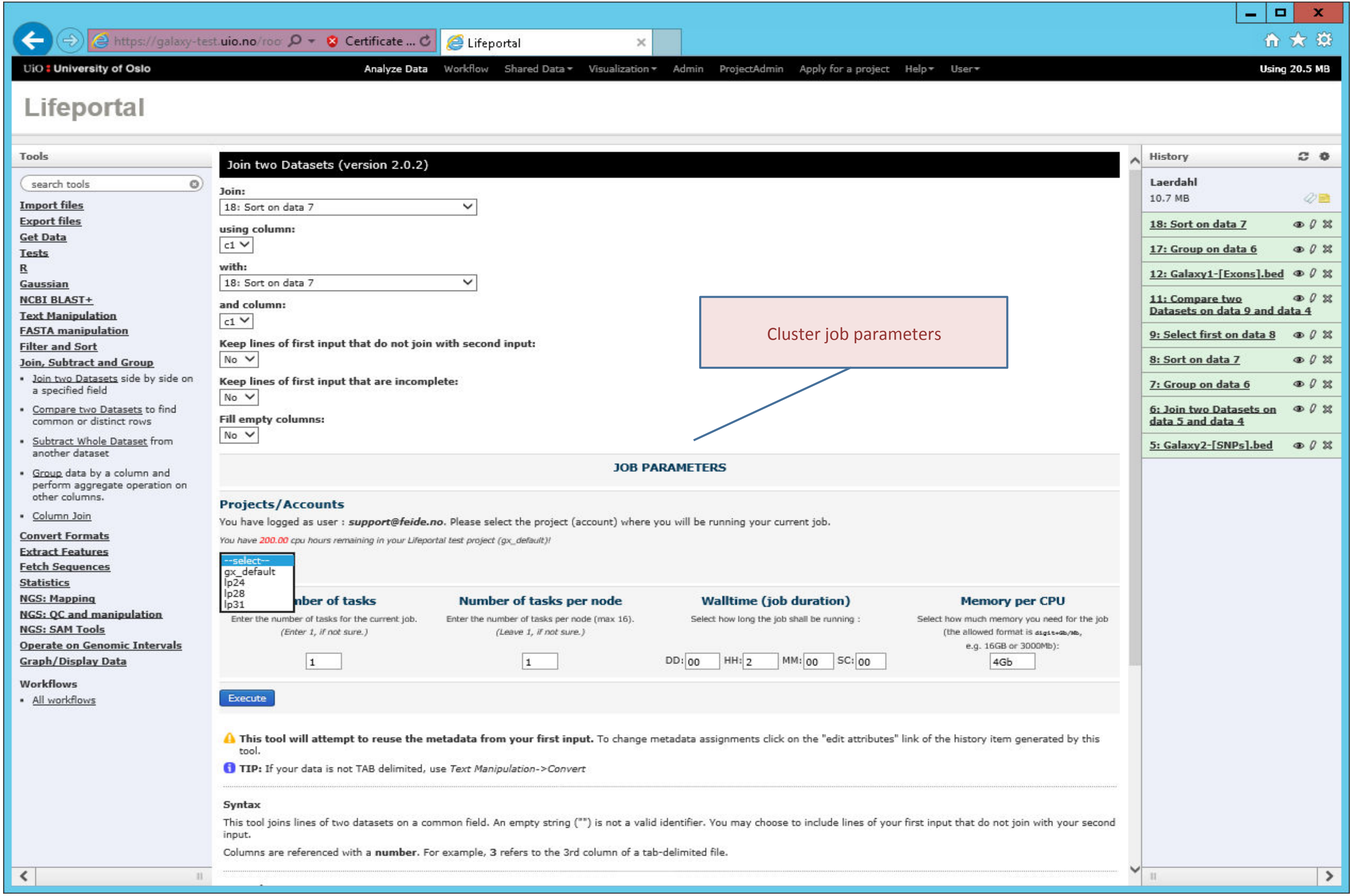


☐ Email ↓	User Name	Groups	Roles	External	Last Login	Status
☐ a.k.brysting@ibv.uio.no	brysting@uio.no	0	1	no	Nov 29, 2013	
☐ aludlow@astro.uni-bonn.de	aludlow	0	1	no	Aug 28, 2013	
☐ anna.vader@unis.no	annavader	0	1	no	Jan 30, 2014	
☐ baltimore2014@lifeportal.no	baltimore	0	1	no	Jul 01, 2014	
☐ dimitry.pokhotelov@fmi.fi	dimitry	0	1	no	Aug 28, 2013	
☐ feide001@feide.no	feide001				Jul 25, 2013	
☐ feide002@feide.no	feide002				Jul 25, 2013	
☐ feide003@feide.no	feide003				Sep 17, 2013	
☐ feide007@feide.no	feide007	0	1	no	Jul 25, 2013	
☐ feide008@feide.no	feide008	0	1	no	Jul 25, 2013	
☐ feide009@feide.no	feide009	0	1	no	Jul 25, 2013	
☐ feide1@feide.no	feide1	0	1	no	Jul 18, 2013	
☐ feide333@feide.no	feide333	0	1	no	Sep 17, 2013	
☐ feide444@feide.no	feide444	0	1	no	Sep 18, 2013	
☐ feide4@feide.no	feide4	0	1	no	Jul 19, 2013	
☐ feide6@feide.no	feide6	0	1	no	Jul 19, 2013	
☐ feide7@feide.no	feide7	0	1	no	Jul 19, 2013	
☐ jesper.friis@sintef.no	totototototo	0	1	no	Aug 28, 2013	
☐ katerina@ulrik.uio.no	katerina	0	1	no	Sep 10, 2013	
☐ kjella@test.feide.no	kjell789faculty@feide.no	0	1	no	Oct 03, 2014	
☐ kristiner@test.feide.no	kristine789student@feide.no	0	1	no	Oct 20, 2014	
☐ n.a.vazov@usit.uio.no	nikolaiv@uio.no	0	1	no	Feb 11, 2014	
☐ nano@toto.bg	nanototo	0	1	no	Jan 30, 2014	
☐ nikolai.vazov@gmail.com	nvgm	0	1	no	Oct 24, 2014	

Invite external collaborators to to *your* projects

Lifeportal features : cluster jobs

- Cluster jobs
  - Users input is minimal
  - Galaxy platform uses the potential of a powerful cluster



Join two Datasets (version 2.0.2)

Join: 18: Sort on data 7

using column: c1

with: 18: Sort on data 7

and column: c1

Keep lines of first input that do not join with second input: No

Keep lines of first input that are incomplete: No

Fill empty columns: No

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***support@feide.no***. Please select the project (account) where you will be running your current job.

You have 200.00 cpu hours remaining in your Lifeportal test project (gx_default)!

--select--
gx_default
lp24
lp28
lp31

...mber of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)
1

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)
1

Walltime (job duration)
Select how long the job shall be running :
DD: 00 HH: 2 MM: 00 SC: 00

Memory per CPU
Select how much memory you need for the job (the allowed format is digit+Gb/Mb, e.g. 16GB or 3000Mb):
4Gb

This tool will attempt to reuse the metadata from your first input. To change metadata assignments click on the "edit attributes" link of the history item generated by this tool.

TIP: If your data is not TAB delimited, use *Text Manipulation->Convert*

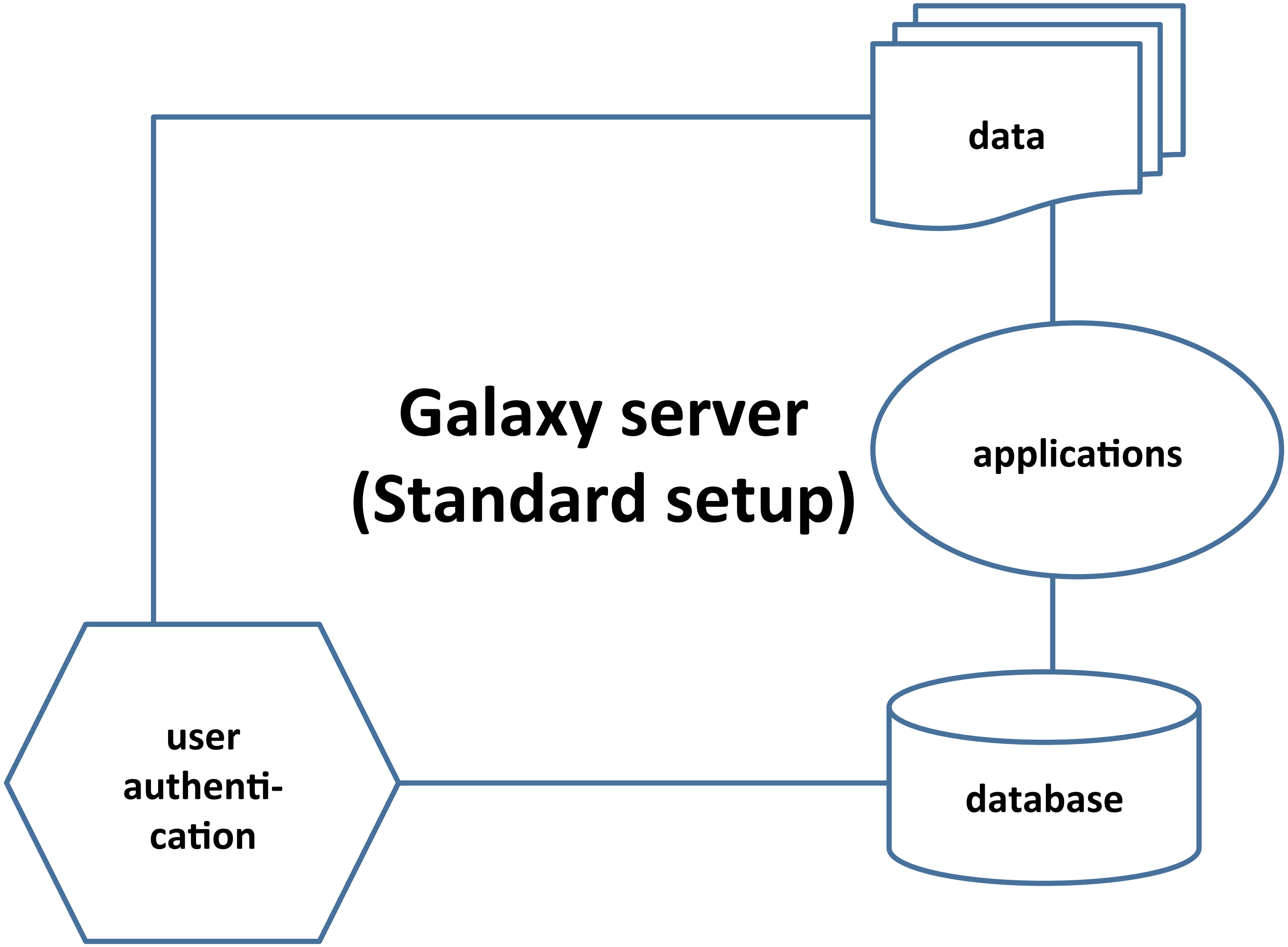
Syntax

This tool joins lines of two datasets on a common field. An empty string ("") is not a valid identifier. You may choose to include lines of your first input that do not join with your second input.

Columns are referenced with a **number**. For example, **3** refers to the 3rd column of a tab-delimited file.

Cluster job parameters

Galaxy server (Standard setup)



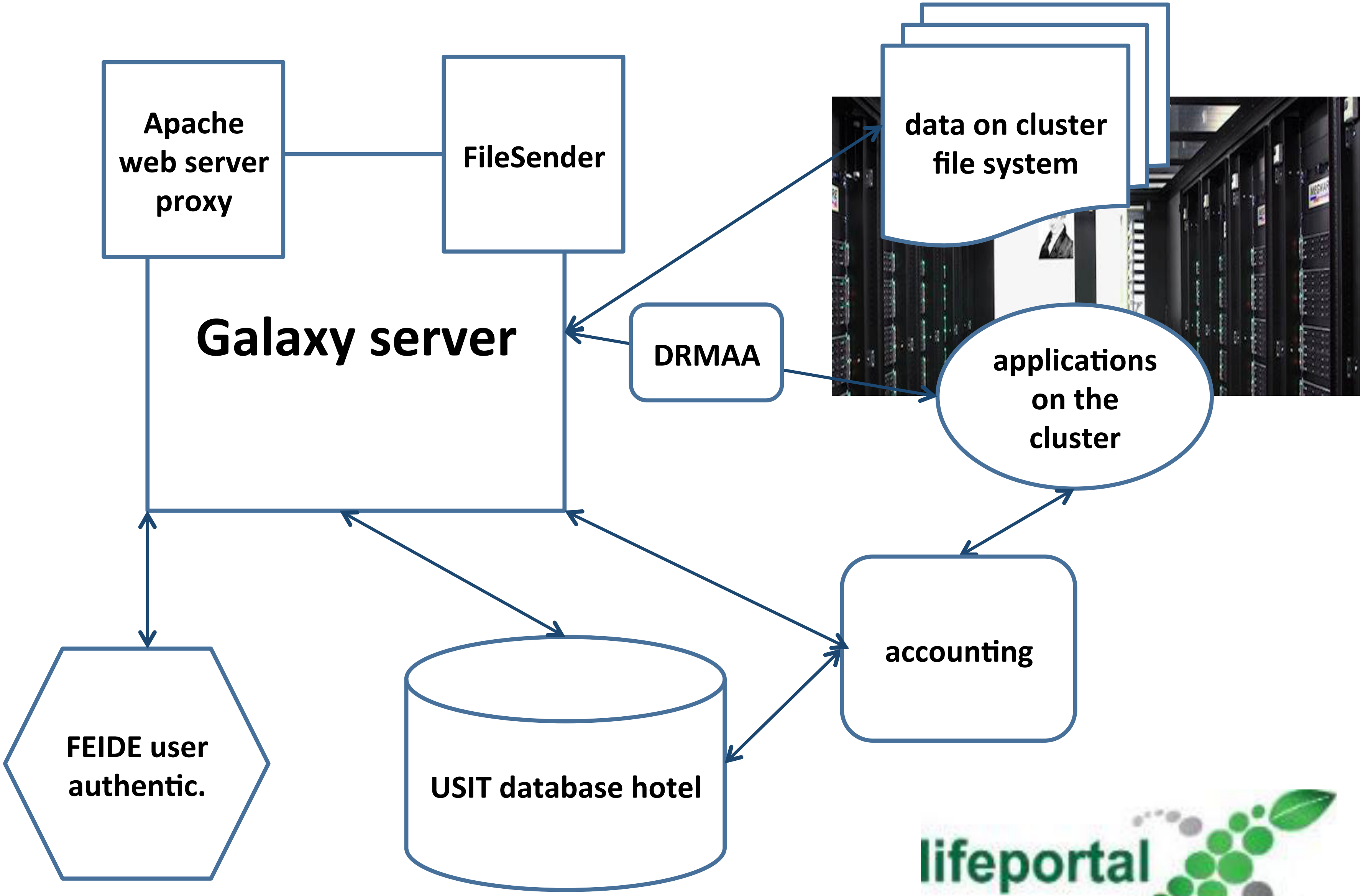
data

applications

database

user authentication

galaxyproject.org



Apache web server proxy

FileSender

Galaxy server

data on cluster file system

DRMAA

applications on the cluster

accounting

FEIDE user authentic.

USIT database hotel

lifeportal

History

© www.xplusone.net

- Bioportal
  - started 2005
  - home-grown from scratch
  - world-wide user base
- Lifeportal
  - opening October 2013 (just 13 months)
  - about 200 users registered
  - more than 50 projects
  - continuously adding new features and tools

Lifeportal vs. Galaxy

	Lifeportal	'Own' Galaxy Instance
Easy access	yes	yes
Support	Well trained helpdesk	-
Collaboration	FEIDE + projects	Limited
Scalability	Abel (10000+ cores)	Depends on your credit card
Customizations	In-house + synergies with emerging portals	Depends on your system level programming skills
Costs	free	Admin, Compute +Storage+Network, Support, Customizations

© iftools.com

Roadmap

- Synchronizing code bases & faster upgrades
  - bulk operations, …
- Deploying dedicated hardware
  - faster turn-around time for short / small tasks
  - better response time of the frontend
  - improved performance for data transfers

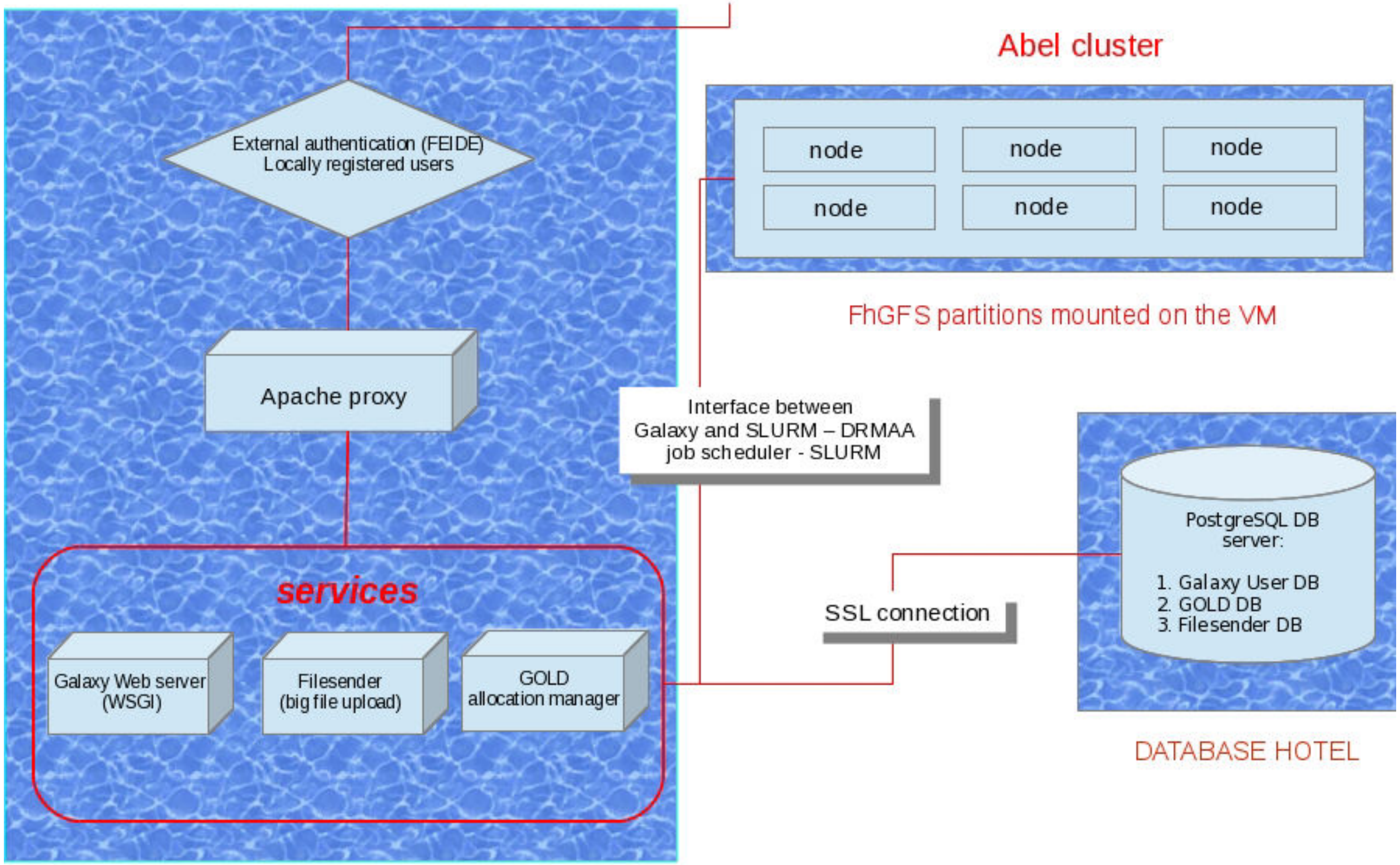
Making Lifeportal even better

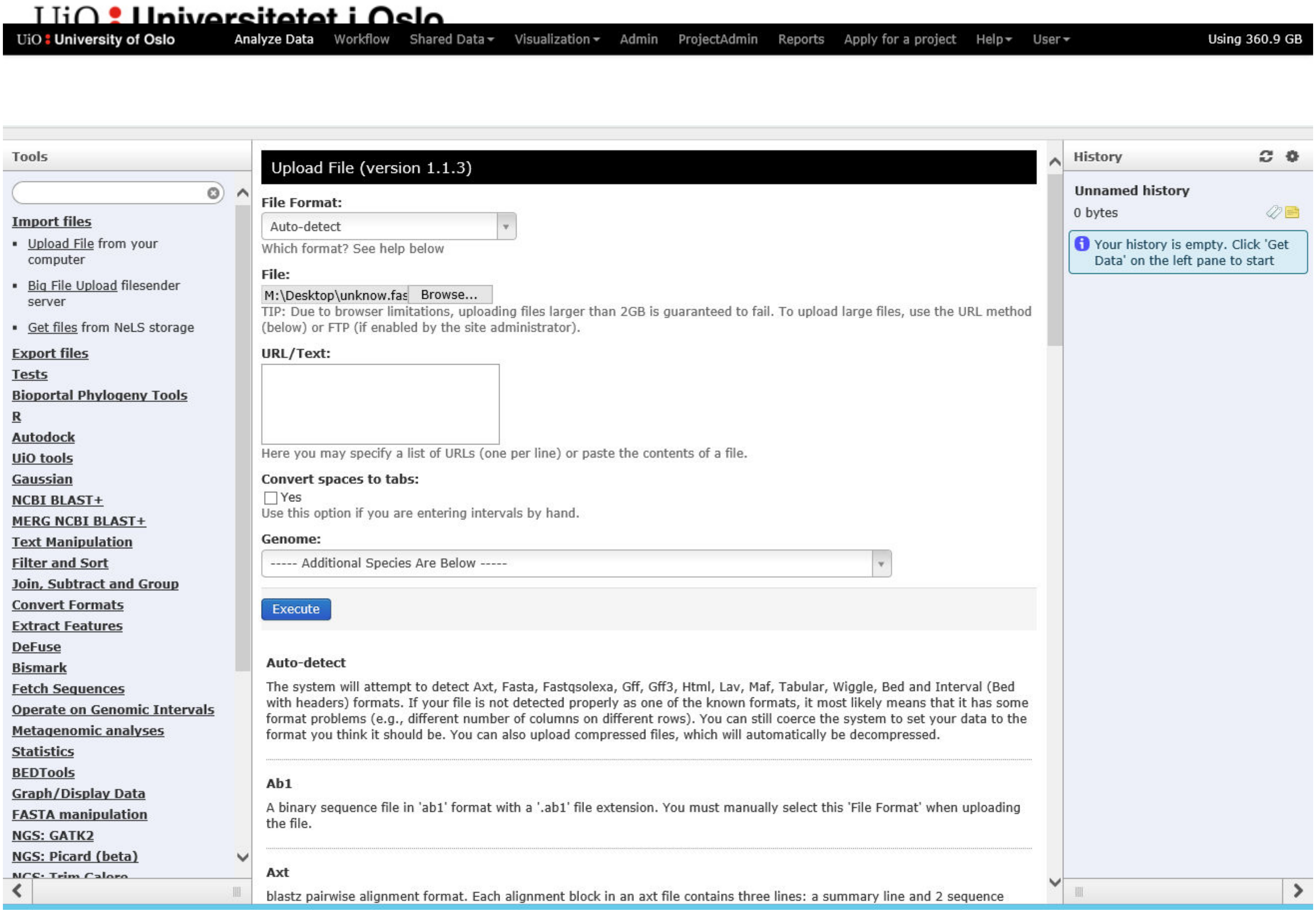
- Lifeportal is great for beginners, but not only for them !
- Opportunities
  - enhancing the service (simpler, richer capabilities, more tools)
  - funding via Sigma2
- Looking forward to your ideas / requirements
  - try and use Lifeportal
  - challenge us with features to add and improve

Summary

- Lifeportal
  - easy, flexible, powerful service
  - sharing, reproducibility
  - support
- Roadmap
  - several significant improvements coming
- USIT / ITF is committed to work with you to make the Lifeportal even better !

Lifeportal Architecture





- Upload File from your computer
- Big File Upload filesender server
- Get files from NeLS storage

Upload File (version 1.1.3)

File Format: Auto-detect

Which format? See help below

File: M:\Desktop\unknow.fas

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Convert spaces to tabs:

☐ Yes

Use this option if you are entering intervals by hand.

Genome: ----- Additional Species Are Below -----

Auto-detect

The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats. If your file is not detected properly as one of the known formats, it most likely means that it has some format problems (e.g., different number of columns on different rows). You can still coerce the system to set your data to the format you think it should be. You can also upload compressed files, which will automatically be decompressed.

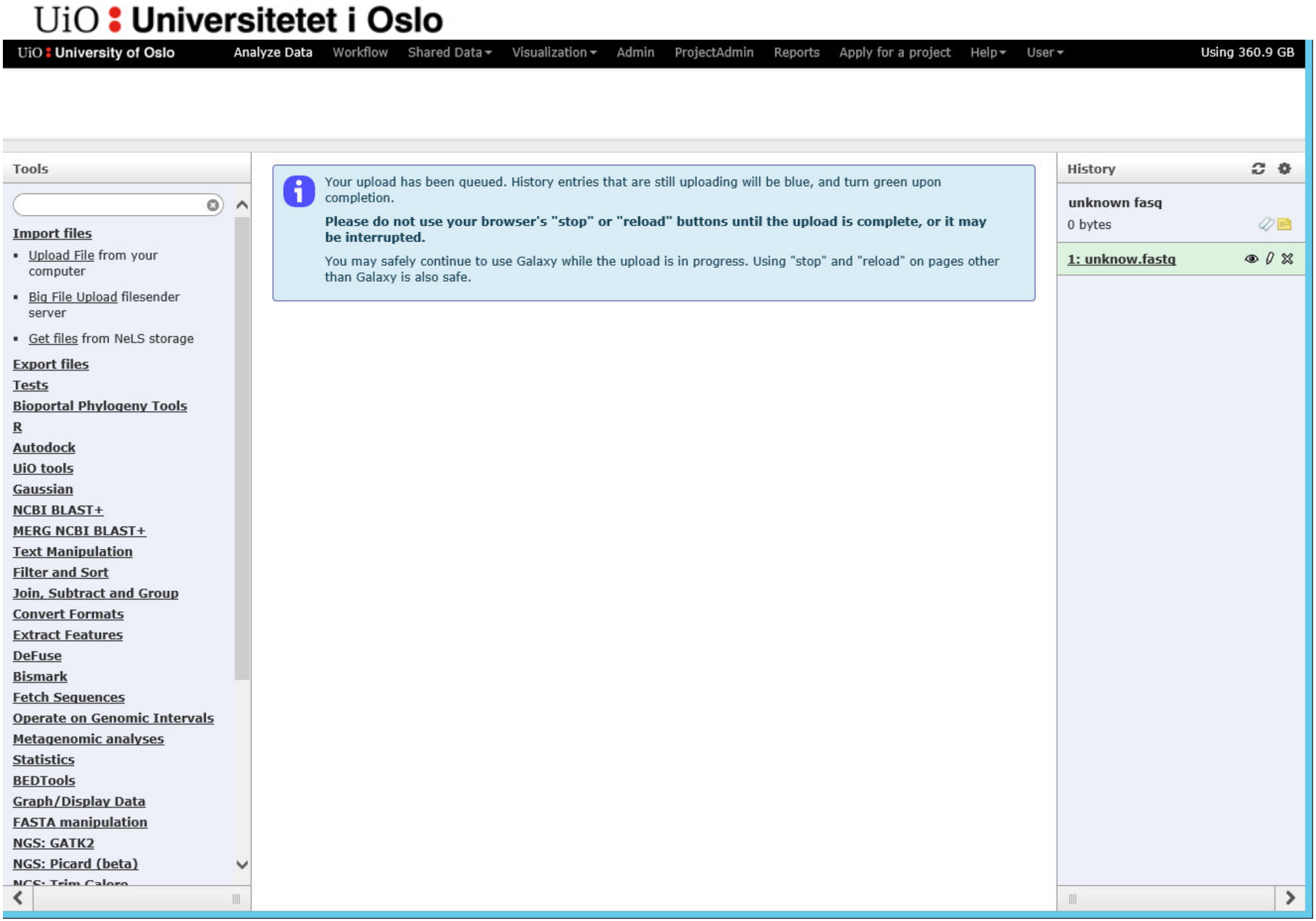
Ab1

A binary sequence file in 'ab1' format with a '.ab1' file extension. You must manually select this 'File Format' when uploading the file.

Axt

blastz pairwise alignment format. Each alignment block in an axt file contains three lines: a summary line and 2 sequence

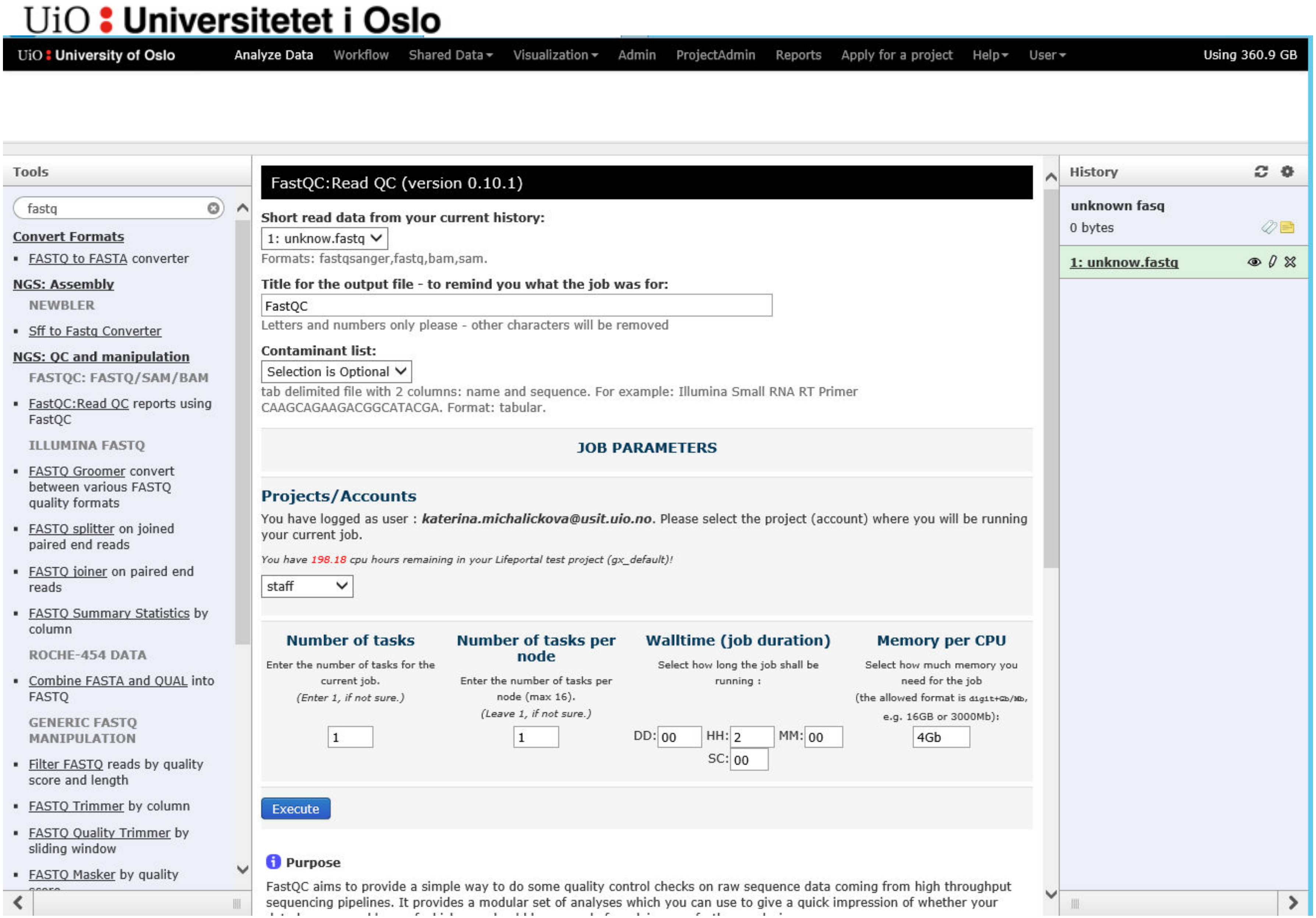
Your history is empty. Click 'Get Data' on the left pane to start



Your upload has been queued. History entries that are still uploading will be blue, and turn green upon completion.

Please do not use your browser's "stop" or "reload" buttons until the upload is complete, or it may be interrupted.

You may safely continue to use Galaxy while the upload is in progress. Using "stop" and "reload" on pages other than Galaxy is also safe.



fastq

Convert Formats

- FASTQ to FASTA converter

NGS: Assembly

NEWBLER

- Sff to Fastq Converter

NGS: QC and manipulation

FASTQC: FASTQ/SAM/BAM

- FastQC:Read QC reports using FastQC

ILLUMINA FASTQ

- FASTQ Groomer convert between various FASTQ quality formats
- FASTQ splitter on joined paired end reads
- FASTQ joiner on paired end reads
- FASTQ Summary Statistics by column

ROCHE-454 DATA

- Combine FASTA and QUAL into FASTQ

GENERIC FASTQ MANIPULATION

- Filter FASTQ reads by quality score and length
- FASTQ Trimmer by column
- FASTQ Quality Trimmer by sliding window
- FASTQ Masker by quality score

FastQC:Read QC (version 0.10.1)

Short read data from your current history:

1: unknow.fastq

Formats: fastqsanger,fastq,bam,sam.

Title for the output file - to remind you what the job was for:

FastQC

Letters and numbers only please - other characters will be removed

Contaminant list:

Selection is Optional

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA. Format: tabular.

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***katerina.michalickova@usit.uio.no***. Please select the project (account) where you will be running your current job.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

staff

Number of tasks	**Number of tasks per node**	**Walltime (job duration)**	**Memory per CPU**
Enter the number of tasks for the current job. *(Enter 1, if not sure.)*	Enter the number of tasks per node (max 16). *(Leave 1, if not sure.)*	Select how long the job shall be running :	Select how much memory you need for the job (the allowed format is digit+Gb/Mb, e.g. 16GB or 3000Mb):
1	1	DD: 00 HH: 2 MM: 00 SC: 00	4Gb

Execute

Purpose

FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your

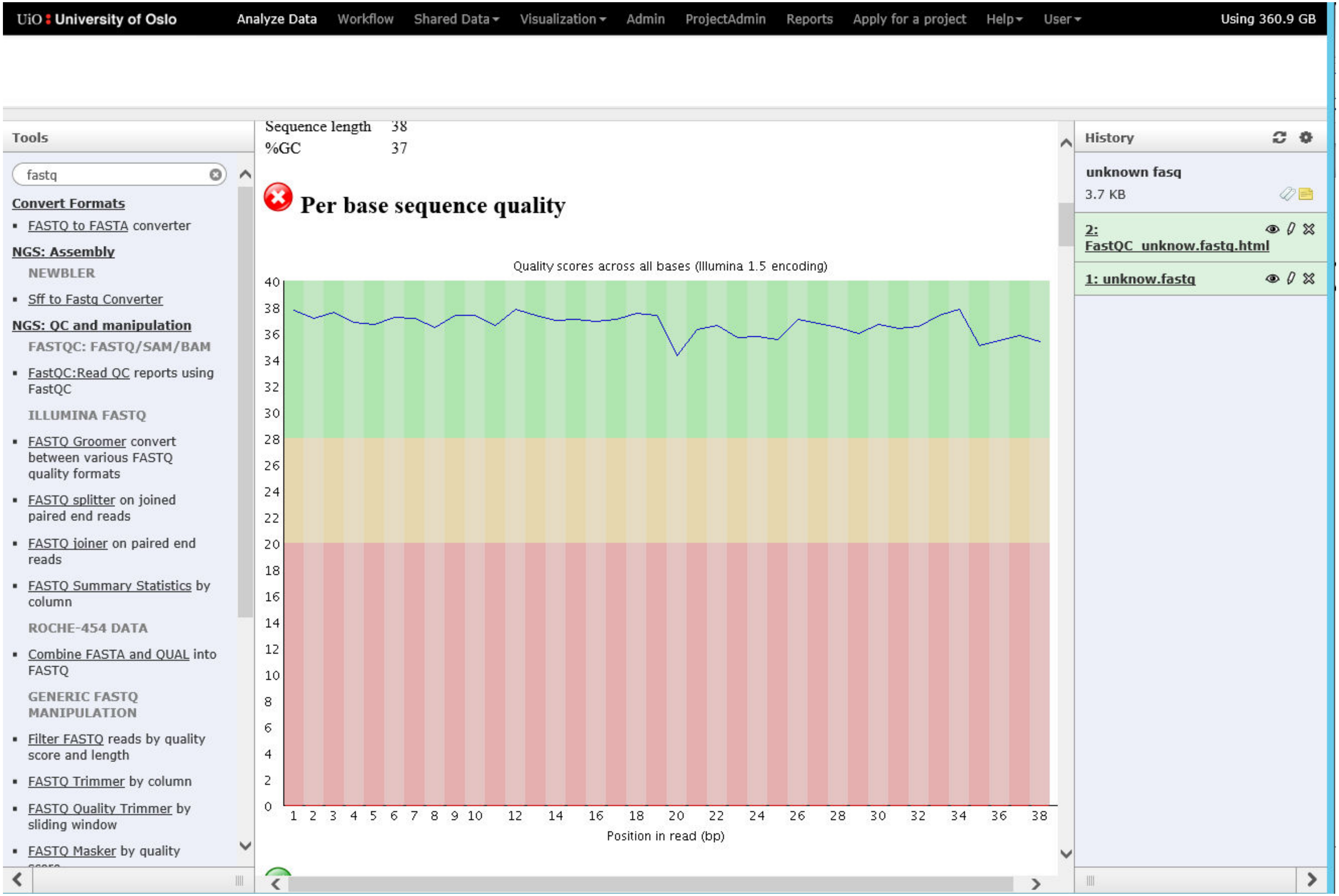
History

unknown fasq

0 bytes

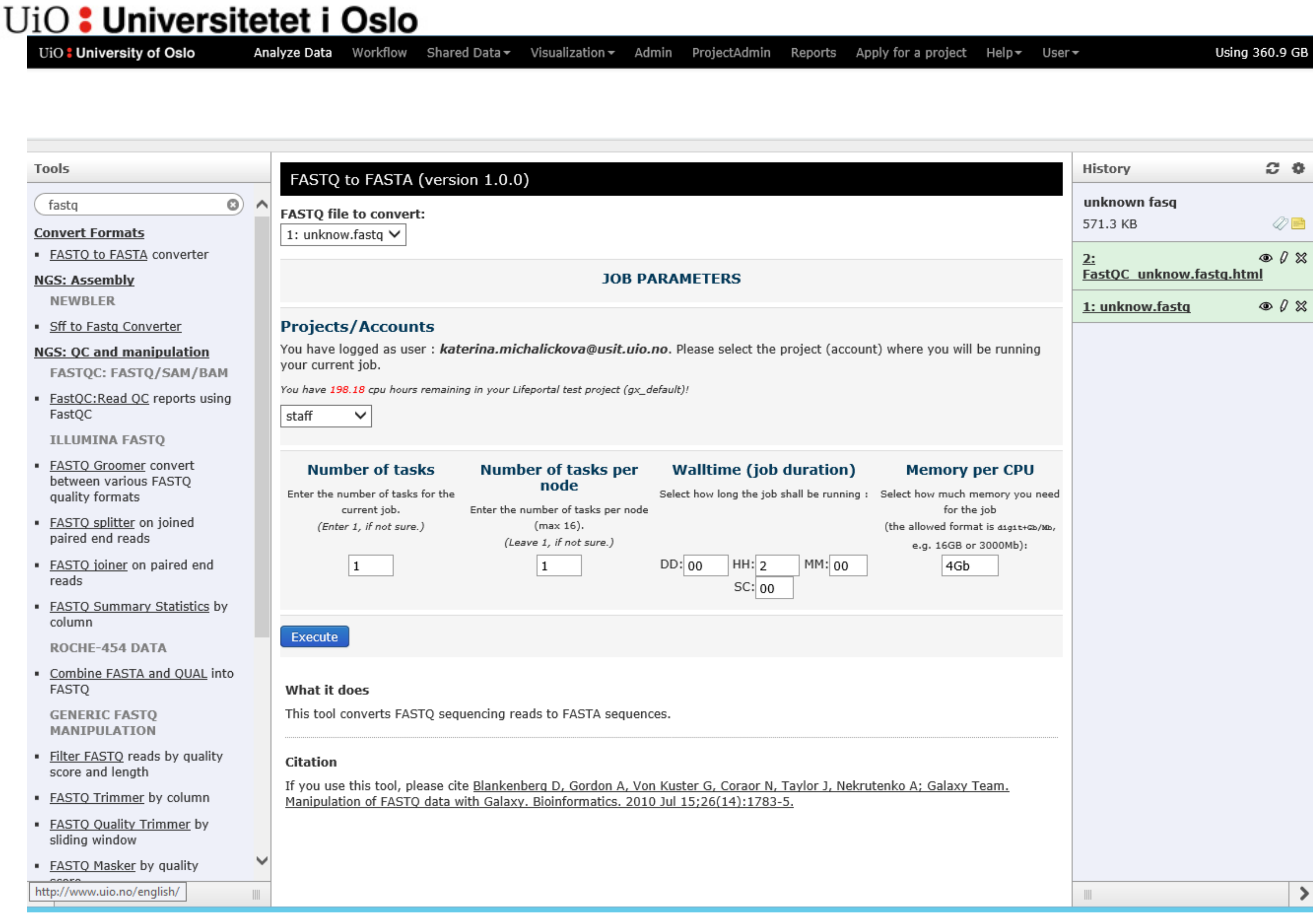
1: unknow.fastq

UiO : Universitetet i Oslo



Sequence length	38
%GC	37

Per base sequence quality



FASTQ to FASTA (version 1.0.0)

FASTQ file to convert:
1: unknow.fastq

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***katerina.michalickova@usit.uio.no***. Please select the project (account) where you will be running your current job.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

staff

Number of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)
1

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)
1

Walltime (job duration)
Select how long the job shall be running :
DD: 00 HH: 2 MM: 00
SC: 00

Memory per CPU
Select how much memory you need for the job
(the allowed format is `digit+Gb/Mb`, e.g. 16GB or 3000Mb):
4Gb

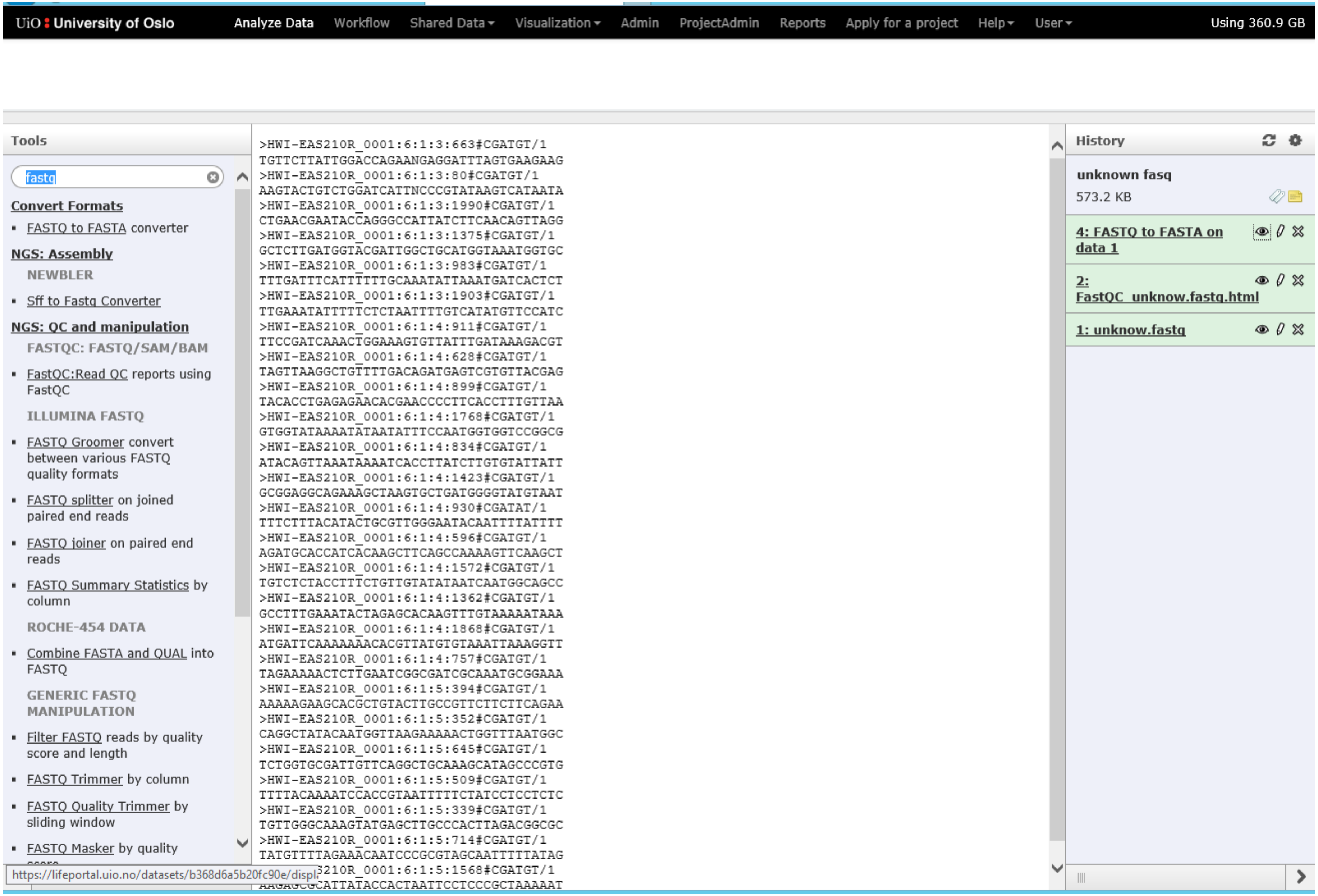
Execute

What it does

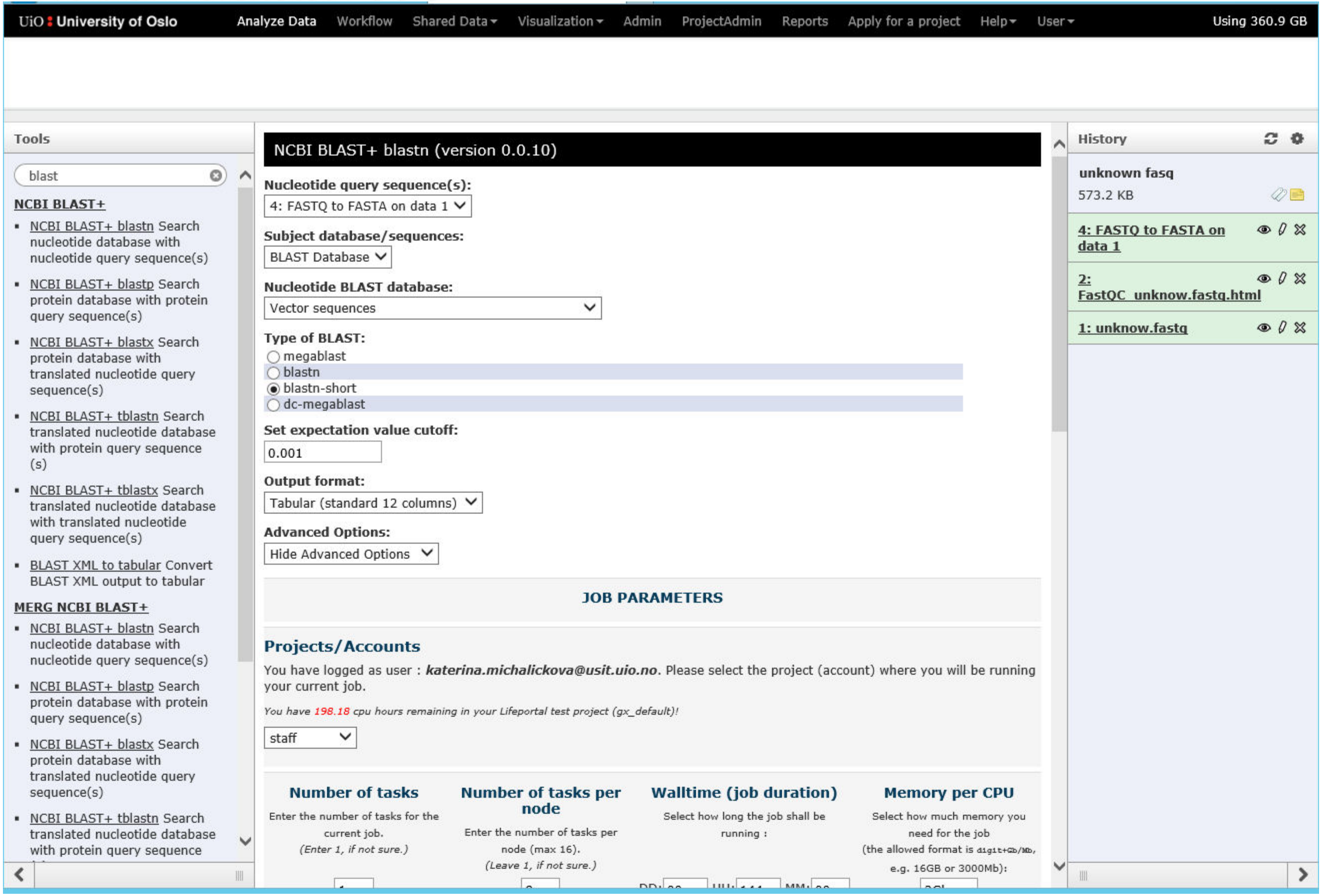
This tool converts FASTQ sequencing reads to FASTA sequences.

Citation

If you use this tool, please cite Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A; Galaxy Team. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010 Jul 15;26(14):1783-5.



```
>HWI-EAS210R_0001:6:1:3:663#CGATGT/1
TGTTCTTATTGGACCAGAANGAGGATTTAGTGAAGAAG
>HWI-EAS210R_0001:6:1:3:80#CGATGT/1
AAGTACTGTCTGGATCATTNCCCGTATAAGTCATAATA
>HWI-EAS210R_0001:6:1:3:1990#CGATGT/1
CTGAACGAATACCAGGGCCATTATCTTCAACAGTTAGG
>HWI-EAS210R_0001:6:1:3:1375#CGATGT/1
GCTCTTGATGGTACGATTGGCTGCATGGTAAATGGTGC
>HWI-EAS210R_0001:6:1:3:983#CGATGT/1
TTTGATTTCATTTTTTGCAAATATTAAATGATCACTCT
>HWI-EAS210R_0001:6:1:3:1903#CGATGT/1
TTGAAATATTTTTTCTCTAATTTTGTCATATGTTCCATC
>HWI-EAS210R_0001:6:1:4:911#CGATGT/1
TTCCGATCAAACTGGAAAGTGTTATTTGATAAAGACGT
>HWI-EAS210R_0001:6:1:4:628#CGATGT/1
TAGTTAAGGCTGTTTTGACAGATGAGTCGTGTTACGAG
>HWI-EAS210R_0001:6:1:4:899#CGATGT/1
TACACCTGAGAGAACACGAACCCCTTCACCTTTGTTAA
>HWI-EAS210R_0001:6:1:4:1768#CGATGT/1
GTGGTATAAAATATAATATTTCCAATGGTGGTCCGGCG
>HWI-EAS210R_0001:6:1:4:834#CGATGT/1
ATACAGTTAAATAAAATCACCTTATCTTGTGTATTATT
>HWI-EAS210R_0001:6:1:4:1423#CGATGT/1
GCGGAGGCAGAAAGCTAAGTGCTGATGGGGTATGTAAT
>HWI-EAS210R_0001:6:1:4:930#CGATAT/1
TTTCTTTACATACTGCGTTGGGAATACAATTTTATTTT
>HWI-EAS210R_0001:6:1:4:596#CGATGT/1
AGATGCACCATCACAAGCTTCAGCCAAAAGTTCAAGCT
>HWI-EAS210R_0001:6:1:4:1572#CGATGT/1
TGTCTCTACCTTTTCTGTTGTATATAATCAATGGCAGCC
>HWI-EAS210R_0001:6:1:4:1362#CGATGT/1
GCCTTTGAAATACTAGAGCACAAGTTTGTAAAAATAAA
>HWI-EAS210R_0001:6:1:4:1868#CGATGT/1
ATGATTCAAAAAAACACGTTATGTGTAAATTAAAGGTT
>HWI-EAS210R_0001:6:1:4:757#CGATGT/1
TAGAAAAACTCTTGAATCGGCGATCGCAAATGCGGAAA
>HWI-EAS210R_0001:6:1:5:394#CGATGT/1
AAAAAGAAGCACGCTGTACTTGCCGTTCTTCTTCAGAA
>HWI-EAS210R_0001:6:1:5:352#CGATGT/1
CAGGCTATACAATGGTTAAGAAAAACTGGTTTAATGGC
>HWI-EAS210R_0001:6:1:5:645#CGATGT/1
TCTGGTGCGATTGTTCAGGCTGCAAAGCATAGCCCGTG
>HWI-EAS210R_0001:6:1:5:509#CGATGT/1
TTTTACAAAATCCACCGTAATTTTTCTATCCTCCTCTC
>HWI-EAS210R_0001:6:1:5:339#CGATGT/1
TGTTGGGCAAAGTATGAGCTTGCCCACTTAGACGGCGC
>HWI-EAS210R_0001:6:1:5:714#CGATGT/1
TATGTTTTAGAAACAATCCCGCGTAGCAATTTTTATAG
S210R_0001:6:1:5:1568#CGATGT/1
AAGAGCGCATTATACCACTAATTCCTCCCGCTAAAAAT
```



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

MERG 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

NCBI BLAST+ blastn (version 0.0.10)

Nucleotide query sequence(s): 4: FASTQ to FASTA on data 1

Subject database/sequences: BLAST Database

Nucleotide BLAST database: Vector sequences

Type of BLAST:
○ megablast
○ blastn
● blastn-short
○ dc-megablast

Set expectation value cutoff: 0.001

Output format: Tabular (standard 12 columns)

Advanced Options: Hide Advanced Options

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***katerina.michalickova@usit.uio.no***. Please select the project (account) where you will be running your current job.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

staff

Number of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)

Walltime (job duration)
Select how long the job shall be running :

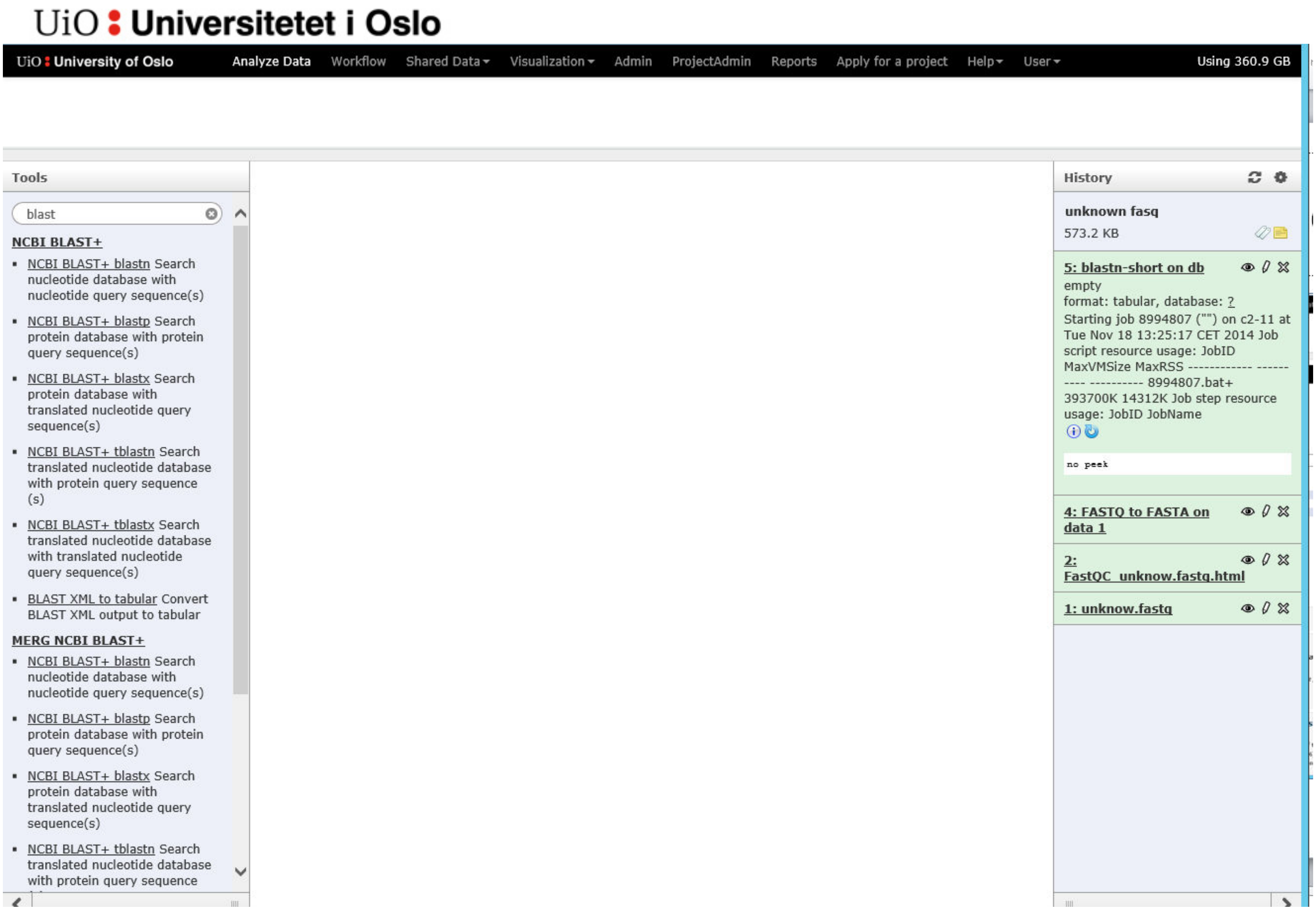
Memory per CPU
Select how much memory you need for the job
(the allowed format is digit+Gb/Mb, e.g. 16GB or 3000Mb):

unknown fasq
573.2 KB

4: FASTQ to FASTA on data 1

2: FastQC_unknow.fastq.html

1: unknow.fastq



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

MERG 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

unknown fasq

573.2 KB

5: blastn-short on db

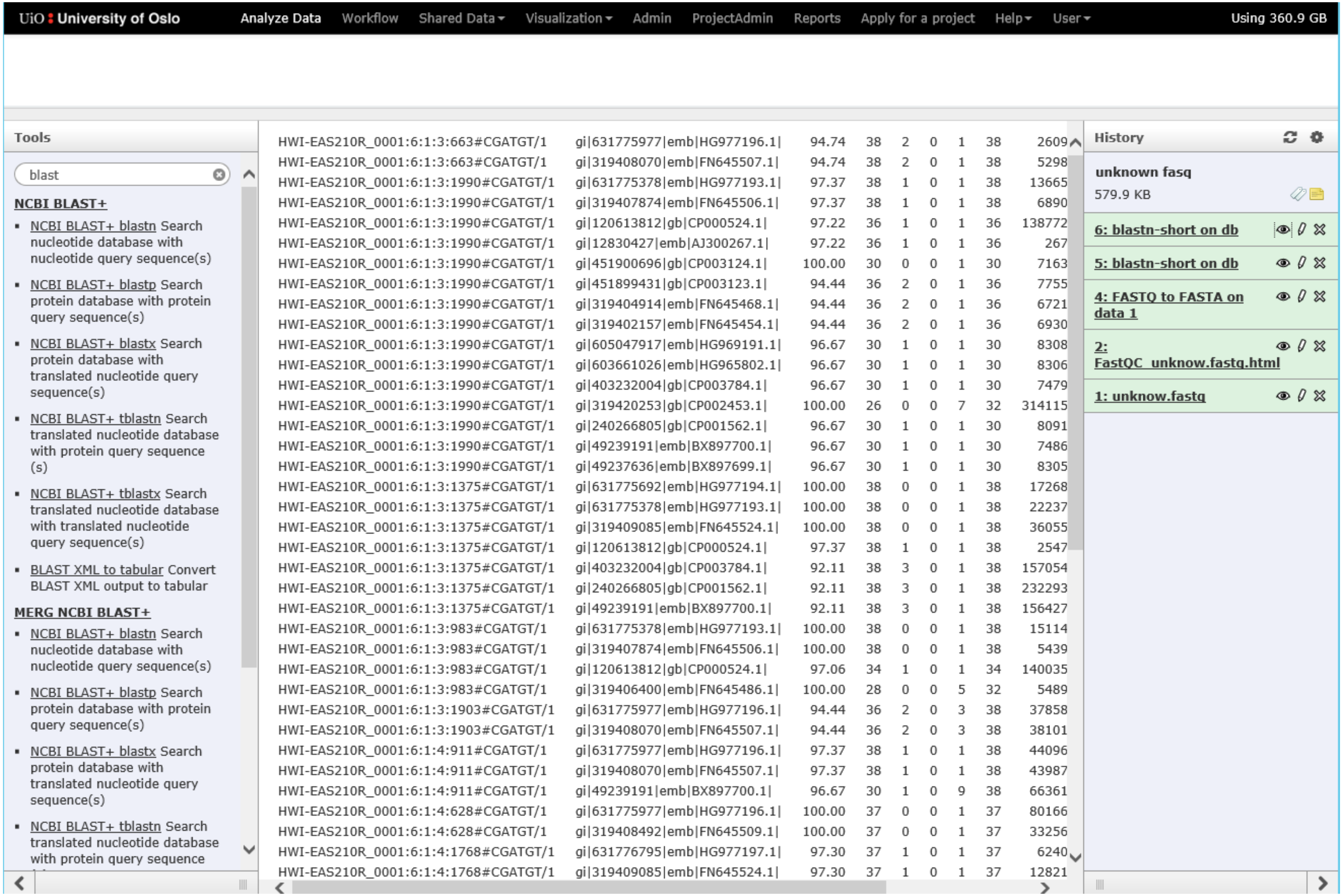
empty
format: tabular, database: ?
Starting job 8994807 ("") on c2-11 at Tue Nov 18 13:25:17 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994807.bat+ 393700K 14312K Job step resource usage: JobID JobName

```
no peek
```

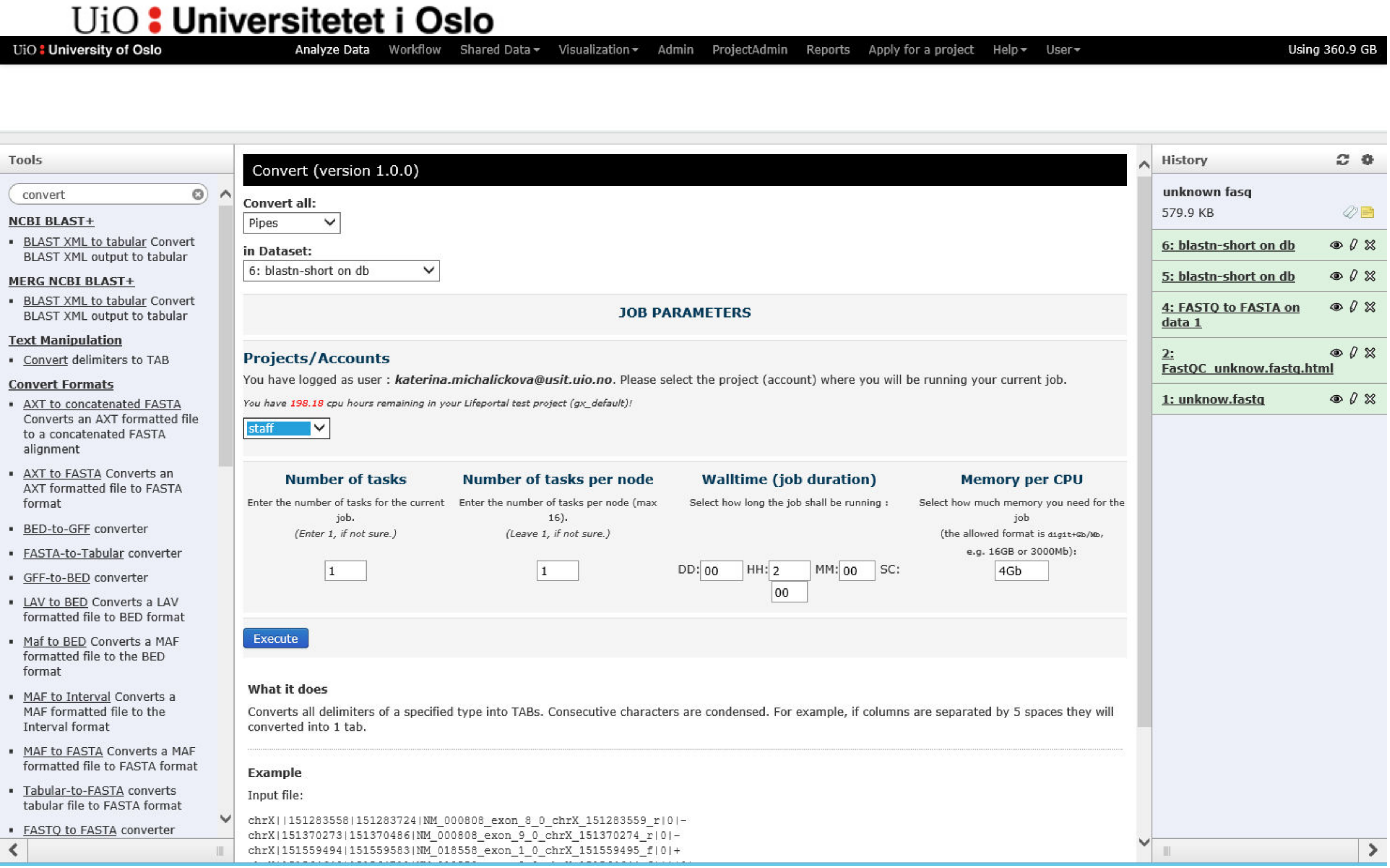
4: FASTQ to FASTA on data 1

2: FastQC_unknow.fastq.html

1: unknow.fastq



```
HWI-EAS210R_0001:6:1:3:663#CGATGT/1    gi|631775977|emb|HG977196.1|   94.74   38   2   0   1   38     2609
HWI-EAS210R_0001:6:1:3:663#CGATGT/1    gi|319408070|emb|FN645507.1|   94.74   38   2   0   1   38     5298
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|631775378|emb|HG977193.1|   97.37   38   1   0   1   38    13665
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|319407874|emb|FN645506.1|   97.37   38   1   0   1   38     6890
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|120613812|gb|CP000524.1|    97.22   36   1   0   1   36   138772
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|12830427|emb|AJ300267.1|    97.22   36   1   0   1   36      267
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|451900696|gb|CP003124.1|   100.00   30   0   0   1   30     7163
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|451899431|gb|CP003123.1|    94.44   36   2   0   1   36     7755
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|319404914|emb|FN645468.1|   94.44   36   2   0   1   36     6721
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|319402157|emb|FN645454.1|   94.44   36   2   0   1   36     6930
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|605047917|emb|HG969191.1|   96.67   30   1   0   1   30     8308
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|603661026|emb|HG965802.1|   96.67   30   1   0   1   30     8306
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|403232004|gb|CP003784.1|    96.67   30   1   0   1   30     7479
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|319420253|gb|CP002453.1|   100.00   26   0   0   7   32   314115
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|240266805|gb|CP001562.1|    96.67   30   1   0   1   30     8091
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|49239191|emb|BX897700.1|    96.67   30   1   0   1   30     7486
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi|49237636|emb|BX897699.1|    96.67   30   1   0   1   30     8305
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|631775692|emb|HG977194.1|  100.00   38   0   0   1   38    17268
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|631775378|emb|HG977193.1|  100.00   38   0   0   1   38    22237
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|319409085|emb|FN645524.1|  100.00   38   0   0   1   38    36055
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|120613812|gb|CP000524.1|    97.37   38   1   0   1   38     2547
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|403232004|gb|CP003784.1|    92.11   38   3   0   1   38   157054
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|240266805|gb|CP001562.1|    92.11   38   3   0   1   38   232293
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi|49239191|emb|BX897700.1|    92.11   38   3   0   1   38   156427
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi|631775378|emb|HG977193.1|  100.00   38   0   0   1   38    15114
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi|319407874|emb|FN645506.1|  100.00   38   0   0   1   38     5439
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi|120613812|gb|CP000524.1|    97.06   34   1   0   1   34   140035
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi|319406400|emb|FN645486.1|  100.00   28   0   0   5   32     5489
HWI-EAS210R_0001:6:1:3:1903#CGATGT/1   gi|631775977|emb|HG977196.1|   94.44   36   2   0   3   38    37858
HWI-EAS210R_0001:6:1:3:1903#CGATGT/1   gi|319408070|emb|FN645507.1|   94.44   36   2   0   3   38    38101
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi|631775977|emb|HG977196.1|   97.37   38   1   0   1   38    44096
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi|319408070|emb|FN645507.1|   97.37   38   1   0   1   38    43987
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi|49239191|emb|BX897700.1|    96.67   30   1   0   9   38    66361
HWI-EAS210R_0001:6:1:4:628#CGATGT/1    gi|631775977|emb|HG977196.1|  100.00   37   0   0   1   37    80166
HWI-EAS210R_0001:6:1:4:628#CGATGT/1    gi|319408492|emb|FN645509.1|  100.00   37   0   0   1   37    33256
HWI-EAS210R_0001:6:1:4:1768#CGATGT/1   gi|631776795|emb|HG977197.1|   97.30   37   1   0   1   37     6240
HWI-EAS210R_0001:6:1:4:1768#CGATGT/1   gi|319409085|emb|FN645524.1|   97.30   37   1   0   1   37    12821
```



NCBI BLAST+

- BLAST XML to tabular Convert BLAST XML output to tabular

MERG NCBI BLAST+

- BLAST XML to tabular Convert BLAST XML output to tabular

Text Manipulation

- Convert delimiters to TAB

Convert Formats

- AXT to concatenated FASTA Converts an AXT formatted file to a concatenated FASTA alignment
- AXT to FASTA Converts an AXT formatted file to FASTA format
- BED-to-GFF converter
- FASTA-to-Tabular converter
- GFF-to-BED converter
- LAV to BED Converts a LAV formatted file to BED format
- Maf to BED Converts a MAF formatted file to the BED format
- MAF to Interval Converts a MAF formatted file to the Interval format
- MAF to FASTA Converts a MAF formatted file to FASTA format
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTQ to FASTA converter

Convert (version 1.0.0)

Convert all: Pipes

in Dataset: 6: blastn-short on db

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***katerina.michalickova@usit.uio.no***. Please select the project (account) where you will be running your current job.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

staff

Number of tasks

Enter the number of tasks for the current job.

(Enter 1, if not sure.)

1

Number of tasks per node

Enter the number of tasks per node (max 16).

(Leave 1, if not sure.)

1

Walltime (job duration)

Select how long the job shall be running :

DD: 00 HH: 2 MM: 00 SC: 00

Memory per CPU

Select how much memory you need for the job

(the allowed format is `digit+Gb/Mb`, e.g. 16GB or 3000Mb):

4Gb

Execute

What it does

Converts all delimiters of a specified type into TABs. Consecutive characters are condensed. For example, if columns are separated by 5 spaces they will converted into 1 tab.

Example

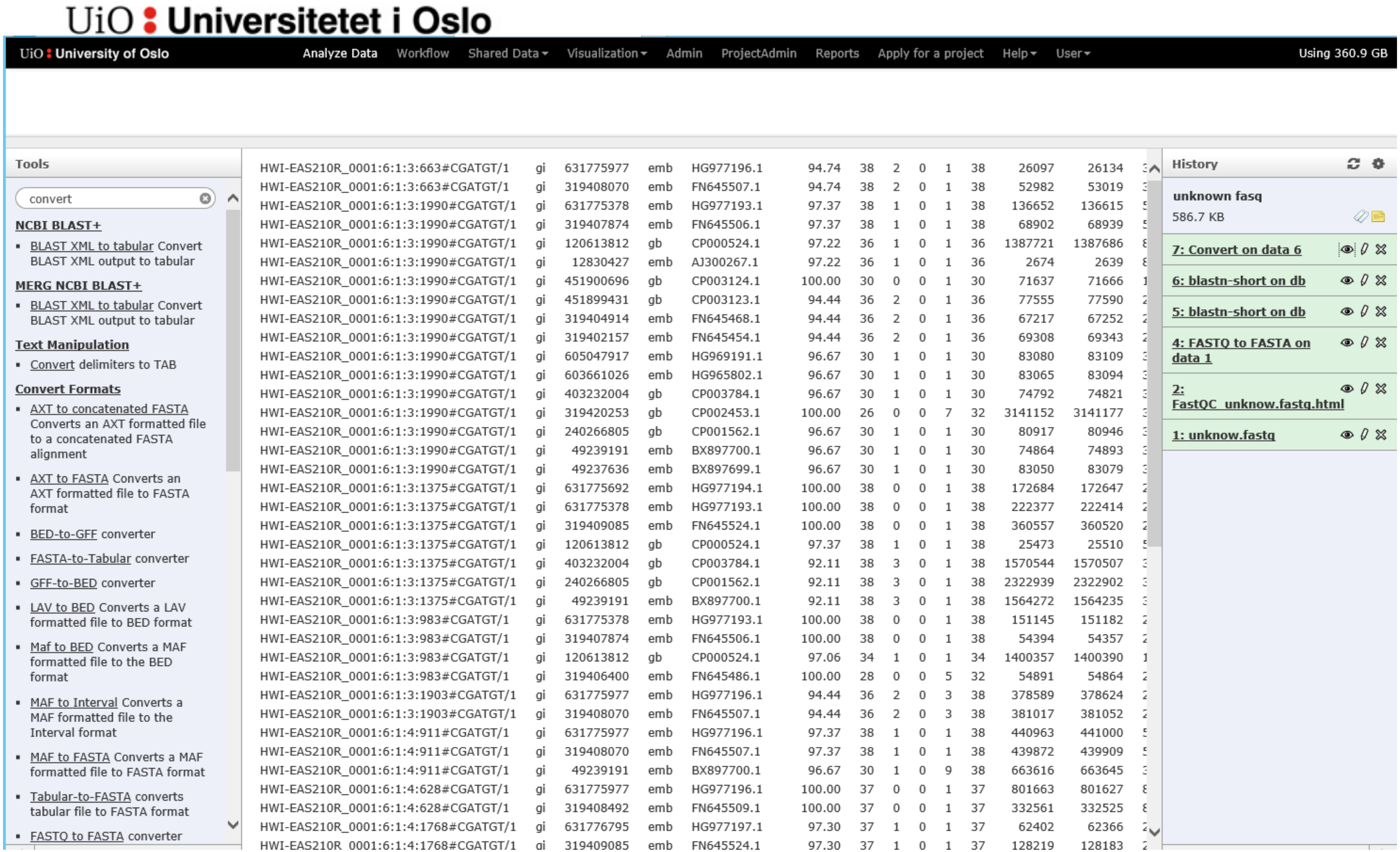
Input file:

```
chrX||151283558|151283724|NM_000808_exon_8_0_chrX_151283559_r|0|-
chrX|151370273|151370486|NM_000808_exon_9_0_chrX_151370274_r|0|-
chrX|151559494|151559583|NM_018558_exon_1_0_chrX_151559495_f|0|+
```

unknown fasq

579.9 KB

- 6: blastn-short on db
- 5: blastn-short on db
- 4: FASTQ to FASTA on data 1
- 2: FastQC_unknow.fastq.html
- 1: unknow.fastq



NCBI BLAST+

- BLAST XML to tabular Convert BLAST XML output to tabular

MERG NCBI BLAST+

- BLAST XML to tabular Convert BLAST XML output to tabular

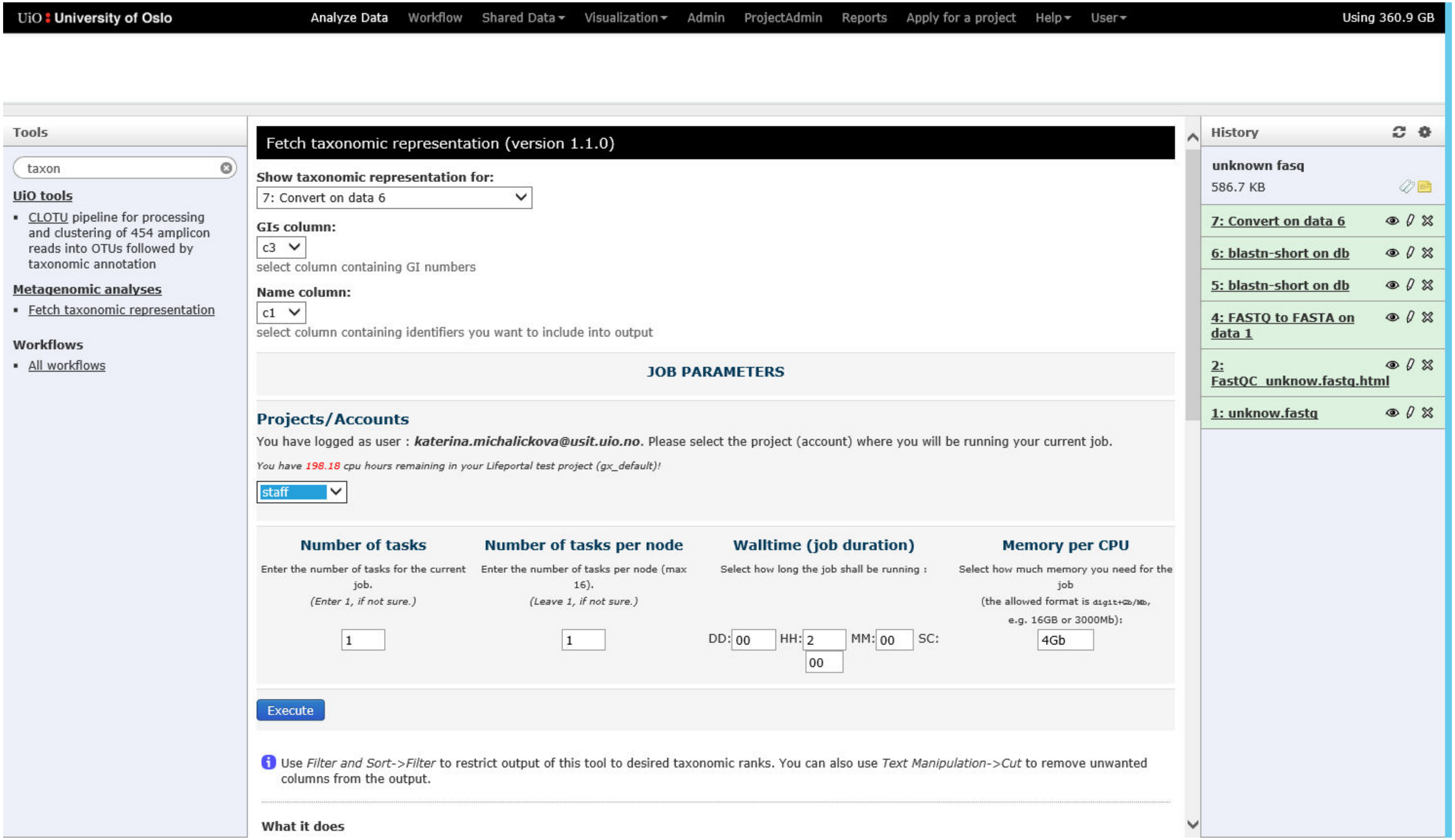
Text Manipulation

- Convert delimiters to TAB

Convert Formats

- AXT to concatenated FASTA Converts an AXT formatted file to a concatenated FASTA alignment
- AXT to FASTA Converts an AXT formatted file to FASTA format
- BED-to-GFF converter
- FASTA-to-Tabular converter
- GFF-to-BED converter
- LAV to BED Converts a LAV formatted file to BED format
- Maf to BED Converts a MAF formatted file to the BED format
- MAF to Interval Converts a MAF formatted file to the Interval format
- MAF to FASTA Converts a MAF formatted file to FASTA format
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTQ to FASTA converter

```
HWI-EAS210R_0001:6:1:3:663#CGATGT/1    gi  631775977  emb  HG977196.1   94.74  38  2  0  1  38    26097    26134
HWI-EAS210R_0001:6:1:3:663#CGATGT/1    gi  319408070  emb  FN645507.1   94.74  38  2  0  1  38    52982    53019
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  631775378  emb  HG977193.1   97.37  38  1  0  1  38   136652   136615
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  319407874  emb  FN645506.1   97.37  38  1  0  1  38    68902    68939
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  120613812  gb   CP000524.1   97.22  36  1  0  1  36  1387721  1387686
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi   12830427  emb  AJ300267.1   97.22  36  1  0  1  36     2674     2639
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  451900696  gb   CP003124.1  100.00  30  0  0  1  30    71637    71666
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  451899431  gb   CP003123.1   94.44  36  2  0  1  36    77555    77590
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  319404914  emb  FN645468.1   94.44  36  2  0  1  36    67217    67252
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  319402157  emb  FN645454.1   94.44  36  2  0  1  36    69308    69343
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  605047917  emb  HG969191.1   96.67  30  1  0  1  30    83080    83109
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  603661026  emb  HG965802.1   96.67  30  1  0  1  30    83065    83094
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  403232004  gb   CP003784.1   96.67  30  1  0  1  30    74792    74821
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  319420253  gb   CP002453.1  100.00  26  0  0  7  32  3141152  3141177
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi  240266805  gb   CP001562.1   96.67  30  1  0  1  30    80917    80946
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi   49239191  emb  BX897700.1   96.67  30  1  0  1  30    74864    74893
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1   gi   49237636  emb  BX897699.1   96.67  30  1  0  1  30    83050    83079
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  631775692  emb  HG977194.1  100.00  38  0  0  1  38   172684   172647
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  631775378  emb  HG977193.1  100.00  38  0  0  1  38   222377   222414
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  319409085  emb  FN645524.1  100.00  38  0  0  1  38   360557   360520
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  120613812  gb   CP000524.1   97.37  38  1  0  1  38    25473    25510
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  403232004  gb   CP003784.1   92.11  38  3  0  1  38  1570544  1570507
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi  240266805  gb   CP001562.1   92.11  38  3  0  1  38  2322939  2322902
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1   gi   49239191  emb  BX897700.1   92.11  38  3  0  1  38  1564272  1564235
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi  631775378  emb  HG977193.1  100.00  38  0  0  1  38   151145   151182
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi  319407874  emb  FN645506.1  100.00  38  0  0  1  38    54394    54357
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi  120613812  gb   CP000524.1   97.06  34  1  0  1  34  1400357  1400390
HWI-EAS210R_0001:6:1:3:983#CGATGT/1    gi  319406400  emb  FN645486.1  100.00  28  0  0  5  32    54891    54864
HWI-EAS210R_0001:6:1:3:1903#CGATGT/1   gi  631775977  emb  HG977196.1   94.44  36  2  0  3  38   378589   378624
HWI-EAS210R_0001:6:1:3:1903#CGATGT/1   gi  319408070  emb  FN645507.1   94.44  36  2  0  3  38   381017   381052
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi  631775977  emb  HG977196.1   97.37  38  1  0  1  38   440963   441000
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi  319408070  emb  FN645507.1   97.37  38  1  0  1  38   439872   439909
HWI-EAS210R_0001:6:1:4:911#CGATGT/1    gi   49239191  emb  BX897700.1   96.67  30  1  0  9  38   663616   663645
HWI-EAS210R_0001:6:1:4:628#CGATGT/1    gi  631775977  emb  HG977196.1  100.00  37  0  0  1  37   801663   801627
HWI-EAS210R_0001:6:1:4:628#CGATGT/1    gi  319408492  emb  FN645509.1  100.00  37  0  0  1  37   332561   332525
HWI-EAS210R_0001:6:1:4:1768#CGATGT/1   gi  631776795  emb  HG977197.1   97.30  37  1  0  1  37    62402    62366
HWI-EAS210R_0001:6:1:4:1768#CGATGT/1   gi  319409085  emb  FN645524.1   97.30  37  1  0  1  37   128219   128183
```



Fetch taxonomic representation (version 1.1.0)

Show taxonomic representation for: 7: Convert on data 6

GIs column: c3

select column containing GI numbers

Name column: c1

select column containing identifiers you want to include into output

JOB PARAMETERS

Projects/Accounts

You have logged as user : ***katerina.michalickova@usit.uio.no***. Please select the project (account) where you will be running your current job.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

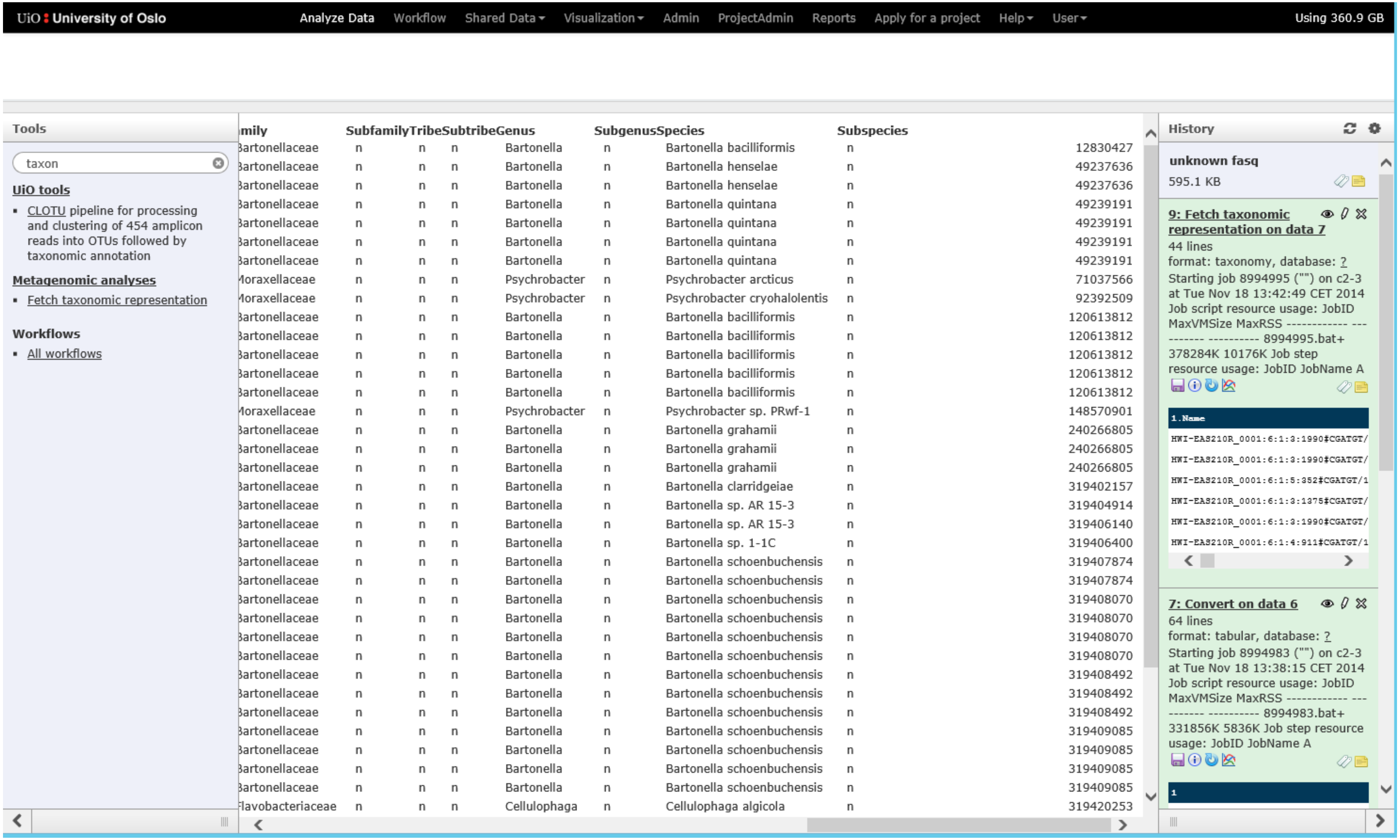
staff

Number of tasks	Number of tasks per node	Walltime (job duration)	Memory per CPU
Enter the number of tasks for the current job. *(Enter 1, if not sure.)*	Enter the number of tasks per node (max 16). *(Leave 1, if not sure.)*	Select how long the job shall be running :	Select how much memory you need for the job (the allowed format is `digit+Gb/Mb`, e.g. 16GB or 3000Mb):
1	1	DD: 00 HH: 2 MM: 00 SC: 00	4Gb

Execute

Use *Filter and Sort->Filter* to restrict output of this tool to desired taxonomic ranks. You can also use *Text Manipulation->Cut* to remove unwanted columns from the output.

What it does



UiO tools

- CLOTU pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation

Metagenomic analyses

- Fetch taxonomic representation

Workflows

- All workflows

mily	Subfamily	Tribe	Subtribe	Genus	Subgenus	Species	Subspecies	
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	12830427
artonellaceae	n	n	n	Bartonella	n	Bartonella henselae	n	49237636
artonellaceae	n	n	n	Bartonella	n	Bartonella henselae	n	49237636
artonellaceae	n	n	n	Bartonella	n	Bartonella quintana	n	49239191
artonellaceae	n	n	n	Bartonella	n	Bartonella quintana	n	49239191
artonellaceae	n	n	n	Bartonella	n	Bartonella quintana	n	49239191
artonellaceae	n	n	n	Bartonella	n	Bartonella quintana	n	49239191
oraxellaceae	n	n	n	Psychrobacter	n	Psychrobacter arcticus	n	71037566
oraxellaceae	n	n	n	Psychrobacter	n	Psychrobacter cryohalolentis	n	92392509
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	120613812
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	120613812
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	120613812
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	120613812
artonellaceae	n	n	n	Bartonella	n	Bartonella bacilliformis	n	120613812
oraxellaceae	n	n	n	Psychrobacter	n	Psychrobacter sp. PRwf-1	n	148570901
artonellaceae	n	n	n	Bartonella	n	Bartonella grahamii	n	240266805
artonellaceae	n	n	n	Bartonella	n	Bartonella grahamii	n	240266805
artonellaceae	n	n	n	Bartonella	n	Bartonella grahamii	n	240266805
artonellaceae	n	n	n	Bartonella	n	Bartonella clarridgeiae	n	319402157
artonellaceae	n	n	n	Bartonella	n	Bartonella sp. AR 15-3	n	319404914
artonellaceae	n	n	n	Bartonella	n	Bartonella sp. AR 15-3	n	319406140
artonellaceae	n	n	n	Bartonella	n	Bartonella sp. 1-1C	n	319406400
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319407874
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319407874
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408070
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408070
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408070
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408070
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408492
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408492
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319408492
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319409085
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319409085
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319409085
artonellaceae	n	n	n	Bartonella	n	Bartonella schoenbuchensis	n	319409085
lavobacteriaceae	n	n	n	Cellulophaga	n	Cellulophaga algicola	n	319420253

unknown fasq

595.1 KB

9: Fetch taxonomic representation on data 7

44 lines

format: taxonomy, database: ?

Starting job 8994995 ("") on c2-3 at Tue Nov 18 13:42:49 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994995.bat+ 378284K 10176K Job step resource usage: JobID JobName A

```
1.Name
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:5:352#CGATGT/1
HWI-EAS210R_0001:6:1:3:1375#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:4:911#CGATGT/1
```

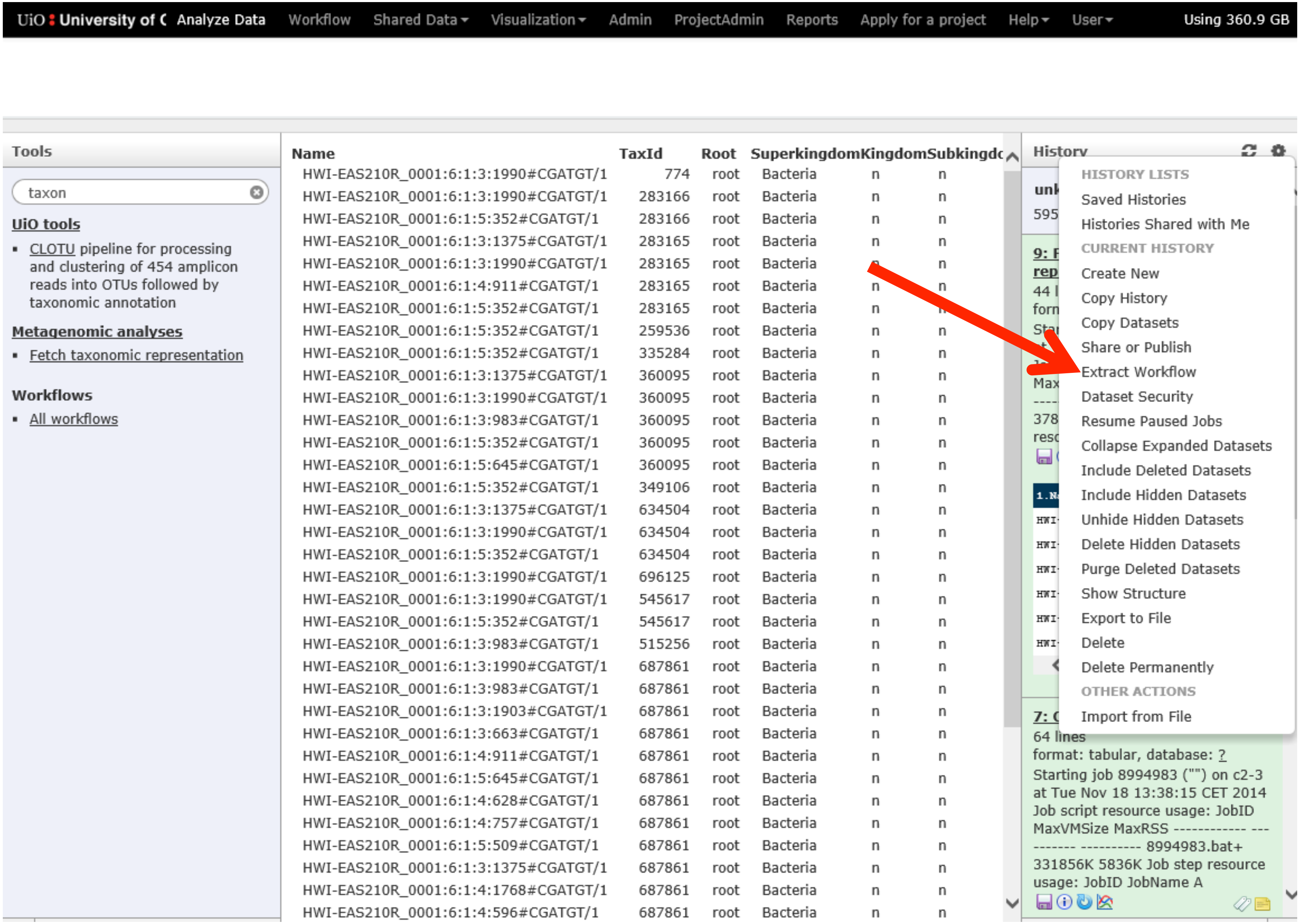
7: Convert on data 6

64 lines

format: tabular, database: ?

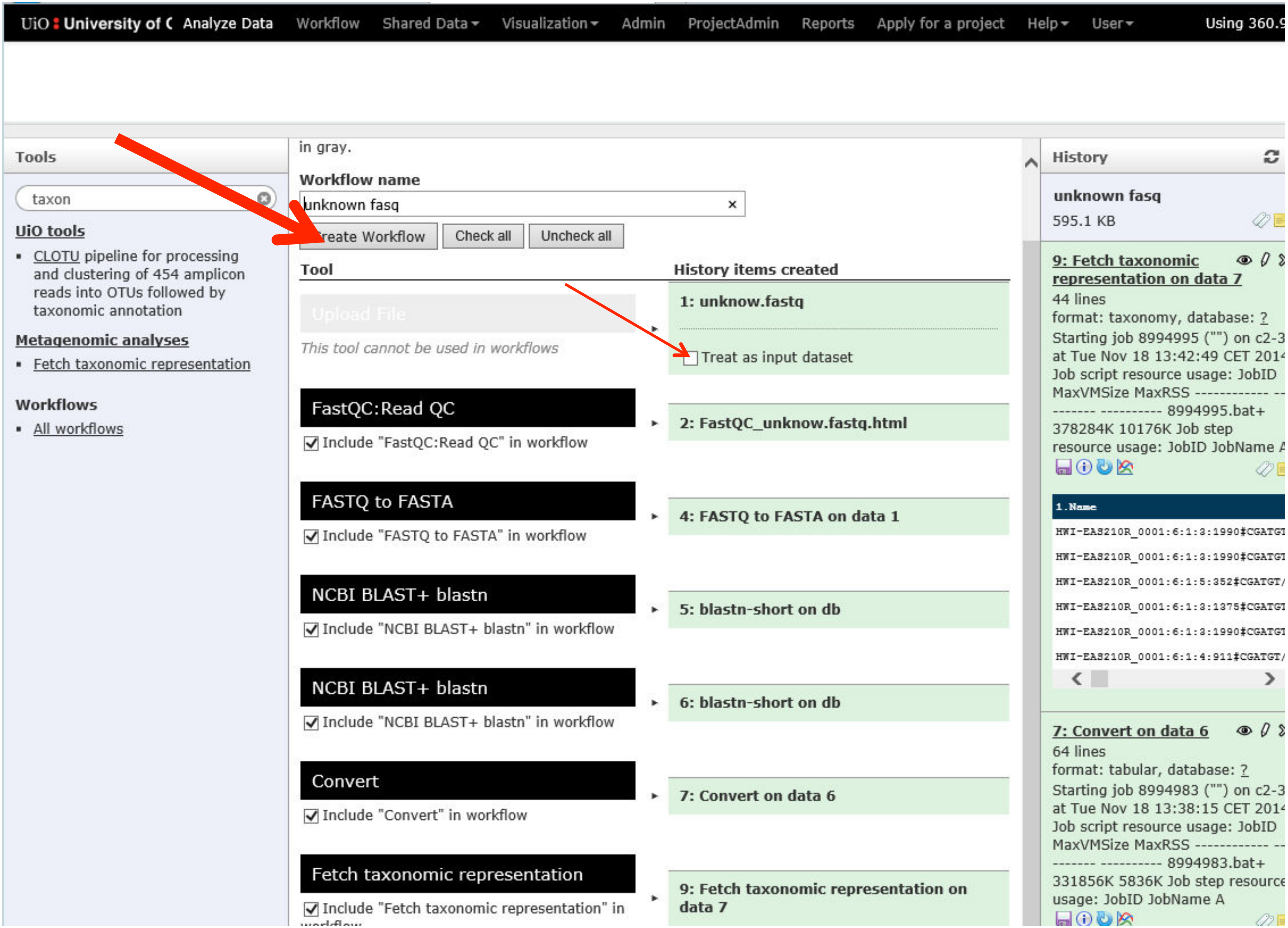
Starting job 8994983 ("") on c2-3 at Tue Nov 18 13:38:15 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994983.bat+ 331856K 5836K Job step resource usage: JobID JobName A

```
1
```



Name	TaxId	Root	Superkingdom	Kingdom	Subkingdc
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	774	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	283166	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	283166	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1	283165	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	283165	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:911#CGATGT/1	283165	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	283165	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	259536	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	335284	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1	360095	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	360095	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:983#CGATGT/1	360095	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	360095	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:645#CGATGT/1	360095	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	349106	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1	634504	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	634504	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	634504	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	696125	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	545617	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:352#CGATGT/1	545617	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:983#CGATGT/1	515256	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1990#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:983#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1903#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:663#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:911#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:645#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:628#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:757#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:5:509#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:3:1375#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:1768#CGATGT/1	687861	root	Bacteria	n	n
HWI-EAS210R_0001:6:1:4:596#CGATGT/1	687861	root	Bacteria	n	n

```
64 lines
format: tabular, database: ?
Starting job 8994983 ("") on c2-3
at Tue Nov 18 13:38:15 CET 2014
Job script resource usage: JobID
MaxVMSize MaxRSS ------------ ---
------- ---------- 8994983.bat+
331856K 5836K Job step resource
usage: JobID JobName A
```



UiO tools

- CLOTU pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation

Metagenomic analyses

- Fetch taxonomic representation

Workflows

- All workflows

in gray.

Workflow name

unknown fasq

Tool	History items created
Upload File *This tool cannot be used in workflows*	1: unknow.fastq ☐ Treat as input dataset
FastQC:Read QC ☑ Include "FastQC:Read QC" in workflow	2: FastQC_unknow.fastq.html
FASTQ to FASTA ☑ Include "FASTQ to FASTA" in workflow	4: FASTQ to FASTA on data 1
NCBI BLAST+ blastn ☑ Include "NCBI BLAST+ blastn" in workflow	5: blastn-short on db
NCBI BLAST+ blastn ☑ Include "NCBI BLAST+ blastn" in workflow	6: blastn-short on db
Convert ☑ Include "Convert" in workflow	7: Convert on data 6
Fetch taxonomic representation ☑ Include "Fetch taxonomic representation" in workflow	9: Fetch taxonomic representation on data 7

unknown fasq

595.1 KB

9: Fetch taxonomic representation on data 7

44 lines

format: taxonomy, database: ?

Starting job 8994995 ("") on c2-3 at Tue Nov 18 13:42:49 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ -------- ---------- 8994995.bat+ 378284K 10176K Job step resource usage: JobID JobName A

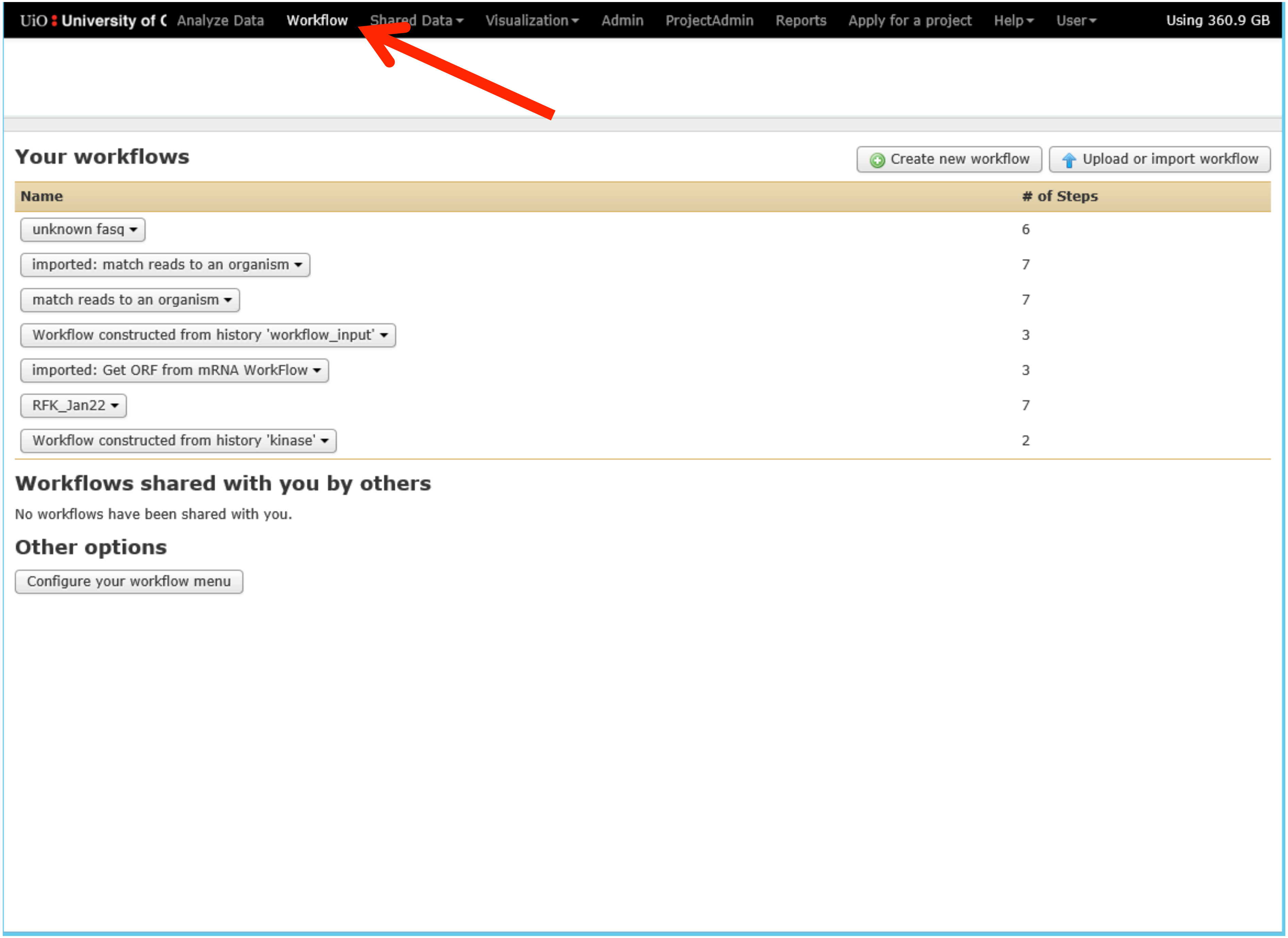
1.Name
HWI-EAS210R_0001:6:1:3:1990#CGATGT
HWI-EAS210R_0001:6:1:3:1990#CGATGT
HWI-EAS210R_0001:6:1:5:352#CGATGT/
HWI-EAS210R_0001:6:1:3:1375#CGATGT
HWI-EAS210R_0001:6:1:3:1990#CGATGT
HWI-EAS210R_0001:6:1:4:911#CGATGT/

7: Convert on data 6

64 lines

format: tabular, database: ?

Starting job 8994983 ("") on c2-3 at Tue Nov 18 13:38:15 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ -------- ---------- 8994983.bat+ 331856K 5836K Job step resource usage: JobID JobName A



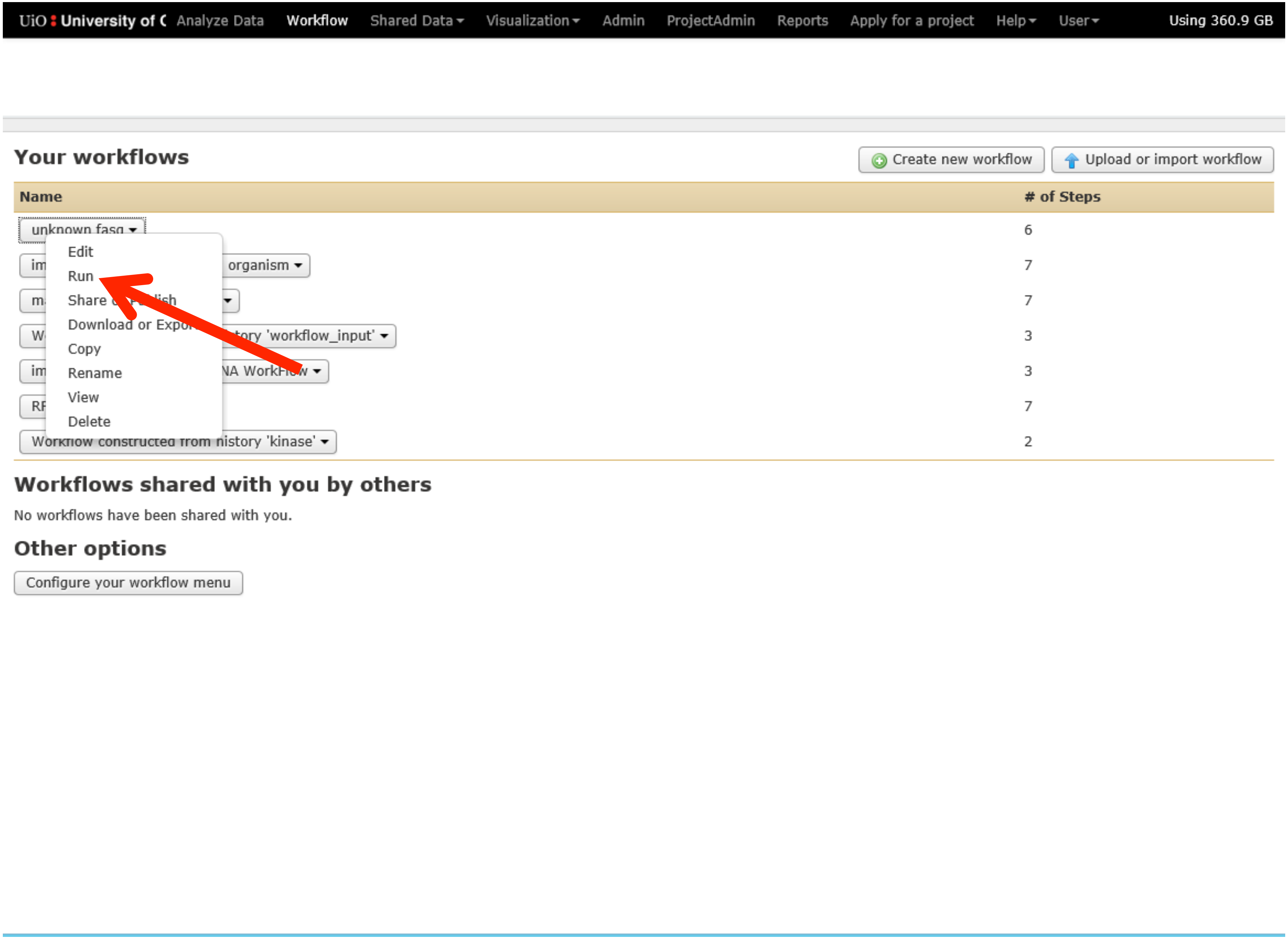
Your workflows

Name	# of Steps
unknown fasq	6
imported: match reads to an organism	7
match reads to an organism	7
Workflow constructed from history 'workflow_input'	3
imported: Get ORF from mRNA WorkFlow	3
RFK_Jan22	7
Workflow constructed from history 'kinase'	2

Workflows shared with you by others

No workflows have been shared with you.

Other options



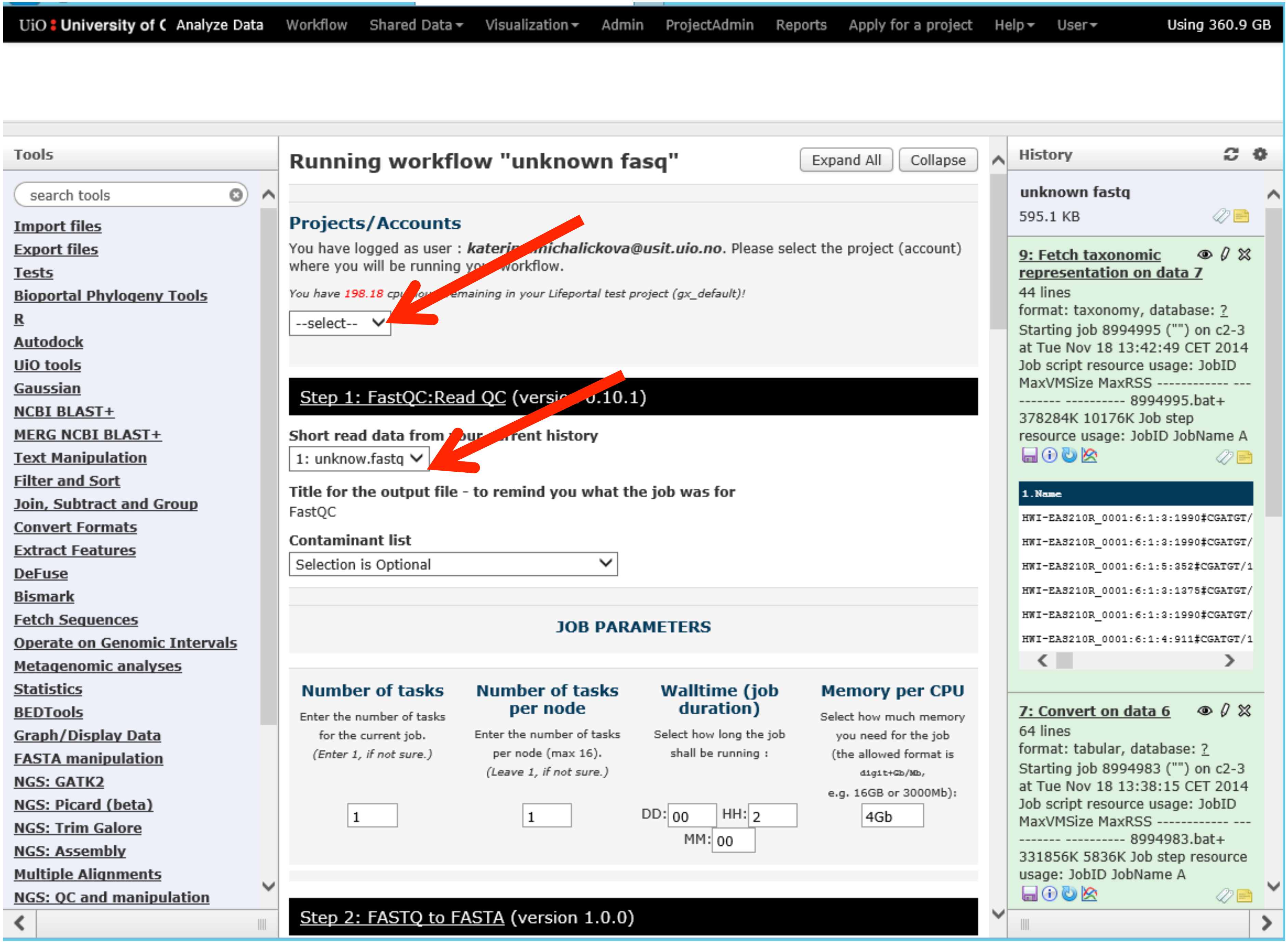
Your workflows

Name	# of Steps
unknown fasq	6
im… organism	7
m…	7
W… tory 'workflow_input'	3
im… NA WorkFlow	3
RF…	7
Workflow constructed from history 'kinase'	2

Workflows shared with you by others

No workflows have been shared with you.

Other options



Running workflow "unknown fasq"

Projects/Accounts

You have logged as user : *katerina.michalickova@usit.uio.no*. Please select the project (account) where you will be running your workflow.

You have 198.18 cpu hours remaining in your Lifeportal test project (gx_default)!

--select--

Step 1: FastQC:Read QC (version 0.10.1)

Short read data from your current history
1: unknow.fastq

Title for the output file - to remind you what the job was for
FastQC

Contaminant list
Selection is Optional

JOB PARAMETERS

Number of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)
1

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)
1

Walltime (job duration)
Select how long the job shall be running :
DD: 00 HH: 2 MM: 00

Memory per CPU
Select how much memory you need for the job (the allowed format is digit+Gb/Mb, e.g. 16GB or 3000Mb):
4Gb

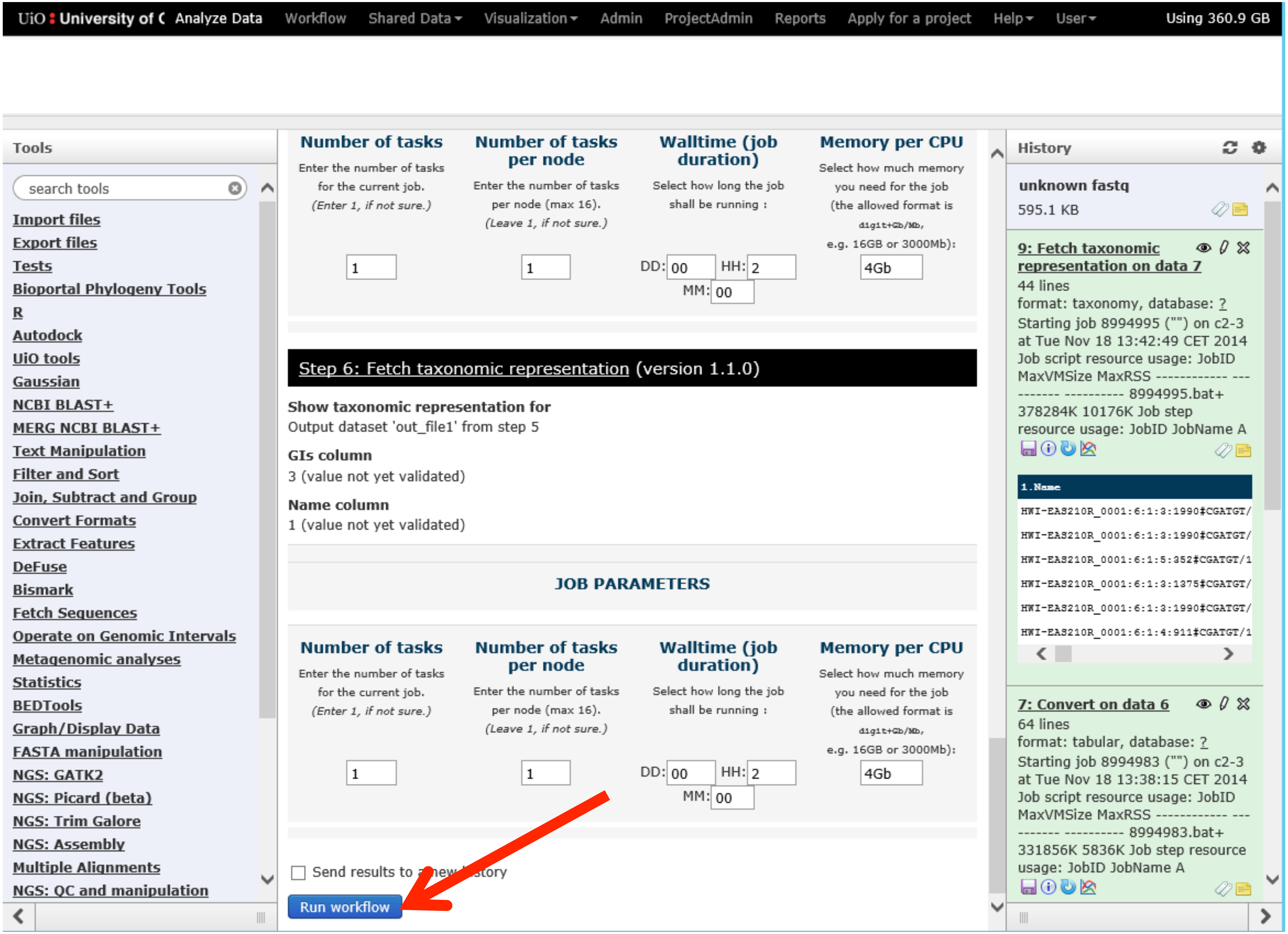
Step 2: FASTQ to FASTA (version 1.0.0)

unknown fastq
595.1 KB

9: Fetch taxonomic representation on data 7
44 lines
format: taxonomy, database: ?
Starting job 8994995 ("") on c2-3 at Tue Nov 18 13:42:49 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994995.bat+ 378284K 10176K Job step resource usage: JobID JobName A

```
1.Name
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:5:352#CGATGT/1
HWI-EAS210R_0001:6:1:3:1375#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:4:911#CGATGT/1
```

7: Convert on data 6
64 lines
format: tabular, database: ?
Starting job 8994983 ("") on c2-3 at Tue Nov 18 13:38:15 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994983.bat+ 331856K 5836K Job step resource usage: JobID JobName A



Number of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)
1

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)
1

Walltime (job duration)
Select how long the job shall be running :
DD: 00 HH: 2
MM: 00

Memory per CPU
Select how much memory you need for the job (the allowed format is `digit+Gb/Mb`, e.g. 16GB or 3000Mb):
4Gb

Step 6: Fetch taxonomic representation (version 1.1.0)

Show taxonomic representation for
Output dataset 'out_file1' from step 5

GIs column
3 (value not yet validated)

Name column
1 (value not yet validated)

JOB PARAMETERS

Number of tasks
Enter the number of tasks for the current job.
(Enter 1, if not sure.)
1

Number of tasks per node
Enter the number of tasks per node (max 16).
(Leave 1, if not sure.)
1

Walltime (job duration)
Select how long the job shall be running :
DD: 00 HH: 2
MM: 00

Memory per CPU
Select how much memory you need for the job (the allowed format is `digit+Gb/Mb`, e.g. 16GB or 3000Mb):
4Gb

☐ Send results to a new history

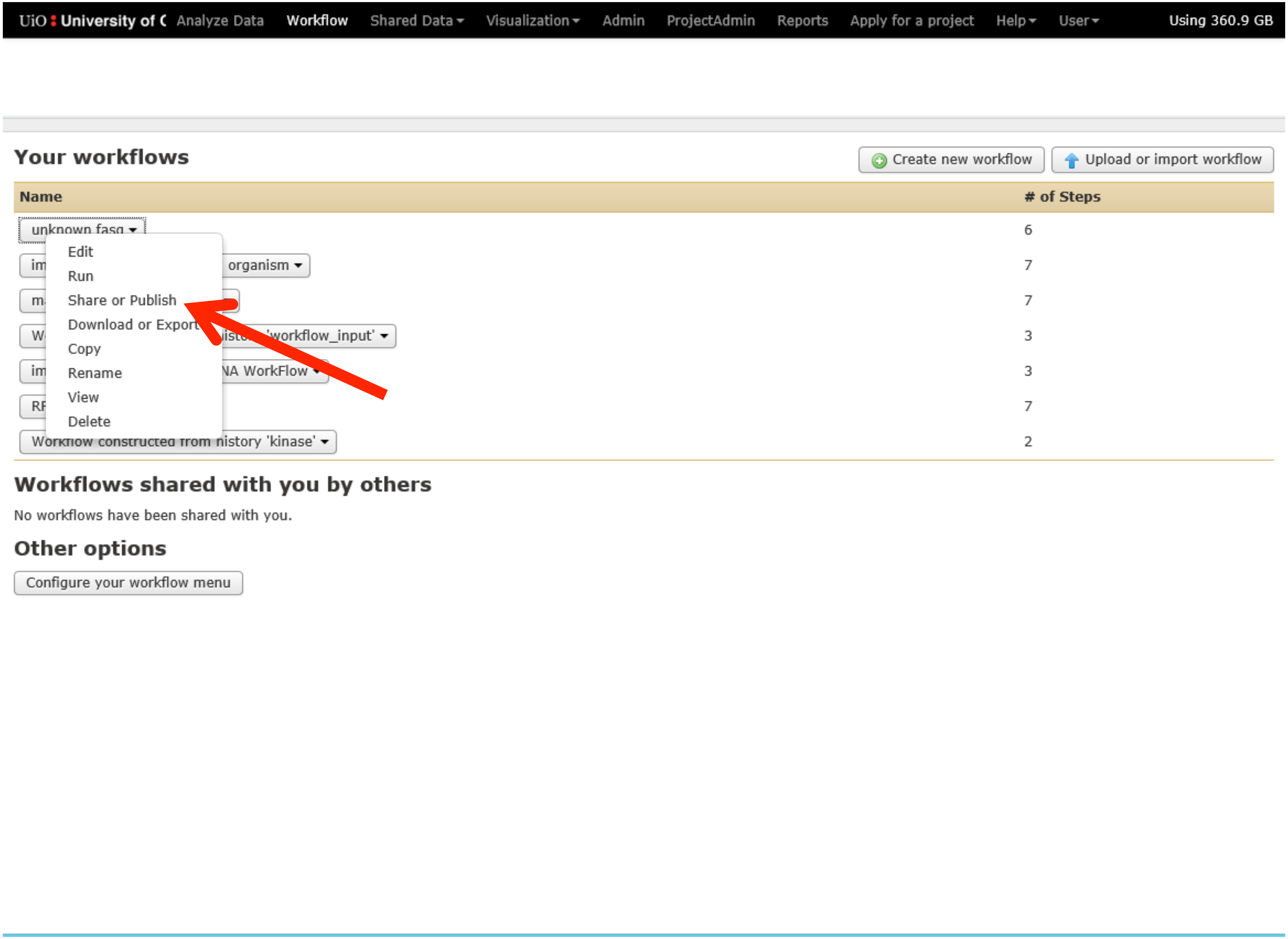
Run workflow

unknown fastq
595.1 KB

9: Fetch taxonomic representation on data 7
44 lines
format: taxonomy, database: ?
Starting job 8994995 ("") on c2-3 at Tue Nov 18 13:42:49 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994995.bat+ 378284K 10176K Job step resource usage: JobID JobName A

```
1.Name
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:5:352#CGATGT/1
HWI-EAS210R_0001:6:1:3:1375#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:4:911#CGATGT/1
```

7: Convert on data 6
64 lines
format: tabular, database: ?
Starting job 8994983 ("") on c2-3 at Tue Nov 18 13:38:15 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994983.bat+ 331856K 5836K Job step resource usage: JobID JobName A



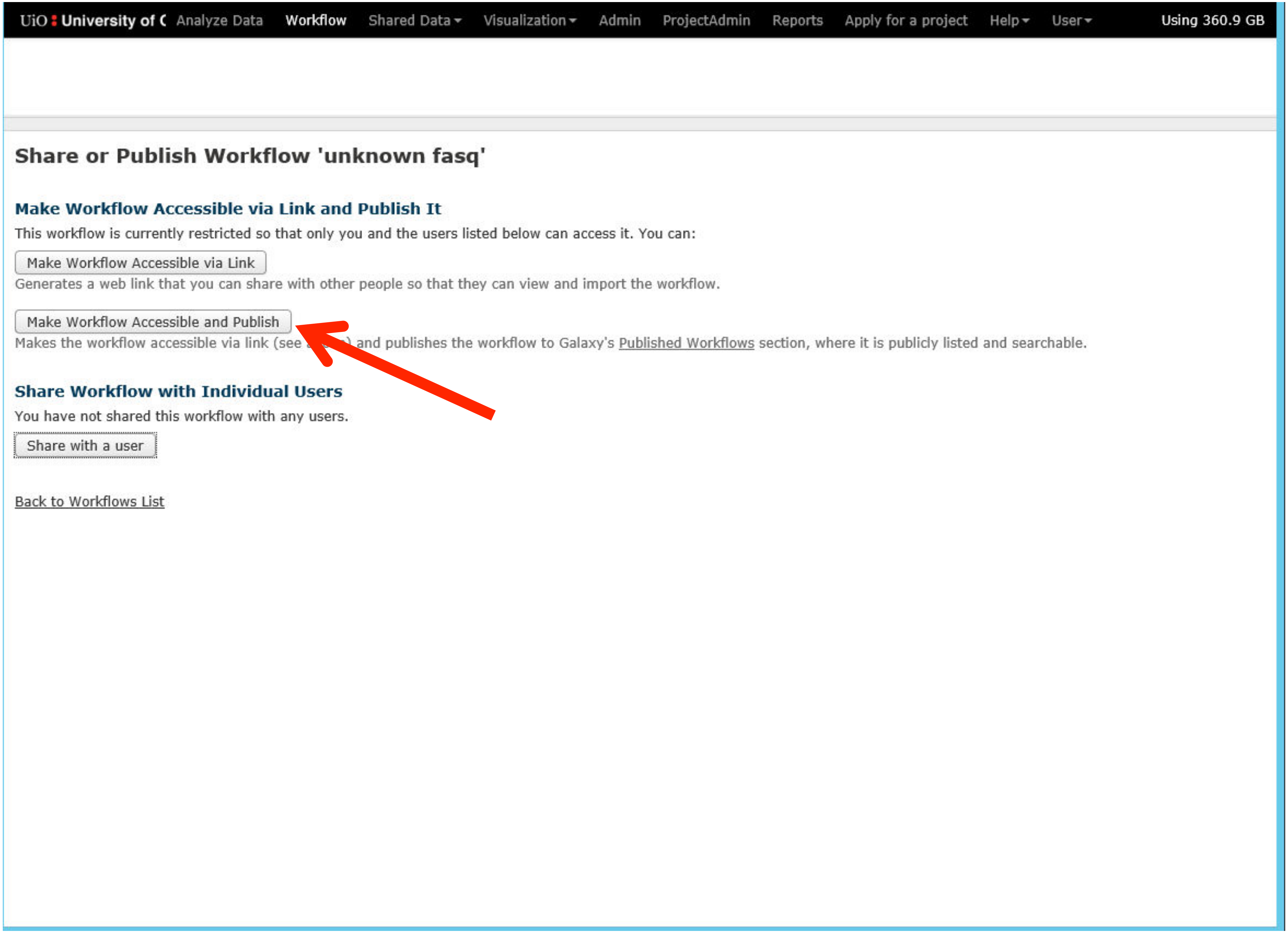
Your workflows

Name	# of Steps
unknown fasq	6
im… organism	7
m…	7
W… istory 'workflow_input'	3
im… NA WorkFlow	3
RF	7
Workflow constructed from history 'kinase'	2

Workflows shared with you by others

No workflows have been shared with you.

Other options



Share or Publish Workflow 'unknown fasq'

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:

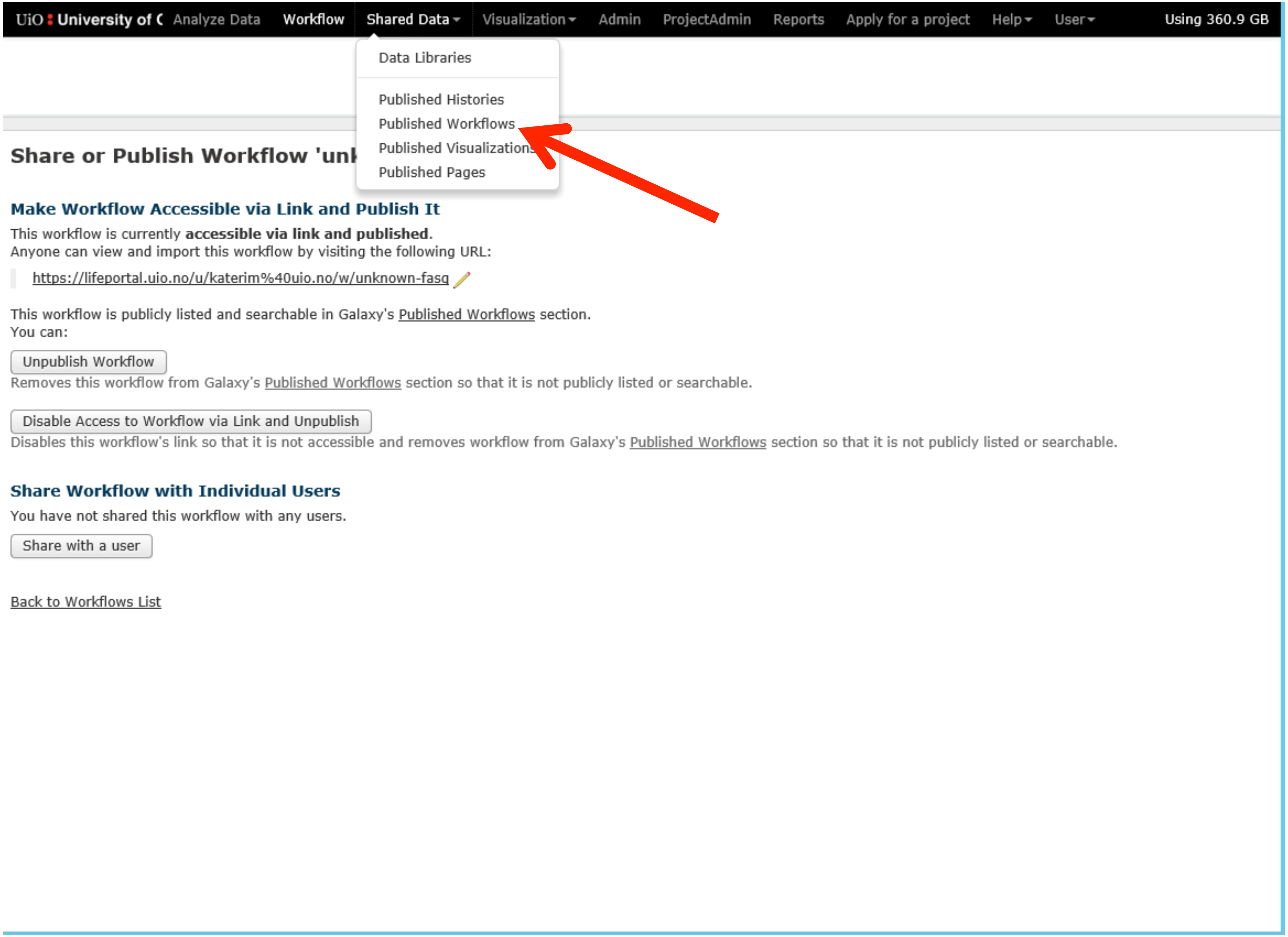
Generates a web link that you can share with other people so that they can view and import the workflow.

Makes the workflow accessible via link (see [illegible]) 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.

Back to Workflows List



Share or Publish Workflow 'unk

Make Workflow Accessible via Link and Publish It

This workflow is currently **accessible via link and published**.
Anyone can view and import this workflow by visiting the following URL:

https://lifeportal.uio.no/u/katerim%40uio.no/w/unknown-fasq

This workflow is publicly listed and searchable in Galaxy's Published Workflows section.
You can:

Unpublish Workflow

Removes this workflow from Galaxy's Published Workflows section so that it is not publicly listed or searchable.

Disable Access to Workflow via Link and Unpublish

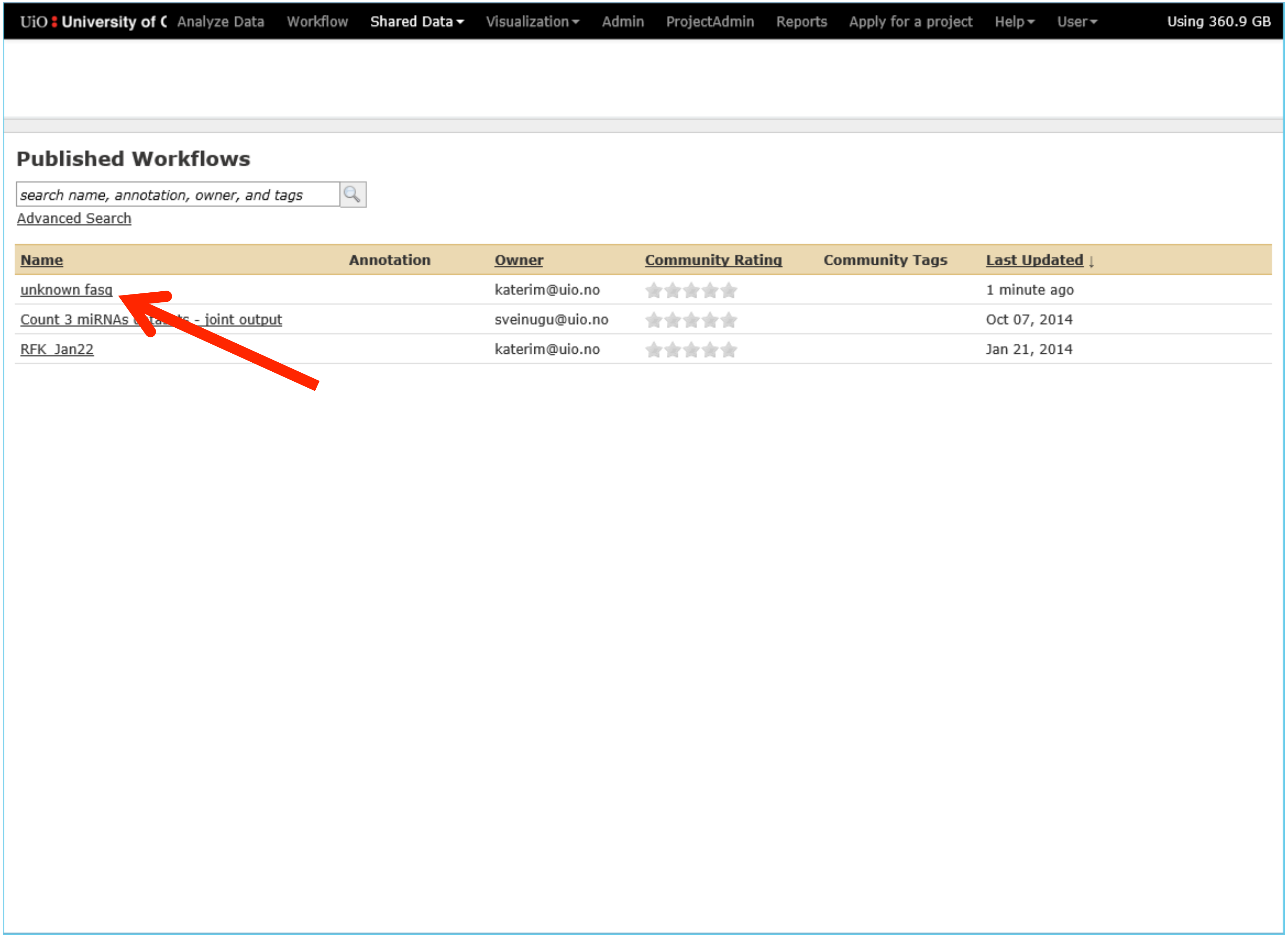
Disables this workflow's link so that it is not accessible and removes workflow from Galaxy's Published Workflows section so that it is not publicly listed or searchable.

Share Workflow with Individual Users

You have not shared this workflow with any users.

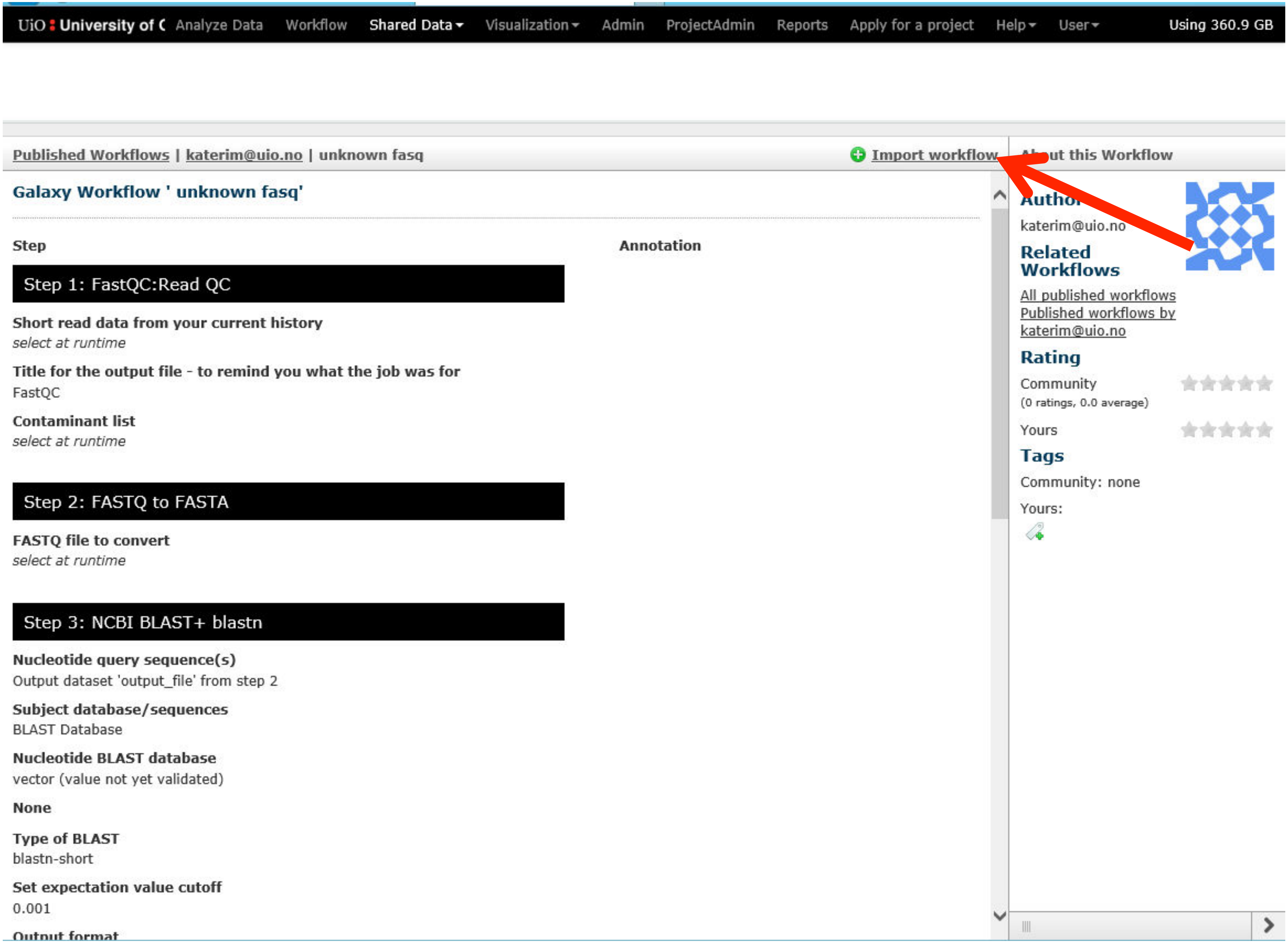
Share with a user

Back to Workflows List



Published Workflows

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
unknown fasq		katerim@uio.no	☆☆☆☆☆		1 minute ago
Count 3 miRNAs datasets - joint output		sveinugu@uio.no	☆☆☆☆☆		Oct 07, 2014
RFK_Jan22		katerim@uio.no	☆☆☆☆☆		Jan 21, 2014



Published Workflows | katerim@uio.no | unknown fasq

About this Workflow

Galaxy Workflow ' unknown fasq'

Step	Annotation
Step 1: FastQC:Read QC	

Short read data from your current history
select at runtime

Title for the output file - to remind you what the job was for
FastQC

Contaminant list
select at runtime

Step 2: FASTQ to FASTA

FASTQ file to convert
select at runtime

Step 3: NCBI BLAST+ blastn

Nucleotide query sequence(s)
Output dataset 'output_file' from step 2

Subject database/sequences
BLAST Database

Nucleotide BLAST database
vector (value not yet validated)

None

Type of BLAST
blastn-short

Set expectation value cutoff
0.001

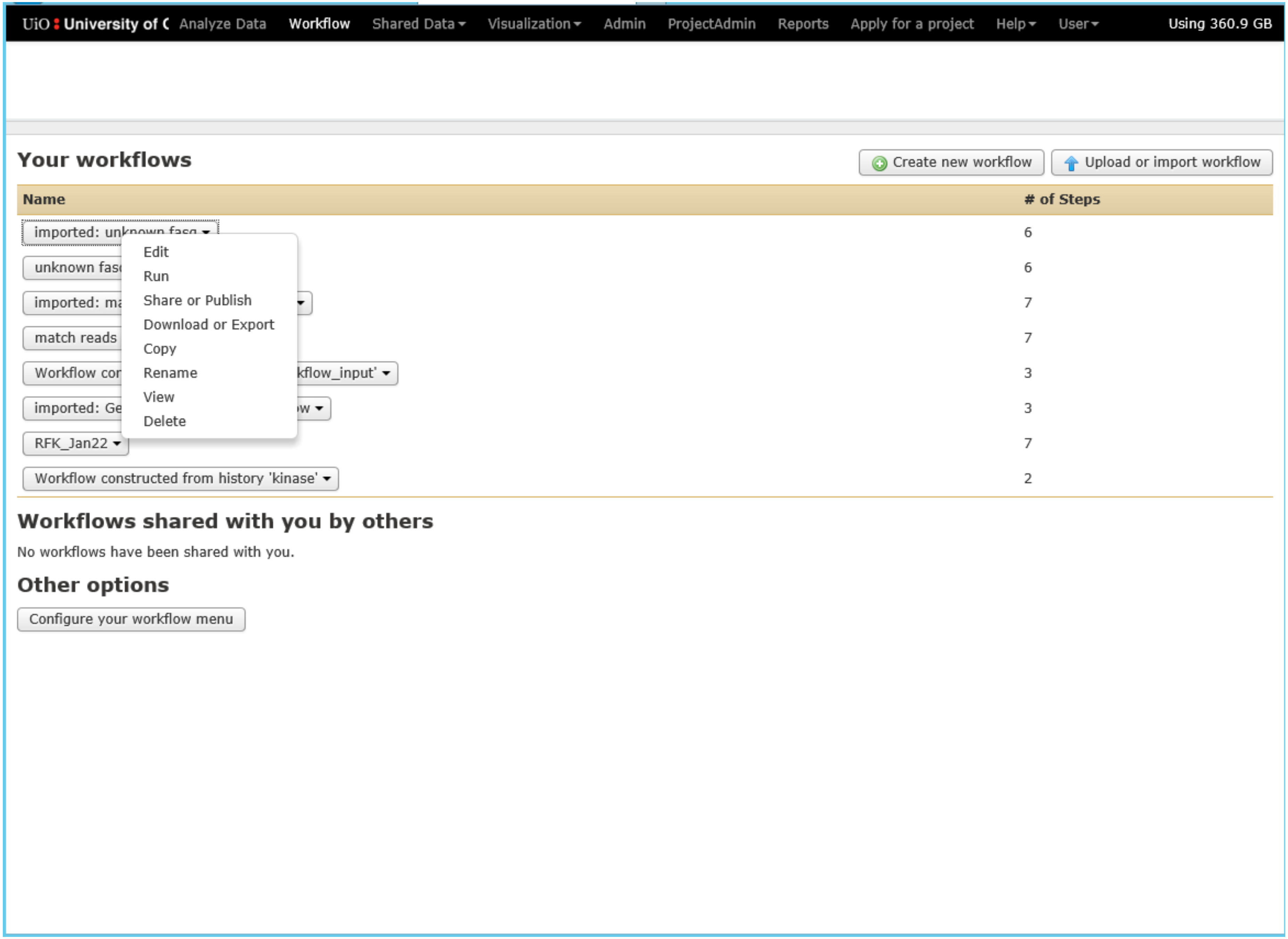
Output format

Author
katerim@uio.no

Related Workflows
All published workflows
Published workflows by katerim@uio.no

Rating
Community ☆☆☆☆☆
(0 ratings, 0.0 average)
Yours ☆☆☆☆☆

Tags
Community: none
Yours:



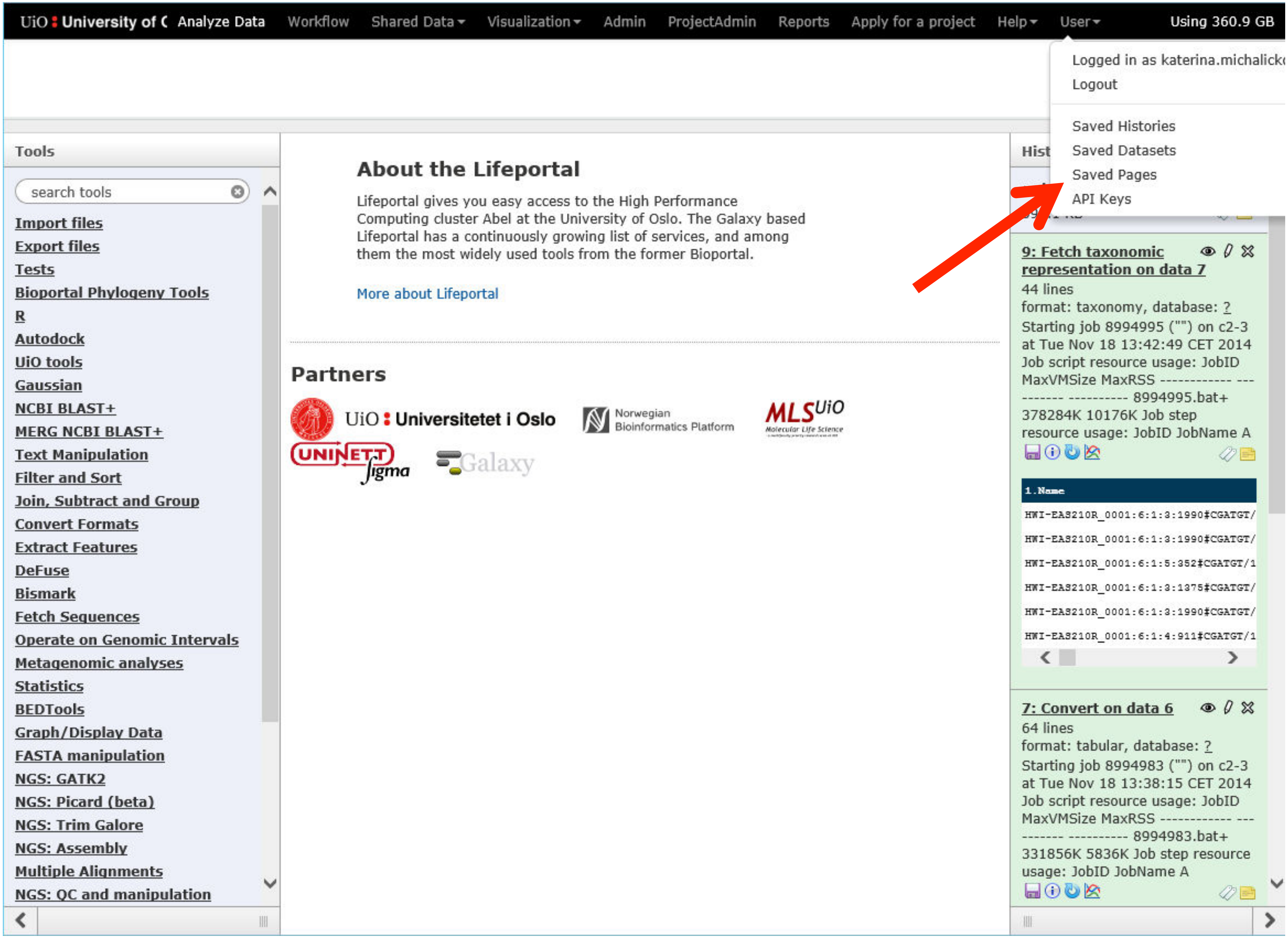
Your workflows

Name	# of Steps
imported: unknown fasq ▾	6
unknown fas	6
imported: ma ▾	7
match reads	7
Workflow cor kflow_input' ▾	3
imported: Ge w ▾	3
RFK_Jan22 ▾	7
Workflow constructed from history 'kinase' ▾	2

Workflows shared with you by others

No workflows have been shared with you.

Other options



About the Lifeport

Lifeport gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeport has a continuously growing list of services, and among them the most widely used tools from the former Bioportal.

More about Lifeport

Partners

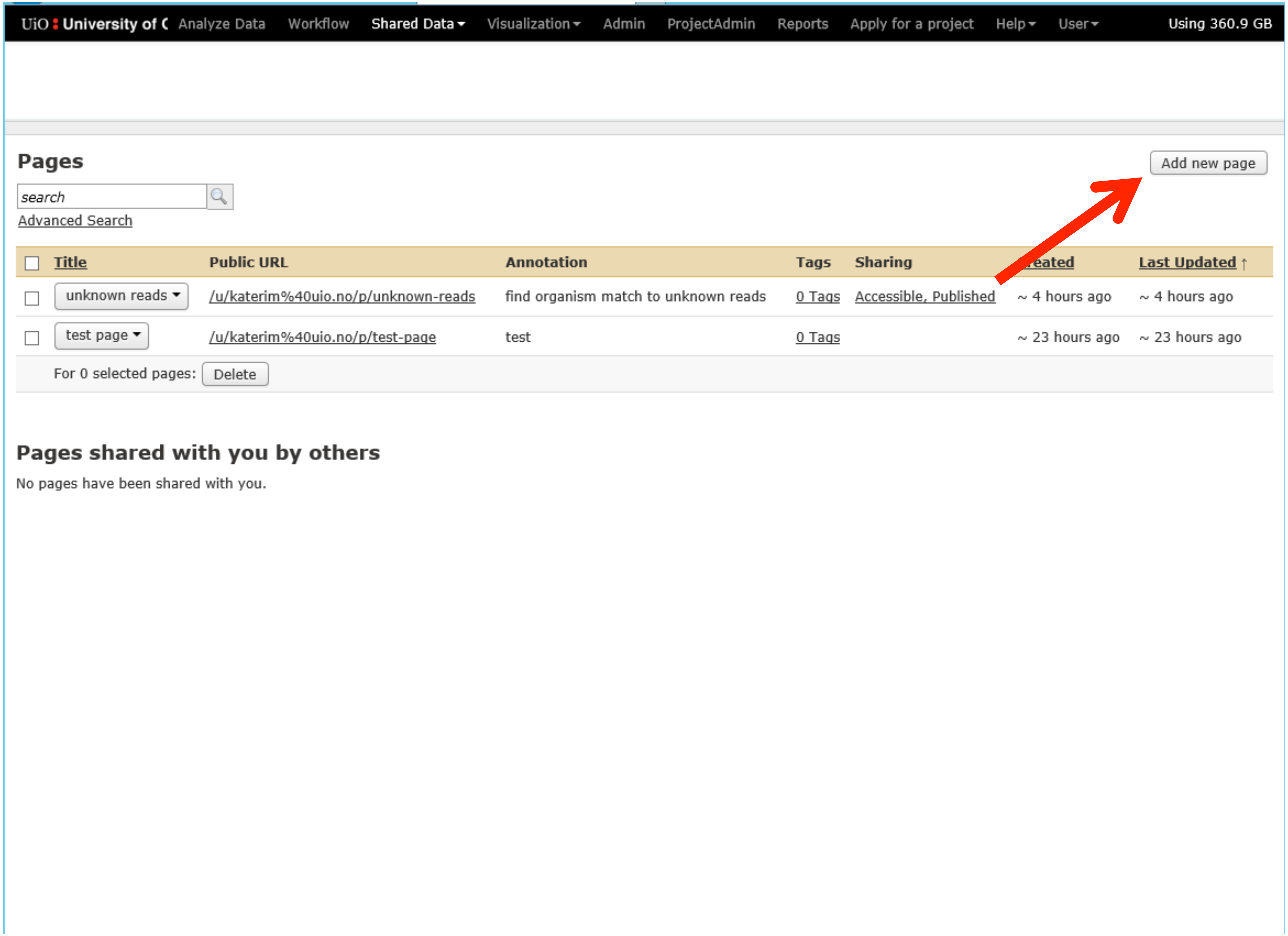
9: Fetch taxonomic representation on data 7

44 lines
format: taxonomy, database: ?
Starting job 8994995 ("") on c2-3 at Tue Nov 18 13:42:49 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994995.bat+ 378284K 10176K Job step resource usage: JobID JobName A

```
1.Name
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:5:352#CGATGT/1
HWI-EAS210R_0001:6:1:3:1375#CGATGT/
HWI-EAS210R_0001:6:1:3:1990#CGATGT/
HWI-EAS210R_0001:6:1:4:911#CGATGT/1
```

7: Convert on data 6

64 lines
format: tabular, database: ?
Starting job 8994983 ("") on c2-3 at Tue Nov 18 13:38:15 CET 2014 Job script resource usage: JobID MaxVMSize MaxRSS ------------ ---------- ---------- 8994983.bat+ 331856K 5836K Job step resource usage: JobID JobName A



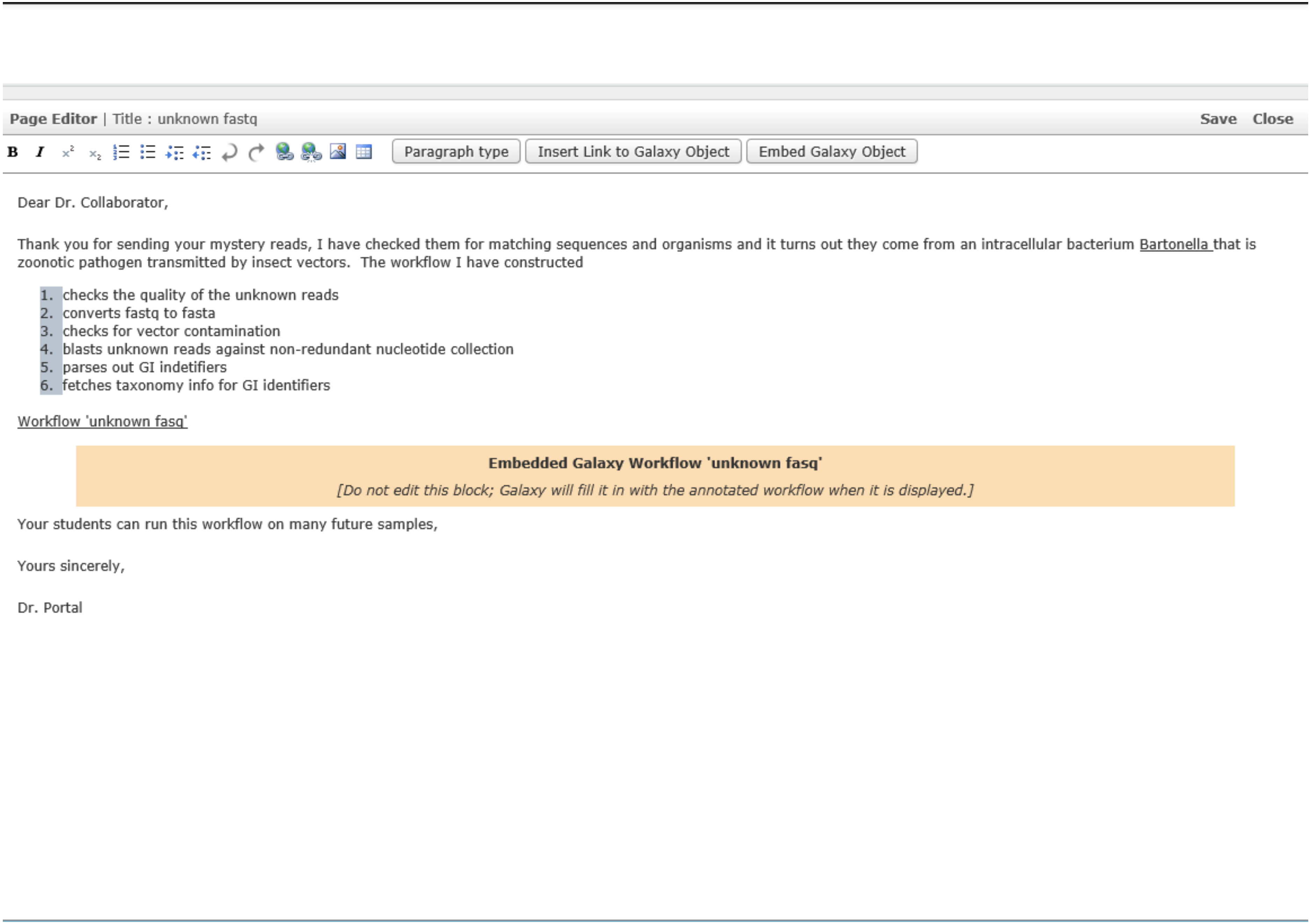
Pages

☐	Title	Public URL	Annotation	Tags	Sharing	eated	Last Updated ↑
☐	unknown reads ▾	/u/katerim%40uio.no/p/unknown-reads	find organism match to unknown reads	0 Tags	Accessible, Published	~ 4 hours ago	~ 4 hours ago
☐	test page ▾	/u/katerim%40uio.no/p/test-page	test	0 Tags		~ 23 hours ago	~ 23 hours ago

For 0 selected pages: Delete

Pages shared with you by others

No pages have been shared with you.



Dear Dr. Collaborator,

Thank you for sending your mystery reads, I have checked them for matching sequences and organisms and it turns out they come from an intracellular bacterium Bartonella that is zoonotic pathogen transmitted by insect vectors. The workflow I have constructed

1. checks the quality of the unknown reads
2. converts fastq to fasta
3. checks for vector contamination
4. blasts unknown reads against non-redundant nucleotide collection
5. parses out GI indetifiers
6. fetches taxonomy info for GI identifiers

Workflow 'unknown fasq'

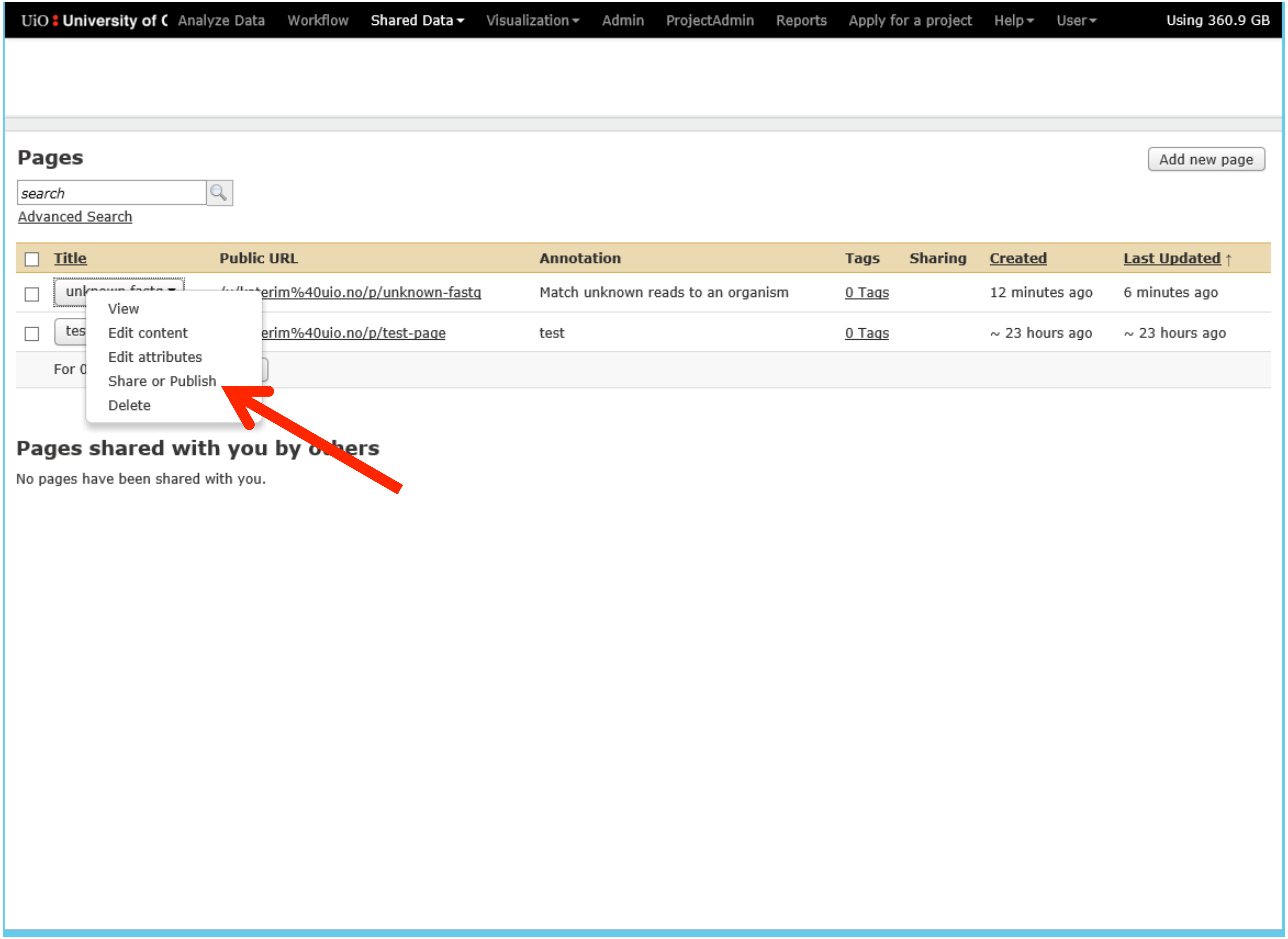
Embedded Galaxy Workflow 'unknown fasq'

[Do not edit this block; Galaxy will fill it in with the annotated workflow when it is displayed.]

Your students can run this workflow on many future samples,

Yours sincerely,

Dr. Portal



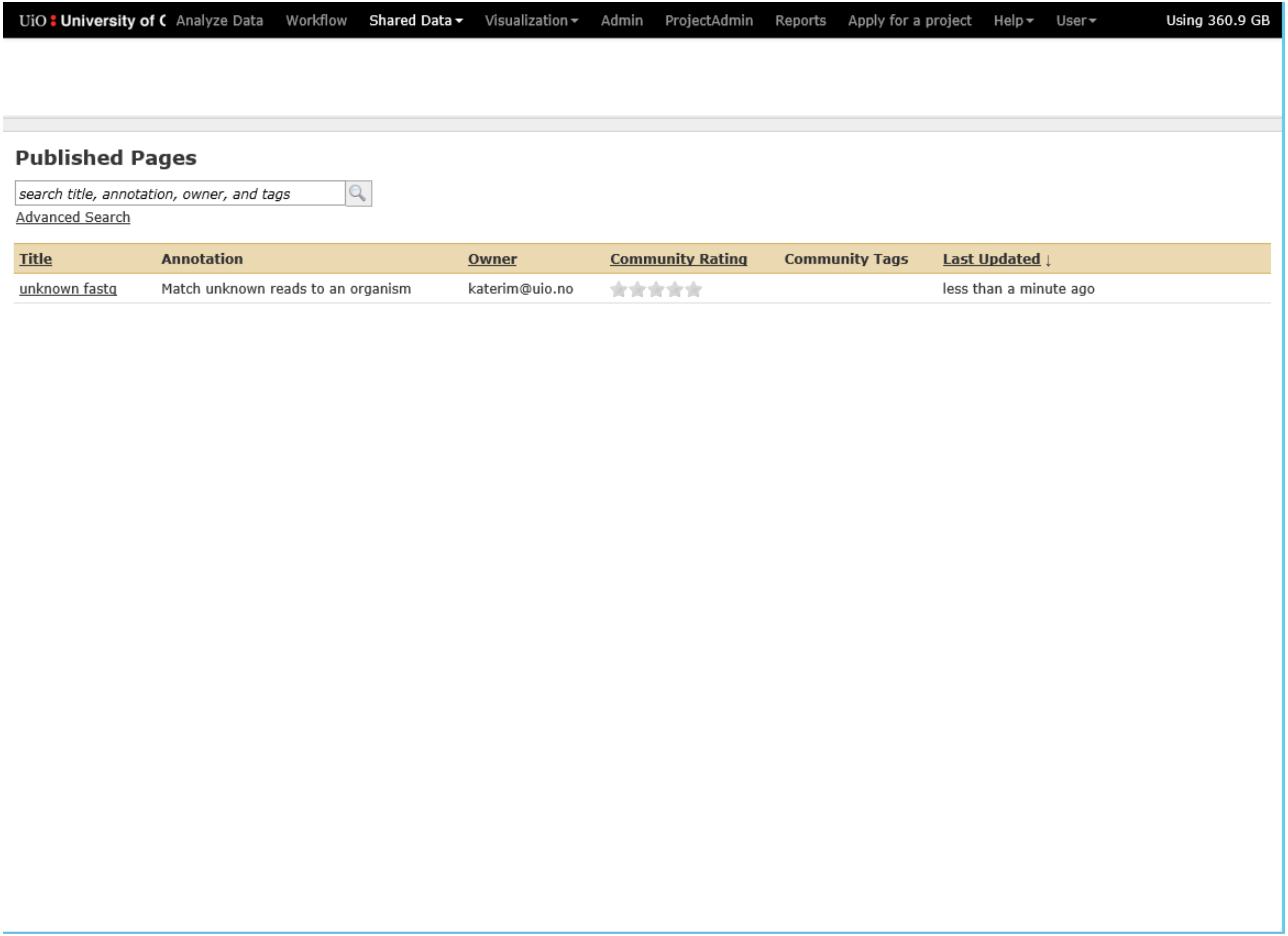
Pages

☐	Title	Public URL	Annotation	Tags	Sharing	Created	Last Updated ↑
☐	unk	erim%40uio.no/p/unknown-fastq	Match unknown reads to an organism	0 Tags		12 minutes ago	6 minutes ago
☐	tes	erim%40uio.no/p/test-page	test	0 Tags		~ 23 hours ago	~ 23 hours ago

For 0

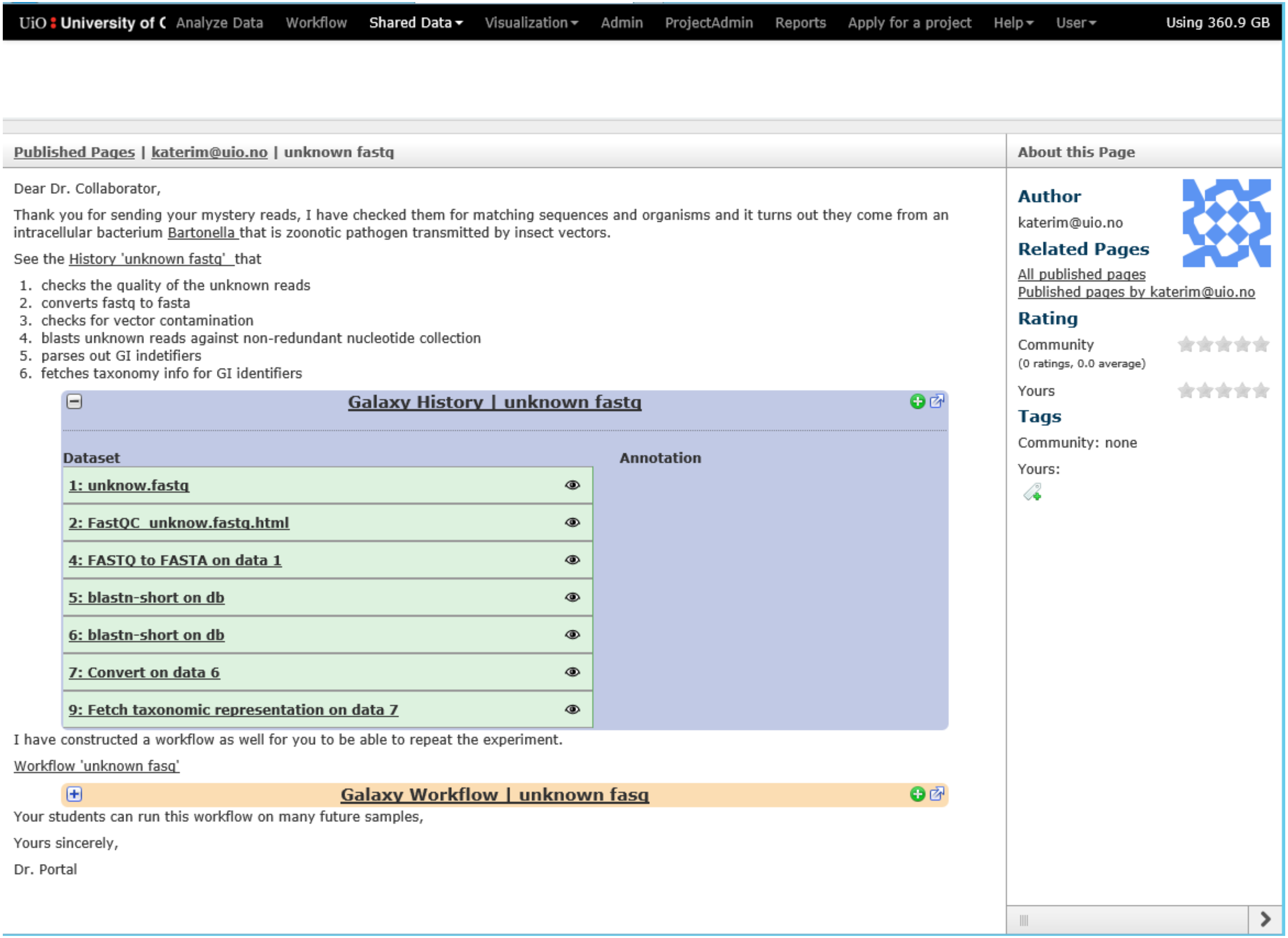
Pages shared with you by others

No pages have been shared with you.



Published Pages

Title	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
unknown fastq	Match unknown reads to an organism	katerim@uio.no	☆☆☆☆☆		less than a minute ago



Published Pages | katerim@uio.no | unknown fastq

Dear Dr. Collaborator,

Thank you for sending your mystery reads, I have checked them for matching sequences and organisms and it turns out they come from an intracellular bacterium Bartonella that is zoonotic pathogen transmitted by insect vectors.

See the History 'unknown fastq' that

1. checks the quality of the unknown reads
2. converts fastq to fasta
3. checks for vector contamination
4. blasts unknown reads against non-redundant nucleotide collection
5. parses out GI indetifiers
6. fetches taxonomy info for GI identifiers

Galaxy History | unknown fastq

Dataset	Annotation
1: unknow.fastq	
2: FastQC_unknow.fastq.html	
4: FASTQ to FASTA on data 1	
5: blastn-short on db	
6: blastn-short on db	
7: Convert on data 6	
9: Fetch taxonomic representation on data 7	

I have constructed a workflow as well for you to be able to repeat the experiment.

Workflow 'unknown fasq'

Galaxy Workflow | unknown fasq

Your students can run this workflow on many future samples,

Yours sincerely,

Dr. Portal

About this Page

Author

katerim@uio.no

Related Pages

All published pages
Published pages by katerim@uio.no

Rating

Community ☆☆☆☆☆
(0 ratings, 0.0 average)

Yours ☆☆☆☆☆

Tags

Community: none

Yours: