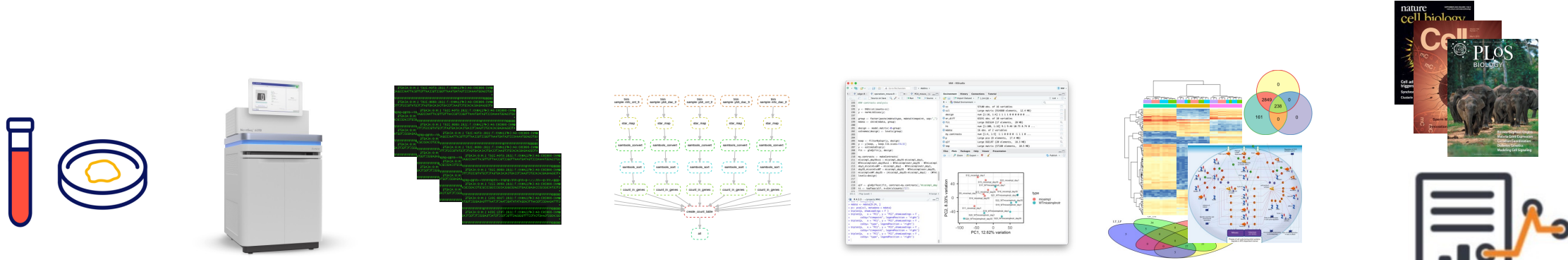


NGSVivo a case of a system for core labs

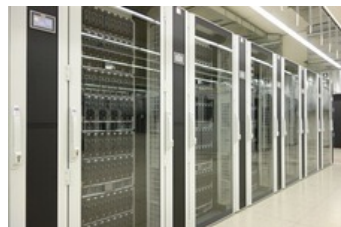
Vincenzo Capece, Michal Okoniewski



Bioinformatic research value chain



F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



RRP



data stewardship



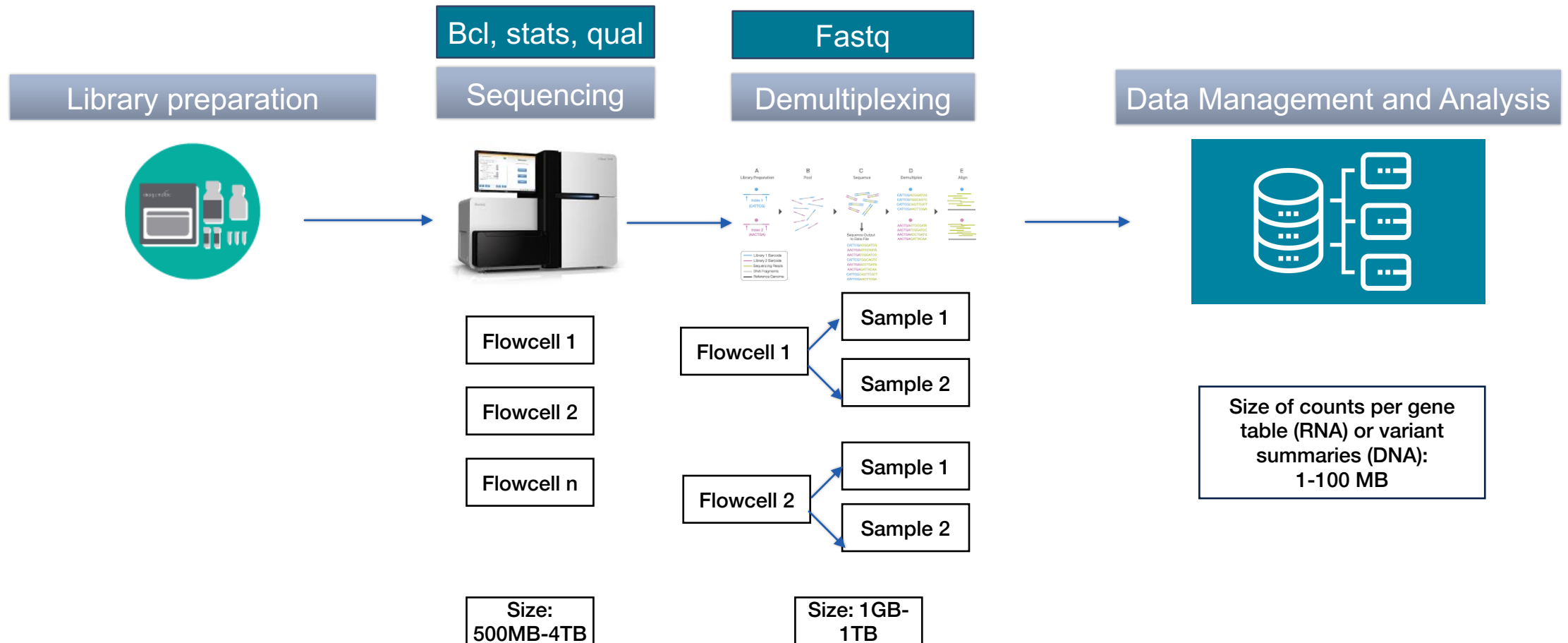
NGSVivo – a case of a system for core labs

1. Next-gen sequencing core facility of ETH D-BSSE in Basel, established 2008
2. Head: Christian Beisel
3. 4 employees
4. 4 sequencers (Illumina, PacBio)
5. 292 groups, 1300 users

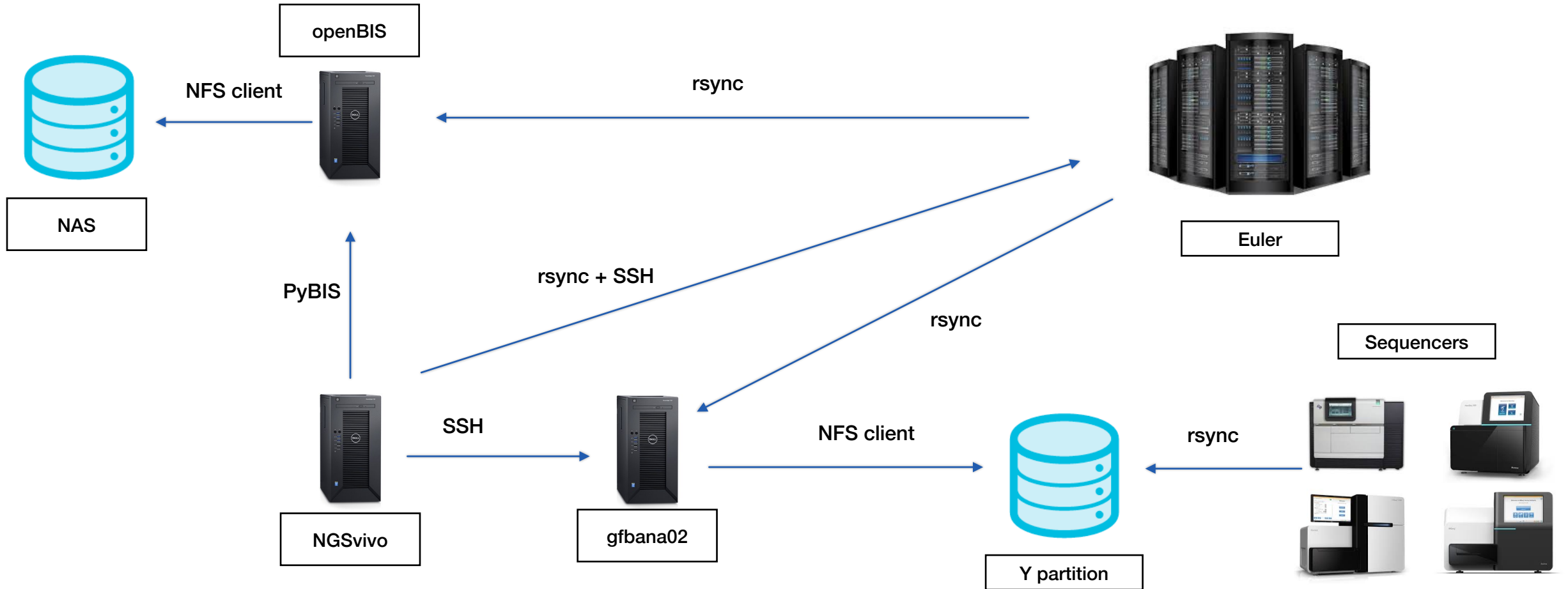


- Expensive technology
 - 1.e.g. sequencer can cost 1M+ CHF, 1 flow cell run 5k CHF minimum
- Large amount of data: (archive: 177 TB, openbis: 110 TB in 2022)

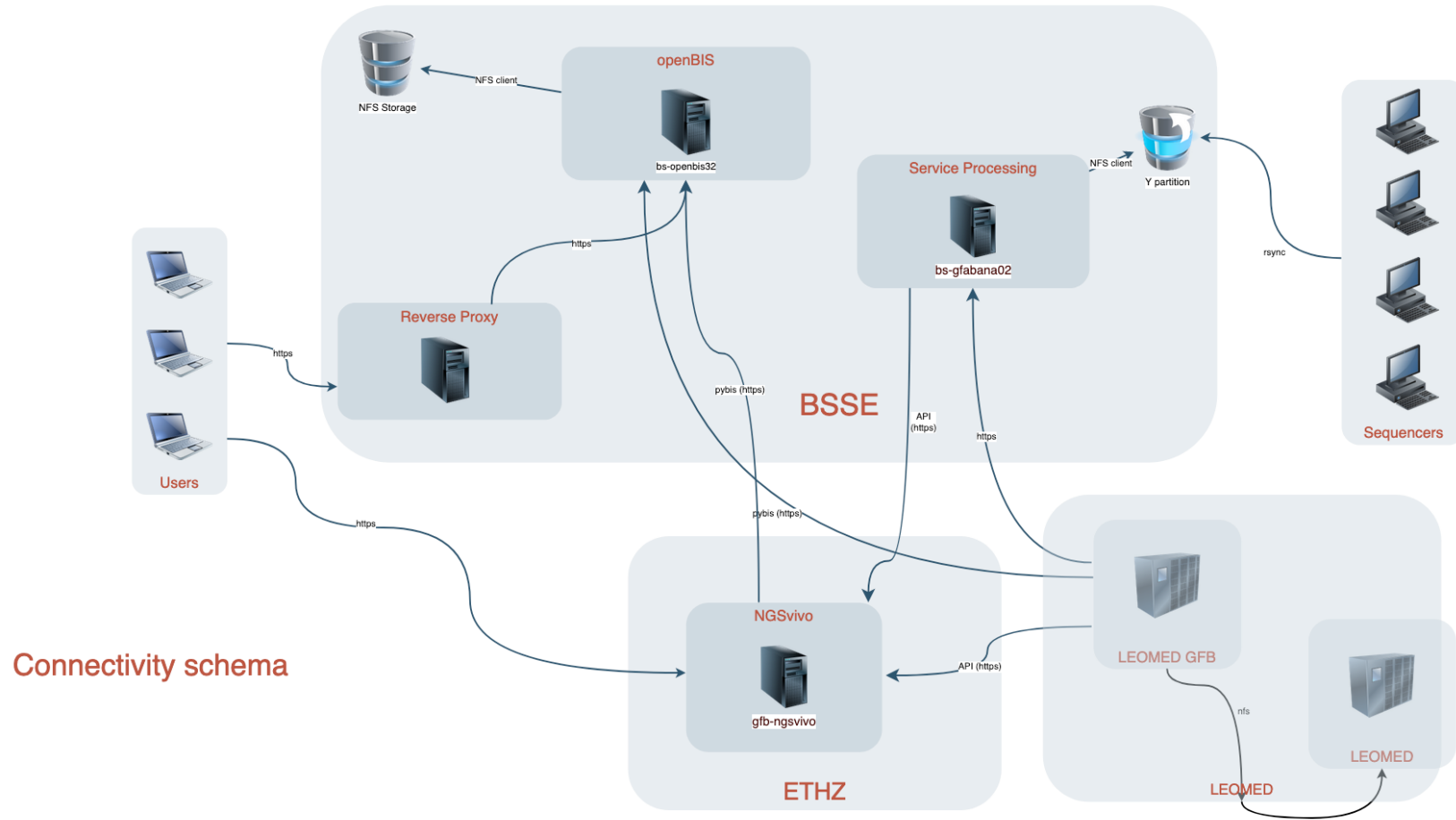
NGS data production workflow



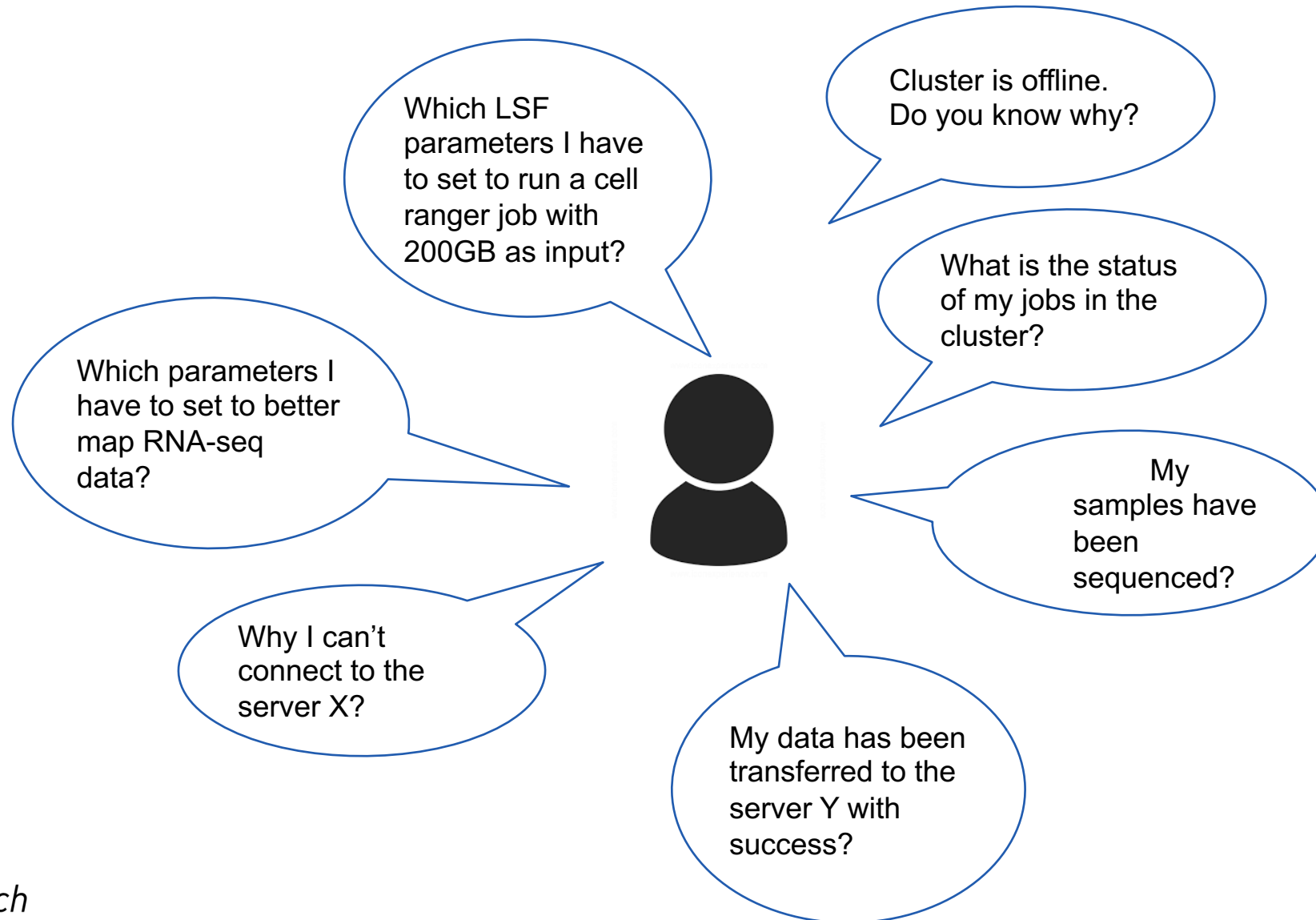
IT infrastructure schema



IT infrastructure schema



NGSVivo from the user perspective



NGSVivo user interface

Ngsvivo administration

NGSVIVO		
Analysis	+ Add	Change
Cluster types	+ Add	Change
Clusters	+ Add	Change
Demultiplexings	+ Add	Change
Groups	+ Add	Change
Job queues	+ Add	Change
Job steps	+ Add	Change
Kits	+ Add	Change
Machine activities	+ Add	Change
Machine types	+ Add	Change
Machines	+ Add	Change
Pipeline step_softwares	+ Add	Change
Pipeline steps	+ Add	Change
Pipelines	+ Add	Change
Queues	+ Add	Change
Reference genome version software versions	+ Add	Change
Reference genome versions	+ Add	Change
Reference genomes	+ Add	Change
Send datas	+ Add	Change
Sequencers	+ Add	Change
Set parameters	+ Add	Change
Software versions	+ Add	Change
Softwares	+ Add	Change
Systems	+ Add	Change
Timeline datas	+ Add	Change
Timelines	+ Add	Change
Todos	+ Add	Change
Transfer datas	+ Add	Change



</> Home

- Lab
- Infrastructure
- HowTo

Contact support

... The demultiplexing of the flowcell 231122_A00730_0898_BHNWKHDRX3 finished with success!

View in openbis
Send data to users
Invoice the samples
Restart demultiplexing

Nov. 23, 2023, 7:30 a.m. 29 minutes ago

... The demultiplexing of the flowcell 231122_NS500318_0004_AHHF7HBGXV finished with success!

View in openbis
Send data to users
Invoice the samples
Restart demultiplexing

Nov. 23, 2023, 4 a.m. 3 hours, 59 minutes ago

... The demultiplexing of the flowcell 231122_NS500318_0004_AHHF7HBGXV started with success.

You can check its running in the right panel in the following sections:

4_HHF7HBGXV_1_ILLUMINA

Nov. 23, 2023, 3:22 a.m. 4 hours, 38 minutes ago

... The demultiplexing of the flowcell 231122_A00730_0898_BHNWKHDRX3 started with success.

You can check its running in the right panel in the following sections:

898_HNWKHDRX3_1_ILLUMINA
898_HNWKHDRX3_2_ILLUMINA

Nov. 23, 2023, 1:36 a.m. 6 hours, 23 minutes ago

capecev

Jobs

- 4_HHF7HBGXV_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 898_HNWKHDRX3_2_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 898_HNWKHDRX3_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 779_000000000-DMKGF_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 1_HNKN2AFX5_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 2_HHF3CBGXV_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 3_HH5JCBGXV_1_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 897_HLCTWDSX7_4_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset
- 897_HLCTWDSX7_3_ILLUMINA
 - ✓ demultiplexing
 - ✓ FastQC per sample
 - ✓ MultiQC on the dataset

NGSVivo user interface

The screenshot displays the NGSvivo user interface. At the top left, the 'NGSvivo' logo is visible, and the user 'ti14264' is logged in. A navigation sidebar on the left contains 'Home', 'Lab', 'Infrastructure', and 'HowTo'. The main area shows a job log with five entries. The first four are green and indicate successful data transfer for flowcell 000000000-KFFGF of lane 1 of the group BSSE_REDDY. The fifth entry is red and shows a terminal log snippet where a command failed due to a delay in the quality control dataset, with a note that the administrator is working to solve it. The right sidebar, titled 'JOBS STATUS', lists several jobs with their progress: '673_000000000-DHMYV_1_ILLUMINA' (running...), '1131_HY5HCBGXK_1_ILLUMINA' (completed), '1134_H3G2LAFX5_1_ILLUMINA' (completed), and '712_HKHFFDRX2_2_ILLUMINA' (demultiplexing is in run). Other jobs are partially visible.

NGSvivo ti14264

<> Home

- Lab
- Infrastructure
- HowTo

📌 Data of the flowcell 000000000-KFFGF of lane 1 of the group BSSE_REDDY sent with success
Oct. 24, 2022, 11:58 a.m. 6 days, 21 hours ago

📌 Data of the flowcell 000000000-KFFGF of lane 1 of the group BSSE_REDDY sent with success
Oct. 24, 2022, 11:58 a.m. 6 days, 21 hours ago

📌 Data of the flowcell 000000000-KFFGF of lane 1 of the group BSSE_REDDY sent with success
Oct. 24, 2022, 11:58 a.m. 6 days, 21 hours ago

```
... The running of the command module load multiqc/1.9 && cd /cluster/work/gfb/ngsvivo/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA && multiqc -o /cluster/work/gfb/ngsvivo/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/qc_dataset_652_000000000-KFFGF_1_ILLUMINA -n 652_000000000-KFFGF_1_ILLUMINA_qc /cluster/work/gfb/ngsvivo/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/GFB-202_000000000-KFFGF_1_ILLUMINA/GFB-197_000000000-KFFGF_1_ILLUMINA/GFB-197_000000000-KFFGF_1_ILLUMINA/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/GFB-198_000000000-KFFGF_1_ILLUMINA/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/GFB-204_000000000-KFFGF_1_ILLUMINA/GFB-199_000000000-KFFGF_1_ILLUMINA/GFB-201_000000000-KFFGF_1_ILLUMINA/GFB-205_000000000-KFFGF_1_ILLUMINA/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/GFB-206_000000000-KFFGF_1_ILLUMINA/GFB-200_000000000-KFFGF_1_ILLUMINA/GFB-195_000000000-KFFGF_1_ILLUMINA/GFB-196_000000000-KFFGF_1_ILLUMINA/GFB-196_000000000-KFFGF_1_ILLUMINA/run/000000000-KFFGF/652_000000000-KFFGF_1_ILLUMINA/GFB-203_000000000-KFFGF_1_ILLUMINA failed and it can cause a delay to the quality control dataset of the data of 652_000000000-KFFGF_1_ILLUMINA  
The administrator is working to solve it.  
Sept. 29, 2022, 12:23 p.m. 1 month ago
```

📌 Data of the flowcell 000000000-KFFGF of lane 1 of the group BSSE_REDDY sent with success
Sept. 9, 2022, 12:27 p.m. 1 month, 3 weeks ago

JOBS STATUS

- 673_000000000-DHMYV_1_ILLUMINA
running...
✔demultiplexing
🚫quality_control_sample
quality_control_dataset
- 1131_HY5HCBGXK_1_ILLUMINA
✔demultiplexing
✔quality_control_sample
✔quality_control_dataset
- 1134_H3G2LAFX5_1_ILLUMINA
✔demultiplexing
✔quality_control_sample
✔quality_control_dataset
- 712_HKHFFDRX2_1_ILLUMINA
demultiplexing is in run
🔄demultiplexing
quality_control_sample
quality_control_dataset
- 712_HKHFFDRX2_2_ILLUMINA
demultiplexing is in run
🔄demultiplexing
quality_control_sample
quality_control_dataset
- 711_HKG3JDRX2_1_10X_SINGLI
BIS...
✔demultiplexing
✔quality_control_sample
✔quality_control_dataset
- 711_HKG3JDRX2_1_10X_SINGLI
BIS...
✔demultiplexing
✔quality_control_sample
✔quality_control_dataset
- 711_HKG3JDRX2_1_10X_SINGLI
BIS...
✔demultiplexing
✔quality_control_sample
✔quality_control_dataset
- 711_HKG3JDRX2_1_ILLUMINA
pic

Data management backbone: openBIS

The screenshot displays the openBIS web application interface. The browser address bar shows the URL: `openbis-gfb.ethz.ch/openbis/webapp/elc-lims/?menuUniqueId={"type":"SAMPLE","id":"20240429121202524-94332"}&viewName=showViewSamplePageFromPe...`

Global Search and **Filter** options are available at the top left.

Pool: POOL-601

Buttons: **+ New**, **Edit**, **Upload**, **More ...**

Navigation: **1-2 of 2**, **Rows per page: 10**, **COLUMNS**, **FILTERS**, **EXPORTS**

Code	Name	Permid	Identifier	Phix(%)	Concentration in lane(pM)	Total PF clusters	Assigned reads	Total raw reads
Operator: AND Filter								
H3WTMDRX5.1		20240429163513532-94889	/LAB/H3WTMDRX5.1	0.5	220	488076265	463547029	638
H3WTMDRX5.2		20240429163514634-94890	/LAB/H3WTMDRX5.2			479943347	454546178	638

Children

Navigation: **1-20 of 24**, **Rows per page: 20**, **COLUMNS**, **FILTERS**, **EXPORTS**

Code	Name	Permid	Identifier	Internal contact	Name of Contact Person	Email of Contact Person	Contact Data Manager Email
Operator: AND Filter							
GFB-30037		20240429121202524-94333	/BSSE_FUSSENEGGER/GFB-30037	MF	Dr. Mohamed Mahameed	mmahameed@ethz.ch	
GFB-		20240429121202524-	/BSSE_FUSSENEGGER/GFB-		Dr.		

Filter sidebar (left):

- P1803_JINBO_RNA
- P1977_ANA_PALMA_TEIXEIRA
- P2109_ADRIAN
- P2739_MOHAMED
 - Objects
 - POOL-601**
 - Bsse Gfb
 - Bsse Held
 - Bsse Hierlemann
 - Bsse Illumina Kits
 - Bsse Moor
 - Bsse Mueller

QC sidebar (bottom left):

- QC : 20240430034229237-94944
 - qc_dataset_POOL-601
 - 955_H3WTMDRX5_1_ILLUMINA_qc_data
 - 955_H3WTMDRX5_2_ILLUMINA_qc_data
 - 955_H3WTMDRX5_1_ILLUMINA_qc.html (1.4 Mb)**
 - 955_H3WTMDRX5_2_ILLUMINA_qc.html (1.4 Mb)
- SEQUENCING : 20240430035219646-94945

Automated tasks in NGSVivo

- Start demultiplexing;
- Release output;
- Wrong barcodes;
- Wrong mismatches;
- Machines not reachable;
- No space in the machines;
- openBIS failed dataset injection;
- Few number of reads per sample;
- Compute stats over lane/pool/samples;
- Users can see the status of their jobs;
- ...



NGSvivo notification