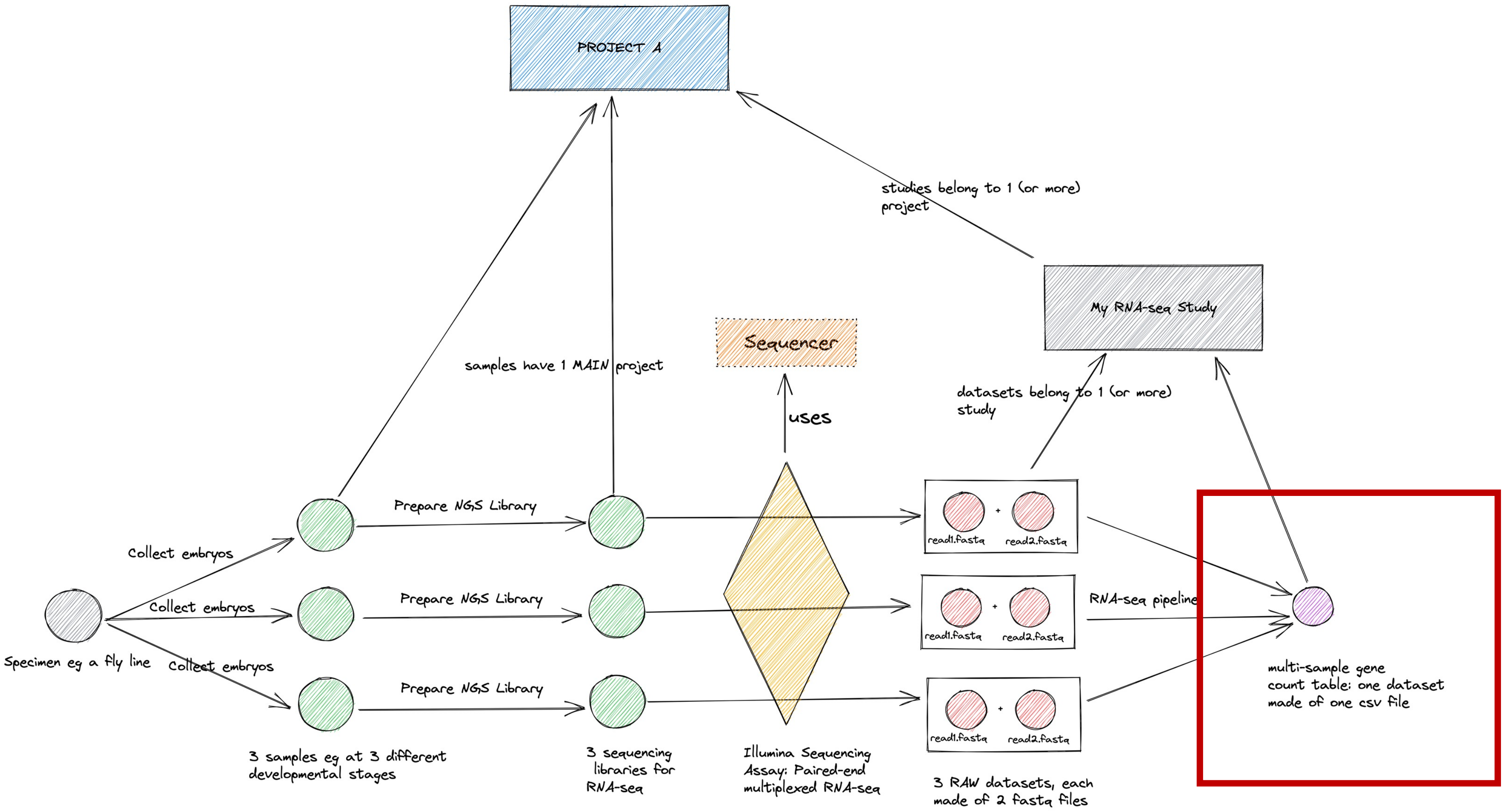
STOCKS Training: Derived Datasets and Analysis

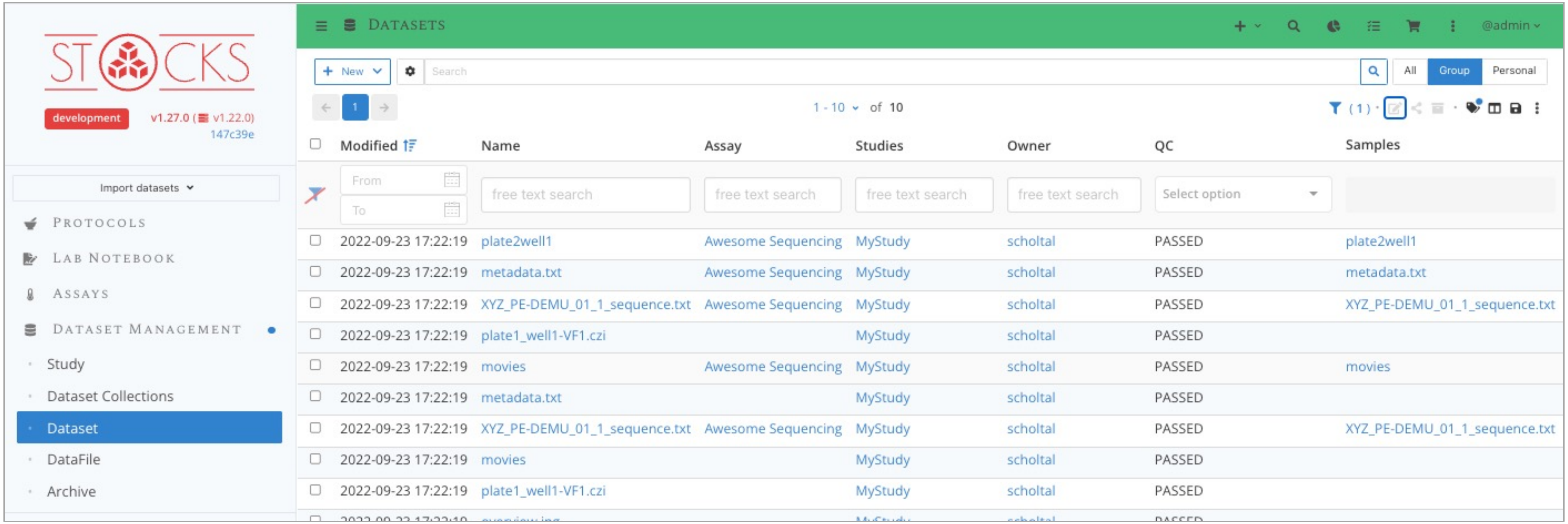
Charles Girardot

Genome Biology Computational Support (GBCS)

STOCKS aims at storing the analyzed datasets

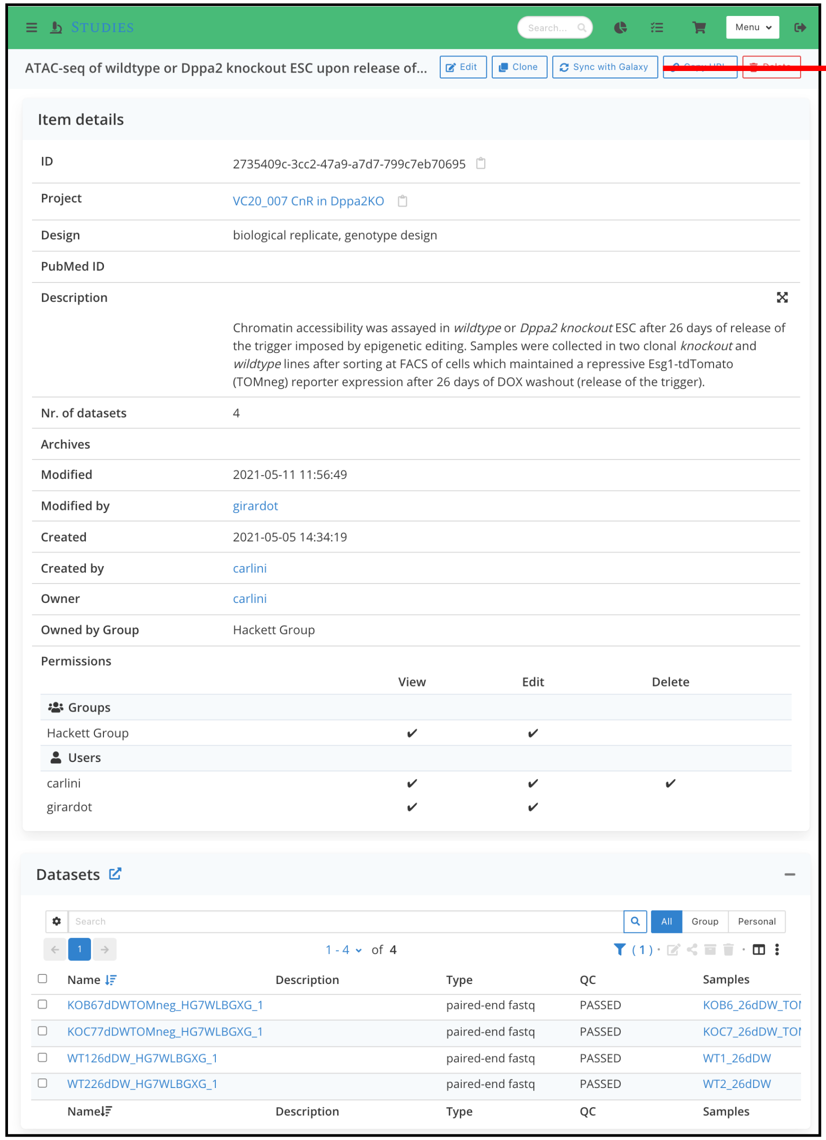


Raw and derived datasets are listed under the same menu



☐	Modified	Name	Assay	Studies	Owner	QC	Samples
☐	2022-09-23 17:22:19	plate2well1	Awesome Sequencing	MyStudy	scholtal	PASSED	plate2well1
☐	2022-09-23 17:22:19	metadata.txt	Awesome Sequencing	MyStudy	scholtal	PASSED	metadata.txt
☐	2022-09-23 17:22:19	XYZ_PE-DEMU_01_1_sequence.txt	Awesome Sequencing	MyStudy	scholtal	PASSED	XYZ_PE-DEMU_01_1_sequence.txt
☐	2022-09-23 17:22:19	plate1_well1-VF1.czi		MyStudy	scholtal	PASSED	
☐	2022-09-23 17:22:19	movies	Awesome Sequencing	MyStudy	scholtal	PASSED	movies
☐	2022-09-23 17:22:19	metadata.txt		MyStudy	scholtal	PASSED	
☐	2022-09-23 17:22:19	XYZ_PE-DEMU_01_1_sequence.txt	Awesome Sequencing	MyStudy	scholtal	PASSED	XYZ_PE-DEMU_01_1_sequence.txt
☐	2022-09-23 17:22:19	movies		MyStudy	scholtal	PASSED	
☐	2022-09-23 17:22:19	plate1_well1-VF1.czi		MyStudy	scholtal	PASSED	

Raw data processing made easy with Galaxy Link



ATAC-seq of wildtype or Dppa2 knockout ESC upon release of...

Item details

ID	2735409c-3cc2-47a9-a7d7-799c7eb70695
Project	VC20_007 CnR in Dppa2KO
Design	biological replicate, genotype design
PubMed ID	
Description	Chromatin accessibility was assayed in *wildtype* or *Dppa2 knockout* ESC after 26 days of release of the trigger imposed by epigenetic editing. Samples were collected in two clonal *knockout* and *wildtype* lines after sorting at FACS of cells which maintained a repressive Esg1-tdTomato (TOMneg) reporter expression after 26 days of DOX washout (release of the trigger).
Nr. of datasets	4
Archives	
Modified	2021-05-11 11:56:49
Modified by	girardot
Created	2021-05-05 14:34:19
Created by	carlini
Owner	carlini
Owned by Group	Hackett Group

Permissions

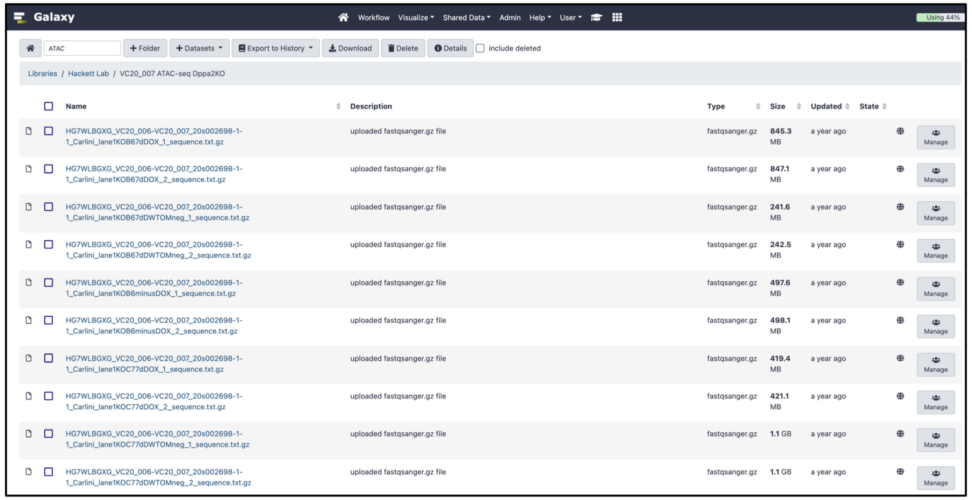
	View	Edit	Delete
Groups			
Hackett Group	✔	✔	
Users			
carlini	✔	✔	✔
girardot	✔	✔	

Datasets

☐	Name	Description	Type	QC	Samples
☐	KOB67dDWTOMneg_HG7WLBGXG_1		paired-end fastq	PASSED	KOB6_26dDW_TO
☐	KOC77dDWTOMneg_HG7WLBGXG_1		paired-end fastq	PASSED	KOC7_26dDW_TO
☐	WT126dDW_HG7WLBGXG_1		paired-end fastq	PASSED	WT1_26dDW
☐	WT226dDW_HG7WLBGXG_1		paired-end fastq	PASSED	WT2_26dDW
	Name	Description	Type	QC	Samples

No data duplication

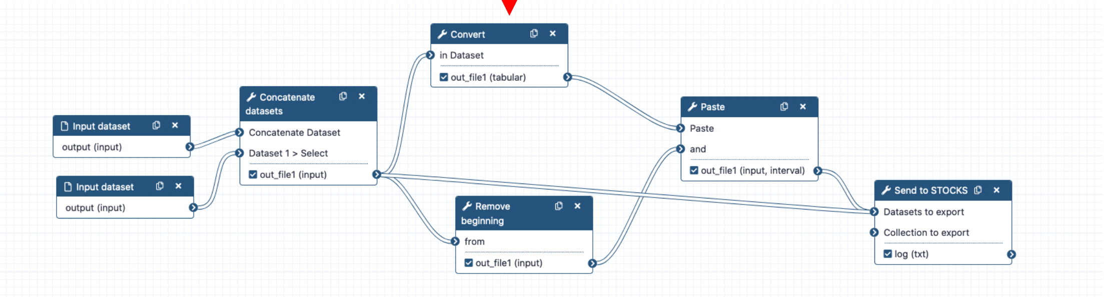
https://galaxy.embl.de



Libraries / Hackett Lab / VC20_007 ATAC-seq Dppa2KO

Name	Description	Type	Size	Updated	State
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB67dDOX_1_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	845.3 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB67dDOX_2_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	847.1 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB67dDWTOMneg_1_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	241.6 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB67dDWTOMneg_2_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	242.5 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB6minusDOX_1_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	497.6 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOB6minusDOX_2_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	498.1 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOC77dDOX_1_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	419.4 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOC77dDOX_2_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	421.1 MB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOC77dDWTOMneg_1_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	1.1 GB	a year ago	
HG7WLBGXG_VC20_006-VC20_007_20s002698-1-1_Carlini_lane1KOC77dDWTOMneg_2_sequence.txt.gz	uploaded fastqsanger.gz file	fastqsanger.gz	1.1 GB	a year ago	

Start ATAC-seq Standard WF

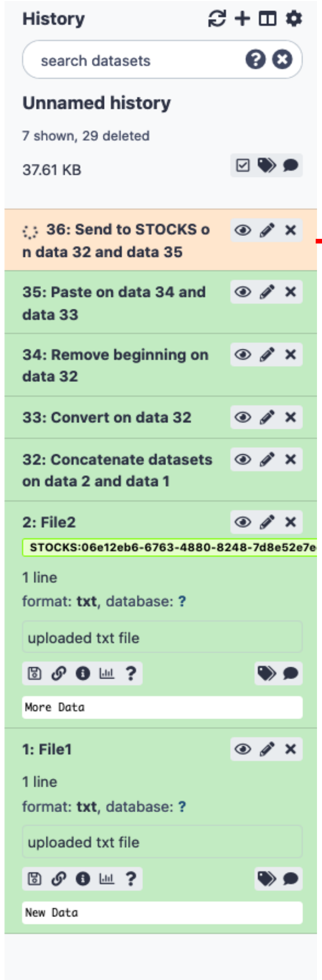


All jobs executed on ITS HPC

Analyzed Data

Sending derived datasets back to STOCKS

Under active dev



History

search datasets

Unnamed history

7 shown, 29 deleted

37.61 KB

36: Send to STOCKS on data 32 and data 35

35: Paste on data 34 and data 33

34: Remove beginning on data 32

33: Convert on data 32

32: Concatenate datasets on data 2 and data 1

2: File2

STOCKS:06e12eb6-6763-4880-8248-7d8e52e7e

1 line

format: **txt**, database: ?

uploaded txt file

```
More Data
```

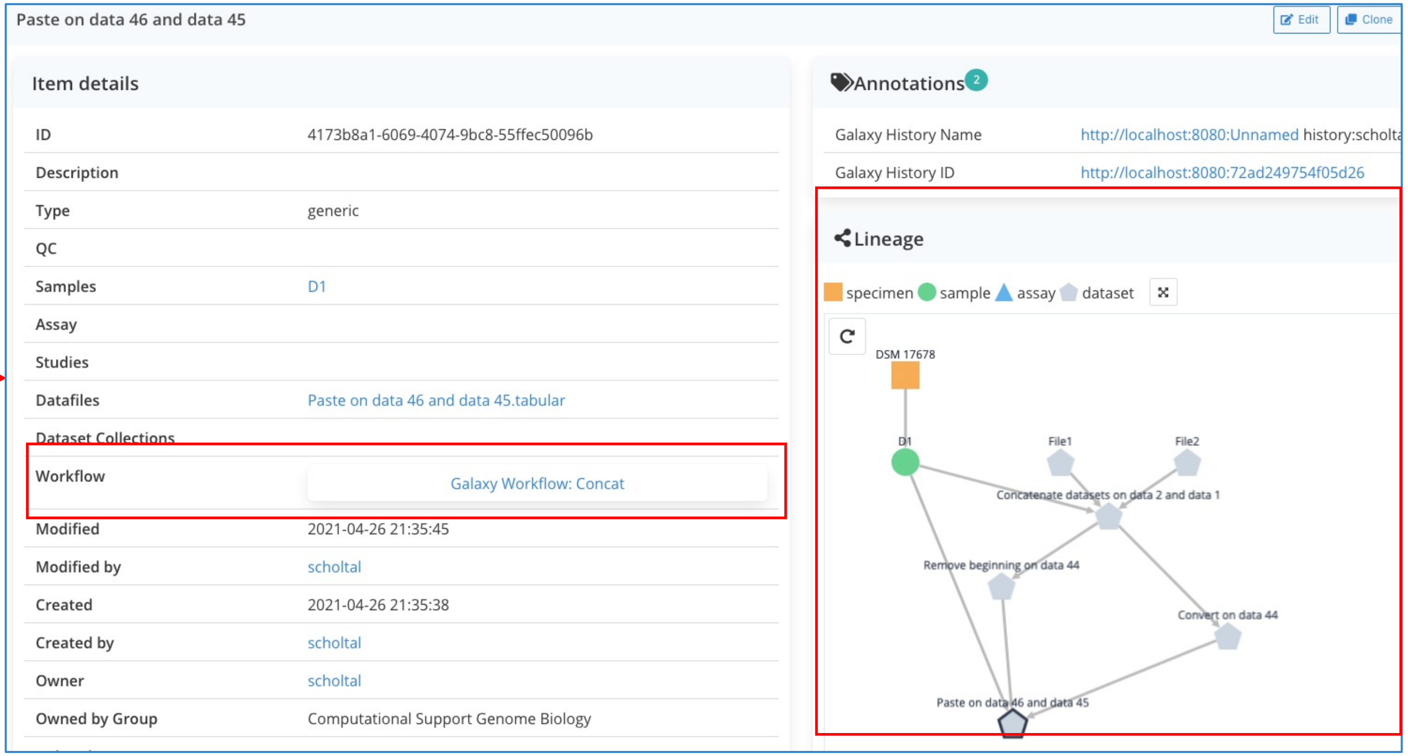
1: File1

1 line

format: **txt**, database: ?

uploaded txt file

```
New Data
```



Paste on data 46 and data 45

Item details

ID	4173b8a1-6069-4074-9bc8-55ffec50096b
Description	
Type	generic
QC	
Samples	D1
Assay	
Studies	
Datafiles	Paste on data 46 and data 45.tabular
Dataset Collections	
Workflow	Galaxy Workflow: Concat
Modified	2021-04-26 21:35:45
Modified by	scholtal
Created	2021-04-26 21:35:38
Created by	scholtal
Owner	scholtal
Owned by Group	Computational Support Genome Biology

Annotations 2

Galaxy History Name	http://localhost:8080:Unnamed history:scholta
Galaxy History ID	http://localhost:8080:72ad249754f05d26

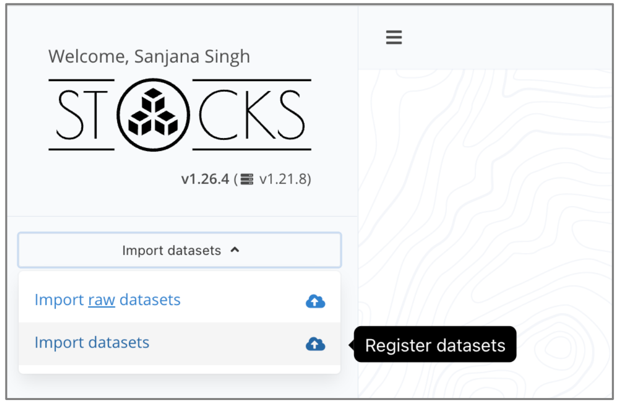
Lineage

🟧 specimen ● sample ▲ assay ⬠ dataset

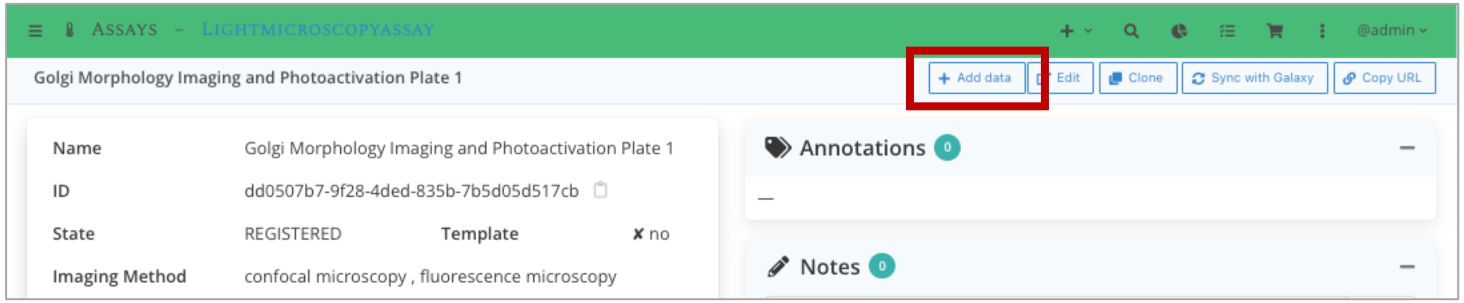
- ➢ Derived datasets are **automatically** linked to their parent
- ➢ The **complete data structure** is reproduced in STOCKS
- ➢ The executed **WF is saved as a protocol**
- ➢ **Links** to Galaxy are maintained

Processed datasets can also be manually imported

From the main Menu



From the Assay Page

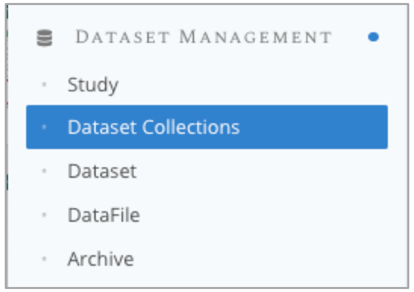


Golgi Morphology Imaging and Photoactivation Plate 1

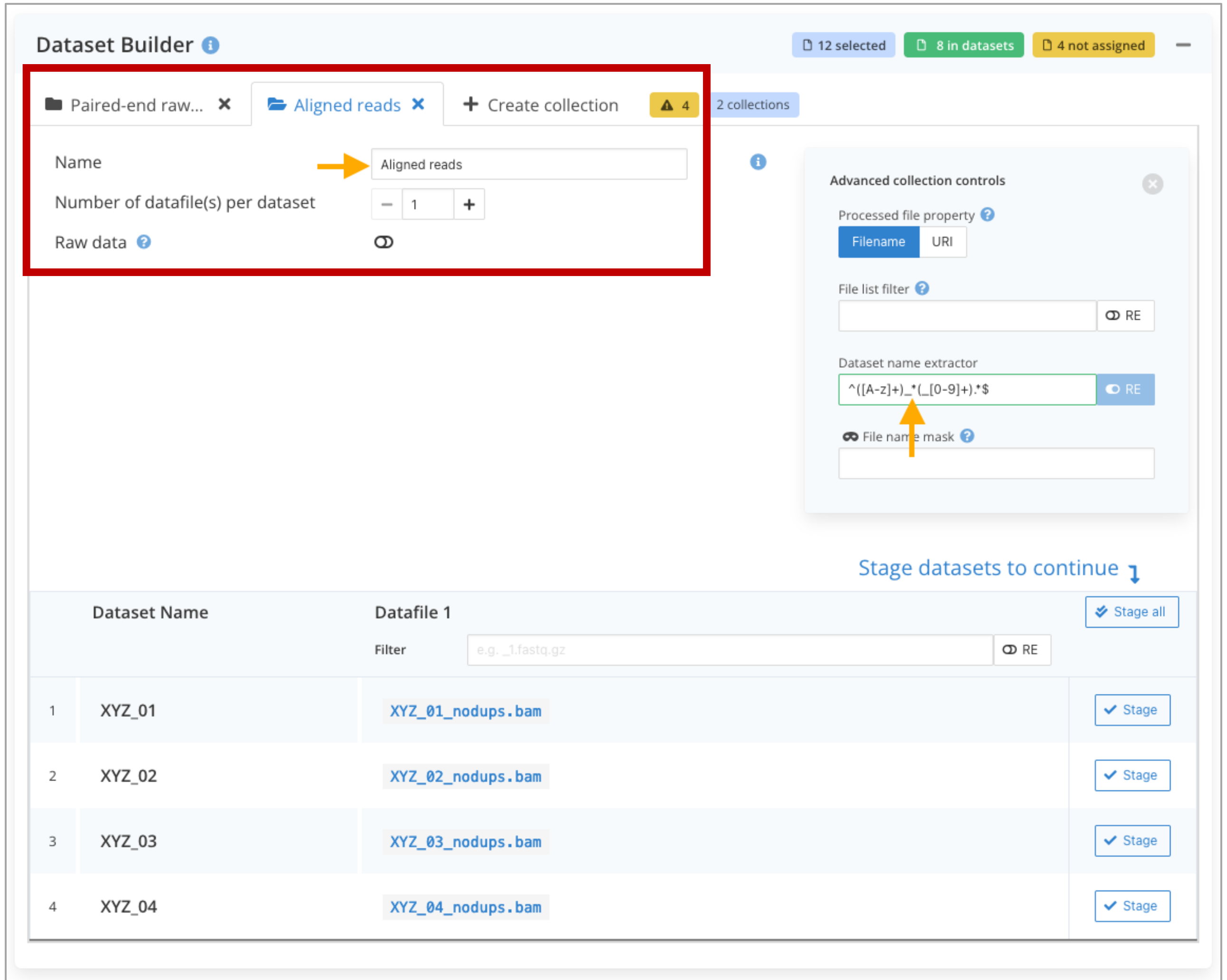
Name: Golgi Morphology Imaging and Photoactivation Plate 1
ID: dd0507b7-9f28-4ded-835b-7b5d05d517cb
State: REGISTERED
Template: ✗ no
Imaging Method: confocal microscopy , fluorescence microscopy

- Loading derived datasets uses the same dataset loader, but simplified
  - No Assay type selection nor Assay creation
  - Datasets can optionally linked to individual samples
- Linking derived datasets to samples is highly recommended (when relevant)

Import multiple Dataset Collection at once



- STOCKS also offers a **DatasetCollection** to (manually) group datasets into meaningful sets
  - e.g. “bam files”, “segmentation maps”
- They are created when importing datasets



Name: Aligned reads

Number of datafile(s) per dataset: 1

Raw data

Processed file property: Filename

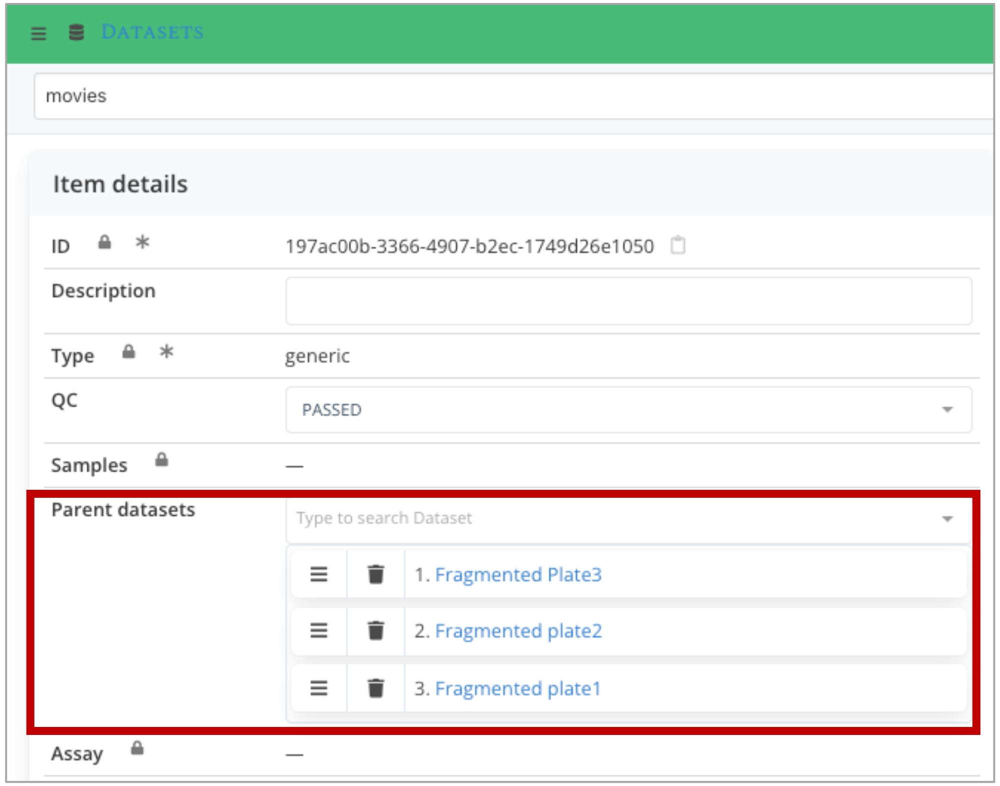
File list filter

Dataset name extractor: ^([A-z]+)_*(_[0-9]+).*$

File name mask

	Dataset Name	Datafile 1
		Filter: e.g. _1.fastq.gz
1	XYZ_01	XYZ_01_nodups.bam
2	XYZ_02	XYZ_02_nodups.bam
3	XYZ_03	XYZ_03_nodups.bam
4	XYZ_04	XYZ_04_nodups.bam

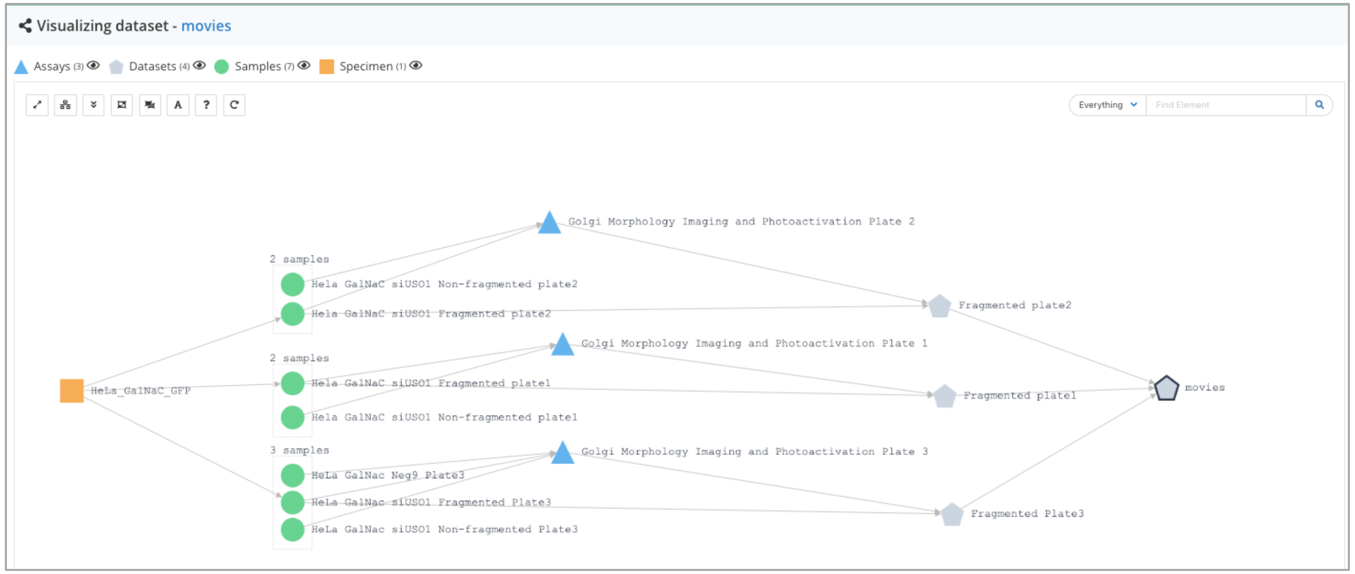
Linking processed datasets to parent datasets



movies

Item details

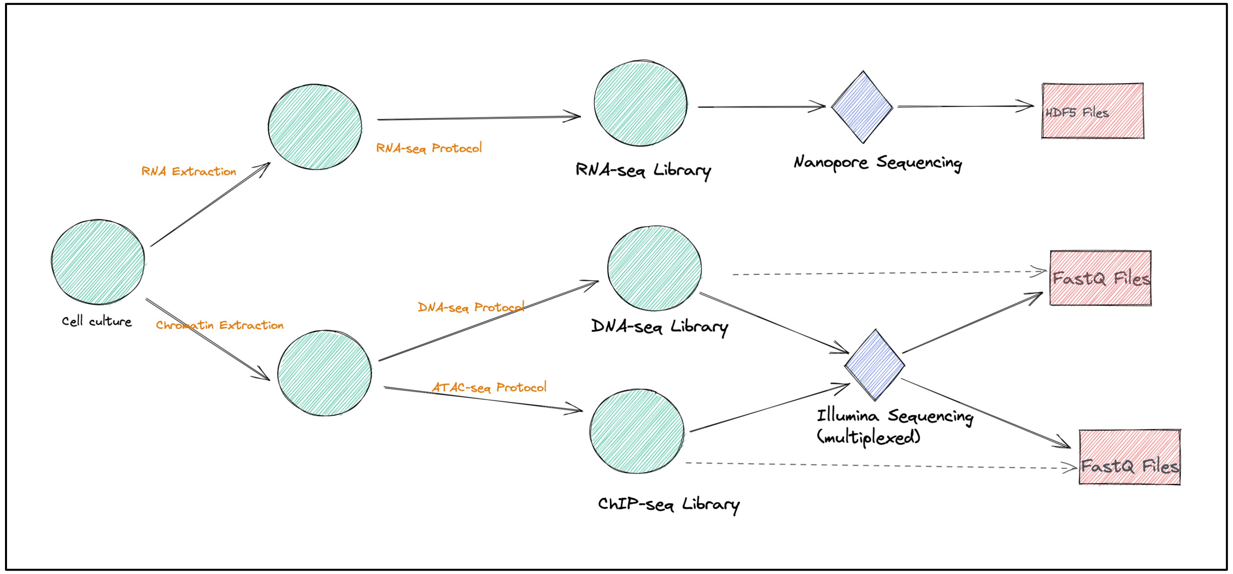
ID	197ac00b-3366-4907-b2ec-1749d26e1050
Description	
Type	generic
QC	PASSED
Samples	—
Parent datasets	Type to search Dataset 1. Fragmented Plate3 2. Fragmented plate2 3. Fragmented plate1
Assay	—

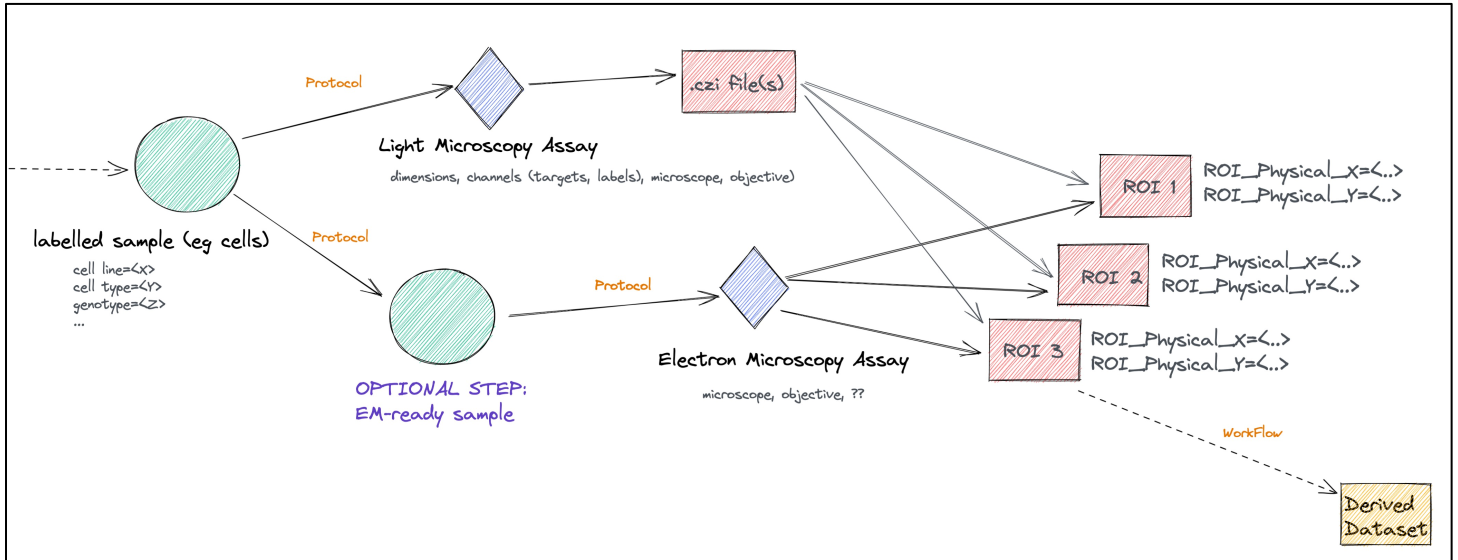


Visualizing dataset - movies

▲ Assays (3) 👁 ⬟ Datasets (4) 👁 ● Samples (7) 👁 ■ Specimen (1) 👁

STOCKS to support complex Multi-omics scenarios

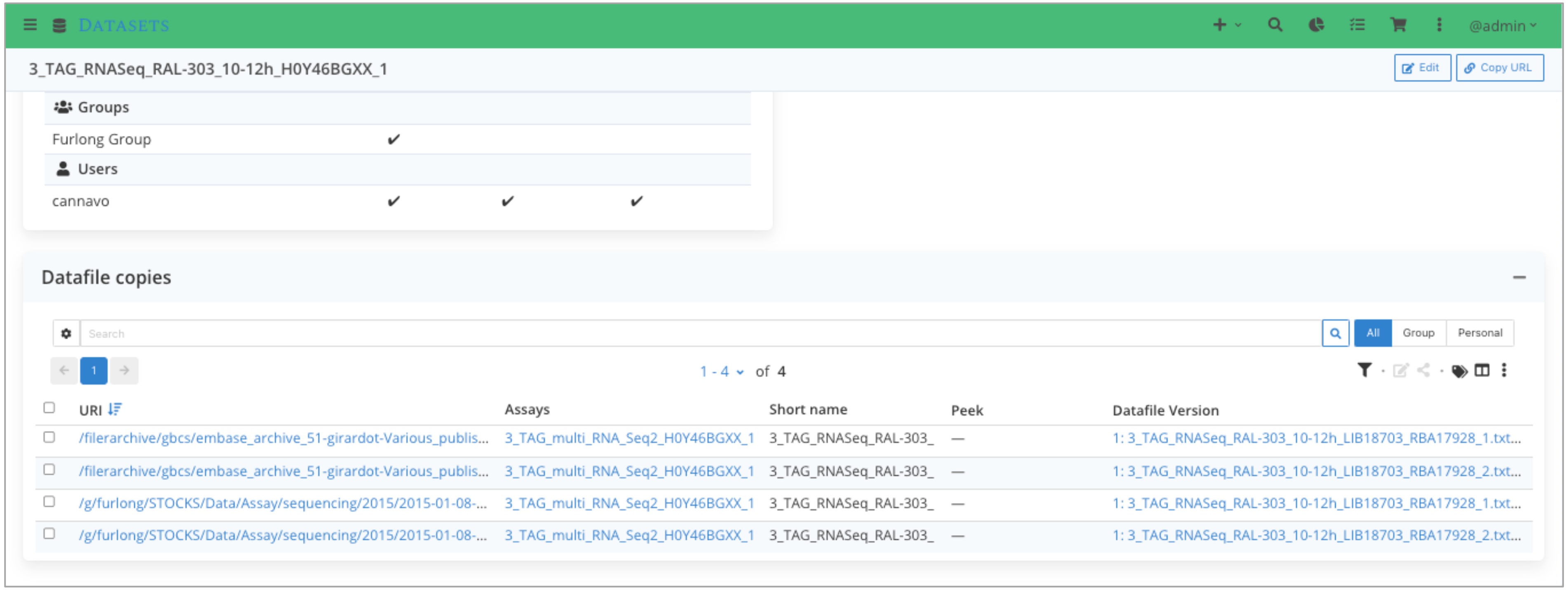




- Aggregate on common samples
- Dependent Datasets can be linked

Datafiles may have multiple copies (different locations)

- A paired-end datasets (== 2 datafiles) with two copies (hence 4 datafile copies)



3_TAG_RNASeq_RAL-303_10-12h_H0Y46BGXX_1

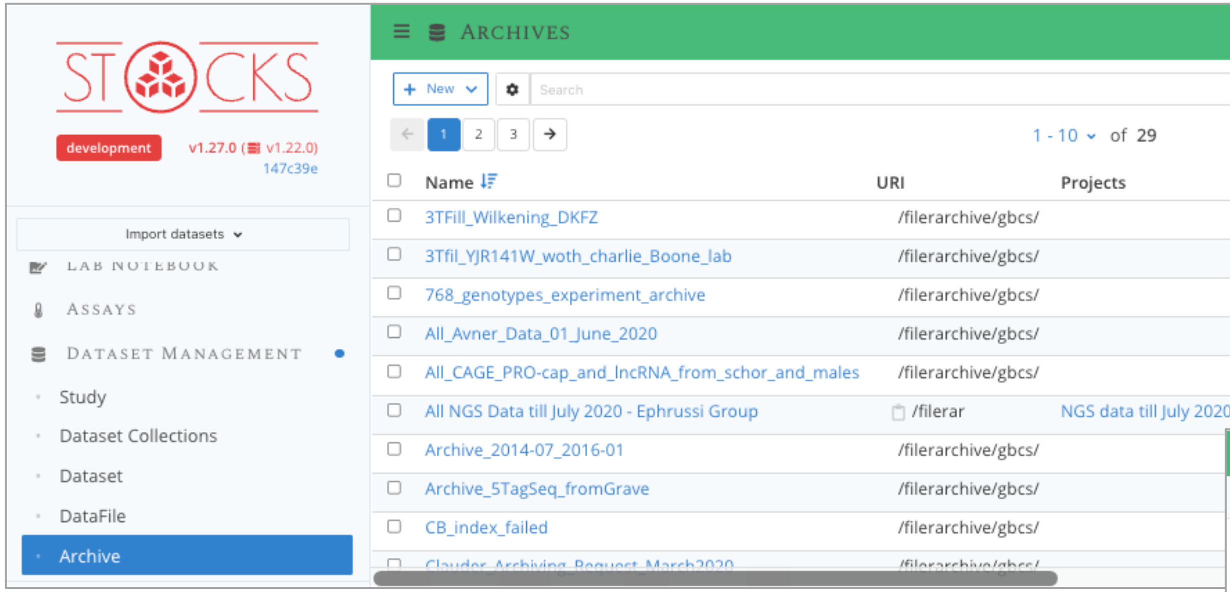
Groups			
Furlong Group	✓		
Users			
cannavo	✓	✓	✓

Datafile copies

☐	URI	Assays	Short name	Peek	Datafile Version
☐	/filerarchive/gbcs/embase_archive_51-girardot-Various_publis...	3_TAG_multi_RNA_Seq2_H0Y46BGXX_1	3_TAG_RNASeq_RAL-303_	—	1: 3_TAG_RNASeq_RAL-303_10-12h_LIB18703_RBA17928_1.txt...
☐	/filerarchive/gbcs/embase_archive_51-girardot-Various_publis...	3_TAG_multi_RNA_Seq2_H0Y46BGXX_1	3_TAG_RNASeq_RAL-303_	—	1: 3_TAG_RNASeq_RAL-303_10-12h_LIB18703_RBA17928_2.txt...
☐	/g/furlong/STOCKS/Data/Assay/sequencing/2015/2015-01-08-...	3_TAG_multi_RNA_Seq2_H0Y46BGXX_1	3_TAG_RNASeq_RAL-303_	—	1: 3_TAG_RNASeq_RAL-303_10-12h_LIB18703_RBA17928_1.txt...
☐	/g/furlong/STOCKS/Data/Assay/sequencing/2015/2015-01-08-...	3_TAG_multi_RNA_Seq2_H0Y46BGXX_1	3_TAG_RNASeq_RAL-303_	—	1: 3_TAG_RNASeq_RAL-303_10-12h_LIB18703_RBA17928_2.txt...

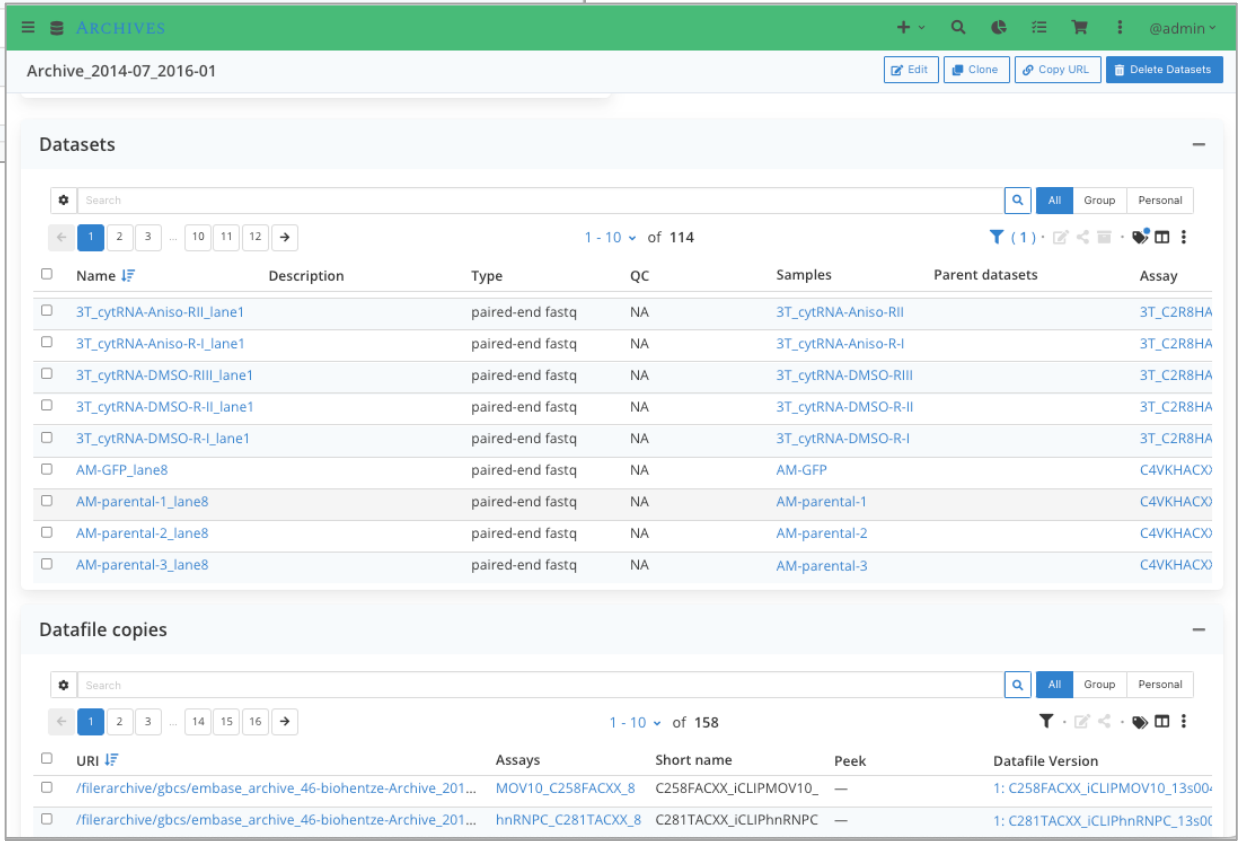
- Tracking copies is important:
  - to save space by deleting useless copies
  - All copies of human data must be deleted when project ends
  - STOCKS has a concept of **primary copy** (cannot be deleted but in special situations)

Data can be archived to free space



☐	Name	URI	Projects	Studies	Nr. of datasets
☐	3TFill_Wilkening_DKFZ	/filerarchive/gbcs/			15
☐	3Tfil_YJR141W_woth_charlie_Boone_lab	/filerarchive/gbcs/			8
☐	768_genotypes_experiment_archive	/filerarchive/gbcs/			18
☐	All_Avner_Data_01_June_2020	/filerarchive/gbcs/			492
☐	All_CAGE_PRO-cap_and_lncRNA_from_schor_and_males	/filerarchive/gbcs/			468
☐	All NGS Data till July 2020 - Ephrussi Group	/filerar	NGS data till July 2020 - Ephrussi Group	view 56 items	129
☐	Archive_2014-07_2016-01	/filerarchive/gbcs/			
☐	Archive_5TagSeq_fromGrave	/filerarchive/gbcs/			
☐	CB_index_failed	/filerarchive/gbcs/			
☐	Clauder_Archiving_Request_March2020	/filerarchive/gbcs/			

- Old data is packed in an **archive and copied to Tape**
- All **metadata** remains live and searchable
- Ultimately handled by ITS
- Taping cost apply (see ITS)



Archive_2014-07_2016-01

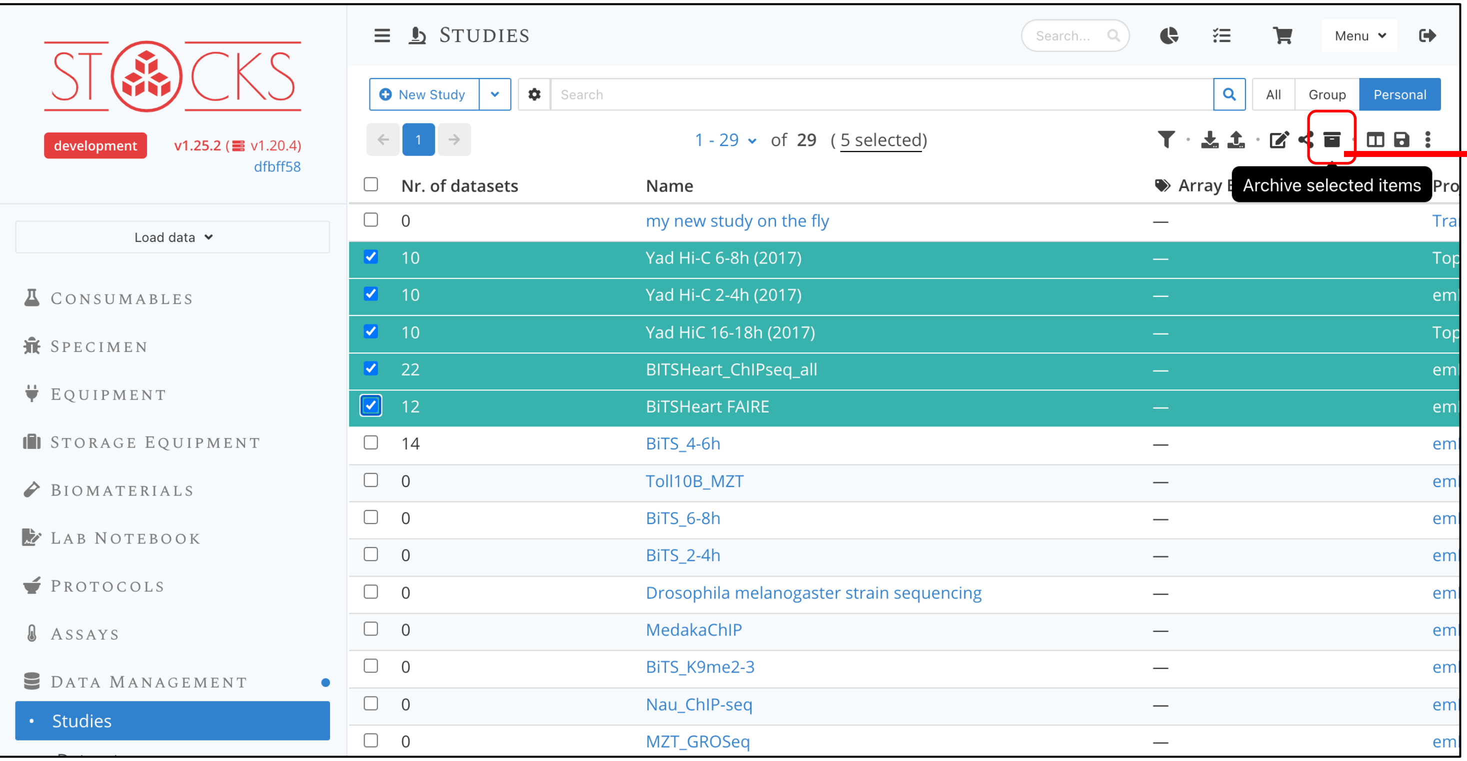
Datasets

☐	Name	Description	Type	QC	Samples	Parent datasets	Assay
☐	3T_cytRNA-Aniso-RII_lane1		paired-end fastq	NA	3T_cytRNA-Aniso-RII		3T_C2R8HA
☐	3T_cytRNA-Aniso-R-I_lane1		paired-end fastq	NA	3T_cytRNA-Aniso-R-I		3T_C2R8HA
☐	3T_cytRNA-DMSO-RIII_lane1		paired-end fastq	NA	3T_cytRNA-DMSO-RIII		3T_C2R8HA
☐	3T_cytRNA-DMSO-R-II_lane1		paired-end fastq	NA	3T_cytRNA-DMSO-R-II		3T_C2R8HA
☐	3T_cytRNA-DMSO-R-I_lane1		paired-end fastq	NA	3T_cytRNA-DMSO-R-I		3T_C2R8HA
☐	AM-GFP_lane8		paired-end fastq	NA	AM-GFP		C4VKHACX
☐	AM-parental-1_lane8		paired-end fastq	NA	AM-parental-1		C4VKHACX
☐	AM-parental-2_lane8		paired-end fastq	NA	AM-parental-2		C4VKHACX
☐	AM-parental-3_lane8		paired-end fastq	NA	AM-parental-3		C4VKHACX

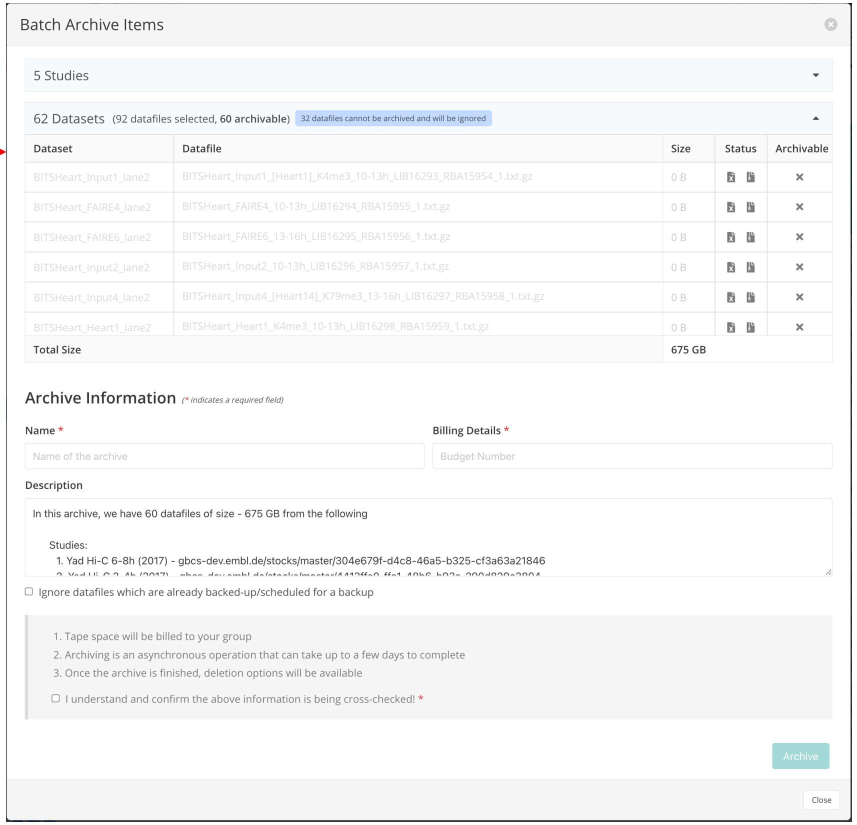
Datafile copies

☐	URI	Assays	Short name	Peek	Datafile Version
☐	/filerarchive/gbcs/embase_archive_46-biohentze-Archive_201...	MOV10_C258FACXX_8	C258FACXX_iCLIPMOV10_	—	1: C258FACXX_iCLIPMOV10_13s00
☐	/filerarchive/gbcs/embase_archive_46-biohentze-Archive_201...	hnRNPC_C281TACXX_8	C281TACXX_iCLIPhnRNPC	—	1: C281TACXX_iCLIPhnRNPC_13s0

Smart Dataset Archiving



	Nr. of datasets	Name	Array E	Pro
☐	0	my new study on the fly	—	Tra
☑	10	Yad Hi-C 6-8h (2017)	—	Top
☑	10	Yad Hi-C 2-4h (2017)	—	em
☑	10	Yad HiC 16-18h (2017)	—	Top
☑	22	BITSHeart_ChIPseq_all	—	em
☑	12	BiTSHeart FAIRE	—	em
☐	14	BiTS_4-6h	—	em
☐	0	Toll10B_MZT	—	em
☐	0	BiTS_6-8h	—	em
☐	0	BiTS_2-4h	—	em
☐	0	Drosophila melanogaster strain sequencing	—	em
☐	0	MedakaChIP	—	em
☐	0	BiTS_K9me2-3	—	em
☐	0	Nau_ChIP-seq	—	em
☐	0	MZT_GROSeq	—	em



62 Datasets (92 datafiles selected, **60 archivable**) 32 datafiles cannot be archived and will be ignored

Dataset	Datafile	Size	Status	Archivable
BITSHeart_Input1_lane2	BITSHeart_Input1_[Heart1]_K4me3_10-13h_LIB16293_RBA15954_1.txt.gz	0 B		×
BITSHeart_FAIRE4_lane2	BITSHeart_FAIRE4_10-13h_LIB16294_RBA15955_1.txt.gz	0 B		×
BITSHeart_FAIRE6_lane2	BITSHeart_FAIRE6_13-16h_LIB16295_RBA15956_1.txt.gz	0 B		×
BITSHeart_Input2_lane2	BITSHeart_Input2_10-13h_LIB16296_RBA15957_1.txt.gz	0 B		×
BITSHeart_Input4_lane2	BITSHeart_Input4_[Heart14]_K79me3_13-16h_LIB16297_RBA15958_1.txt.gz	0 B		×
BITSHeart_Heart1_lane2	BITSHeart_Heart1_K4me3_10-13h_LIB16298_RBA15959_1.txt.gz	0 B		×
Total Size		**675 GB**		

Archive Information *(* indicates a required field)*

Name *: Name of the archive

Billing Details *: Budget Number

Description:

In this archive, we have 60 datafiles of size - 675 GB from the following

Studies:
1. Yad Hi-C 6-8h (2017) - gbcs-dev.embl.de/stocks/master/304e679f-d4c8-46a5-b325-cf3a63a21846

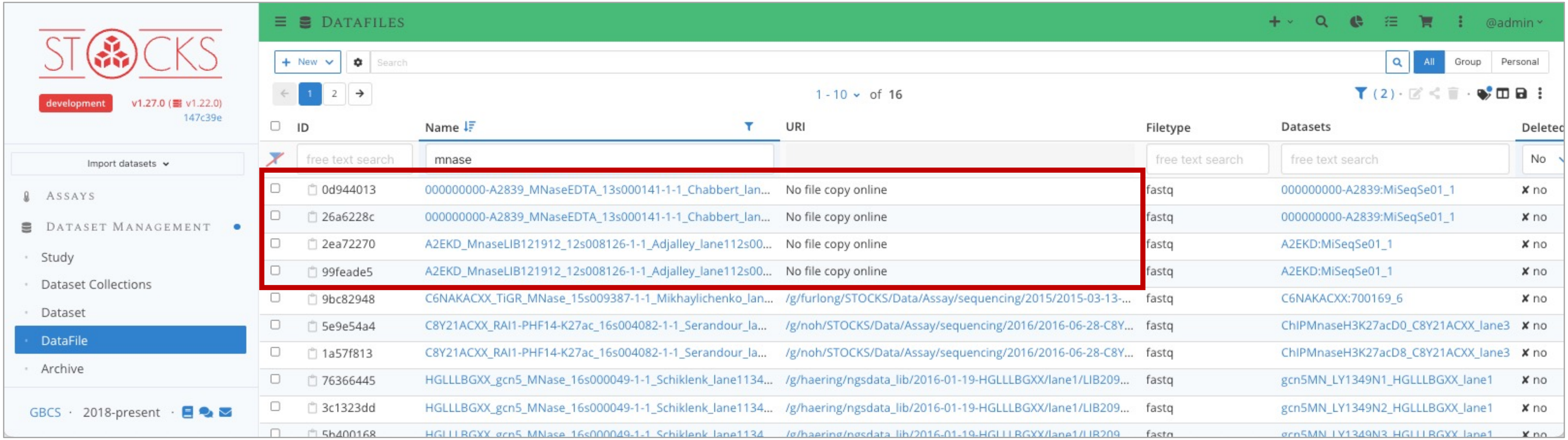
☐ Ignore datafiles which are already backed-up/scheduled for a backup

1. Tape space will be billed to your group
2. Archiving is an asynchronous operation that can take up to a few days to complete
3. Once the archive is finished, deletion options will be available

☐ I understand and confirm the above information is being cross-checked! *

- ➢ Send Datasets to tape by **selecting Studies or Projects**
- ➢ Datasets have **a new copy** on tape
- ➢ Datasets can be deleted from disks **after archiving**
  - This removes the main copy of the datasets
  - The **tape copy is immutable** and cannot be deleted; kept for 10 years
- ➢ Already archived can be skipped

Datafiles are always listed, even when offline



☐	ID	Name	URI	Filetype	Datasets	Deleted
	free text search	mnase		free text search	free text search	No
☐	0d944013	000000000-A2839_MNaseEDTA_13s000141-1-1_Chabbert_lan...	No file copy online	fastq	000000000-A2839:MiSeqSe01_1	✗ no
☐	26a6228c	000000000-A2839_MNaseEDTA_13s000141-1-1_Chabbert_lan...	No file copy online	fastq	000000000-A2839:MiSeqSe01_1	✗ no
☐	2ea72270	A2EKD_MnaseLIB121912_12s008126-1-1_Adjalley_lane112s00...	No file copy online	fastq	A2EKD:MiSeqSe01_1	✗ no
☐	99feade5	A2EKD_MnaseLIB121912_12s008126-1-1_Adjalley_lane112s00...	No file copy online	fastq	A2EKD:MiSeqSe01_1	✗ no
☐	9bc82948	C6NAKACXX_TiGR_MNase_15s009387-1-1_Mikhaylichenko_lan...	/g/furlong/STOCKS/Data/Assay/sequencing/2015/2015-03-13-...	fastq	C6NAKACXX:700169_6	✗ no
☐	5e9e54a4	C8Y21ACXX_RAI1-PHF14-K27ac_16s004082-1-1_Serandour_la...	/g/noh/STOCKS/Data/Assay/sequencing/2016/2016-06-28-C8Y...	fastq	ChIPMnaseH3K27acD0_C8Y21ACXX_lane3	✗ no
☐	1a57f813	C8Y21ACXX_RAI1-PHF14-K27ac_16s004082-1-1_Serandour_la...	/g/noh/STOCKS/Data/Assay/sequencing/2016/2016-06-28-C8Y...	fastq	ChIPMnaseH3K27acD8_C8Y21ACXX_lane3	✗ no
☐	76366445	HGLLLBGXX_gcn5_MNase_16s000049-1-1_Schiklenk_lane1134...	/g/haering/ngsdata_lib/2016-01-19-HGLLLBGXX/lane1/LIB209...	fastq	gcn5MN_LY1349N1_HGLLLBGXX_lane1	✗ no
☐	3c1323dd	HGLLLBGXX_gcn5_MNase_16s000049-1-1_Schiklenk_lane1134...	/g/haering/ngsdata_lib/2016-01-19-HGLLLBGXX/lane1/LIB209...	fastq	gcn5MN_LY1349N2_HGLLLBGXX_lane1	✗ no
☐	5b400168	HGLLLBGXX_gcn5_MNase_16s000049-1-1_Schiklenk_lane1134	/g/haering/ngsdata_lib/2016-01-19-HGLLLBGXX/lane1/LIB209	fastq	gcn5MN_LY1349N3_HGLLLBGXX_lane1	✗ no

Preparing data for public repository submission

- Data submission == Study submission
- Make sure to have a study containing all the datasets to submit
  - Give a good name and description (will be public)
  - Set the design terms; which also hint at necessary sample annotations
- Annotate your samples according to the study design
- Make sure all samples’ relevant fields are filled
- Link protocols to your samples (all the way up to specimen)
  - Make sure each protocol has a **summary** (no format, few sentences)
  - Growth protocol is always mandatory
  - Extraction protocol is always mandatory for omics
  - NGS Library preparation is always mandatory for sequencing
- One ready get in touch with GBCS
- Get in touch <u>a few weeks </u>before you want to submit your paper
  - Get in touch at project start if you use human samples

Data Submission as done today

Usual submission steps (after initial discussion):

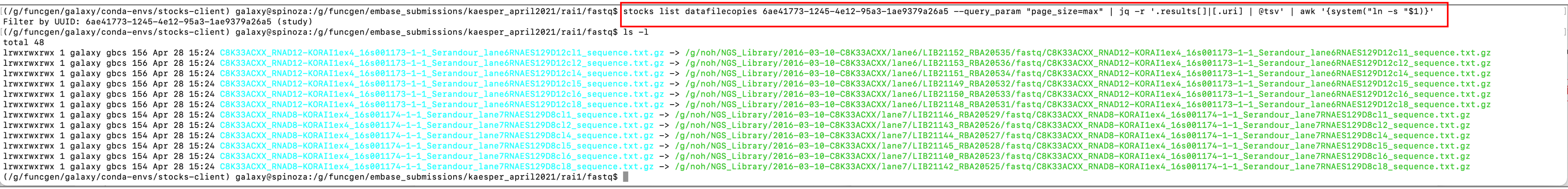
- Annotate samples
- Curate samples (e.g. merging redundant samples)
- Linking protocols
- *Visualization of the complete Study Graph*

STOCKS UI

- Export MAGETAB
    - $ stocks arrayexpress-export export.xlsx <studyid>
- Transfer Files
    - Create local links to files
    - FTP *or* Aspera
- Email the MAGE-TAB to annotare

STOCKS CLI (uses REST API)

https://git.embl.de/grp-gbcs/stocks-python-client

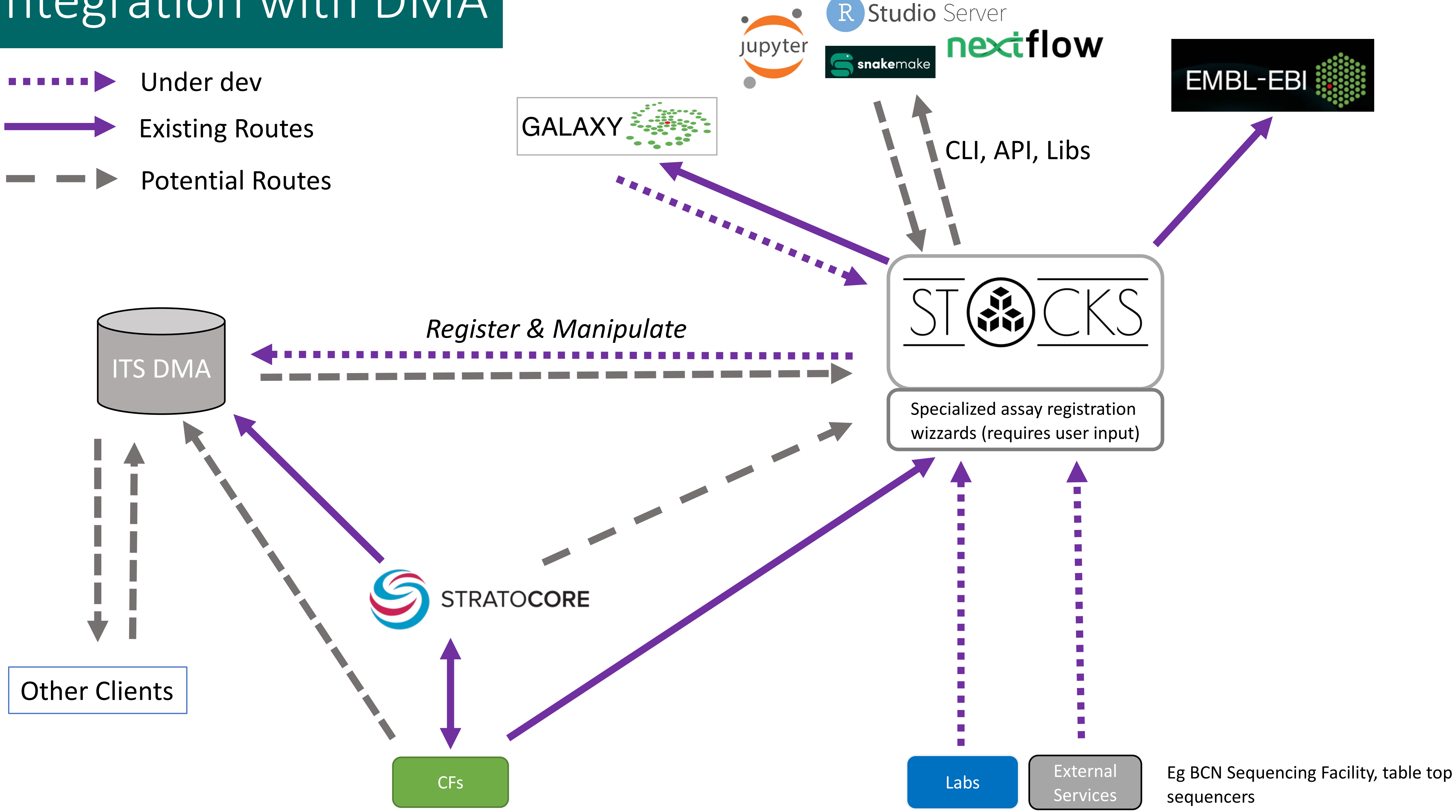


```
(/g/funcgen/galaxy/conda-envs/stocks-client) galaxy@spinoza:/g/funcgen/embase_submissions/kaesper_april2021/rai1/fastq$ stocks list datafilecopies 6ae41773-1245-4e12-95a3-1ae9379a26a5 --query_param "page_size=max" | jq -r '.results[]|[.uri] | @tsv' | awk '{system("ln -s "$1)}'
Filter by UUID: 6ae41773-1245-4e12-95a3-1ae9379a26a5 (study)
(/g/funcgen/galaxy/conda-envs/stocks-client) galaxy@spinoza:/g/funcgen/embase_submissions/kaesper_april2021/rai1/fastq$ ls -l
total 48
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl1_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21152_RBA20535/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl1_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl2_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21153_RBA20536/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl2_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl4_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21151_RBA20534/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl4_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl5_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21149_RBA20532/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl5_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl6_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21150_RBA20533/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl6_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 156 Apr 28 15:24 C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl8_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane6/LIB21148_RBA20531/fastq/C8K33ACXX_RNAD12-KORAI1ex4_16s001173-1-1_Serandour_lane6RNAES129D12cl8_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl1_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21146_RBA20529/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl1_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl2_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21143_RBA20526/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl2_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl4_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21144_RBA20527/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl4_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl5_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21145_RBA20528/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl5_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl6_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21140_RBA20523/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl6_sequence.txt.gz
lrwxrwxrwx 1 galaxy gbcs 154 Apr 28 15:24 C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl8_sequence.txt.gz -> /g/noh/NGS_Library/2016-03-10-C8K33ACXX/lane7/LIB21142_RBA20525/fastq/C8K33ACXX_RNAD8-KORAI1ex4_16s001174-1-1_Serandour_lane7RNAES129D8cl8_sequence.txt.gz
(/g/funcgen/galaxy/conda-envs/stocks-client) galaxy@spinoza:/g/funcgen/embase_submissions/kaesper_april2021/rai1/fastq$
```

The STOCKS REST API enables interoperability

- STOCKS REST API available
  - allow integration with other tools e.g. Galaxy and LOOL
  - automate data registration e.g. GC Bridge
  - STOCKS GUI communicates with the server only using the API
    - Write your own client !
- CLI available
  - also uses the REST API
  - extract study information into MAGE-TAB
- Integration with the ITS DMA (soon):
  - Datasets created in STOCKS are pushed to DMA
  - Will also allow import from DMA

Integration with DMA



Under dev

Existing Routes

Potential Routes

GALAXY

jupyter

RStudio Server

snakemake

nextflow

EMBL-EBI

CLI, API, Libs

STOCKS

Specialized assay registration wizzards (requires user input)

Register & Manipulate

ITS DMA

STRATOCORE

Other Clients

CFs

Labs

External Services

Eg BCN Sequencing Facility, table top sequencers

Thank you for your attention

GBCS

Jelle Scholtalbers

Matthias Monfort

Nayeem Reza

Questions ?