High Throughput Sequencing Facility

02/12/2022

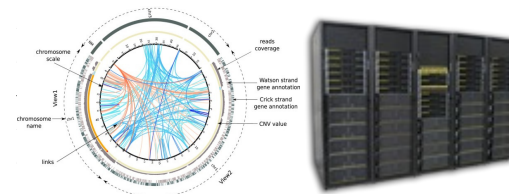# High Throughput Sequencing Facility

Molecular biology

Sequencing

Bioinformatics

- Created at the CGM in February 2010 (IMAGIF)
- I2BC since 2015
- Open to both academic and industrial users, whether in France or abroad

IBiSA. Infrastructures en Biologie Santé et Agronomie

FRANCE GENOMIQUE

LRQA CERTIFIED
ISO 9001 · NFX 50-900

Services in molecular biology

Molecular biology	Sequencing	Bioinformatics

Illumina library preparation

- DNA-seq, ChIP-seq
- RNA-seq
- Small RNA-seq, miRNA-seq

- Single cell sequencing (10X genomics)

ONT library preparation

- DNA-seq
- RNA-seq (direct, cDNA)
- tRNA sequencing

- Cas9 enrichment
- ULR

Custom-made libraries if needed

Sequencing platforms

Molecular biology	**Sequencing**	Bioinformatics

Versatile set of equipment:

Illumina NS 550

Illumina NS 2000

ONT GridION

ONT PromethION P2 solo
(spring 2023)

Bioinformatics services

Molecular biology	Sequencing	**Bioinformatics**

- Differential gene expression
- Genome assembly
- Genome annotation
- SNPs, indels, rearrangement detection

- DNA/RNA modifications (ONT)
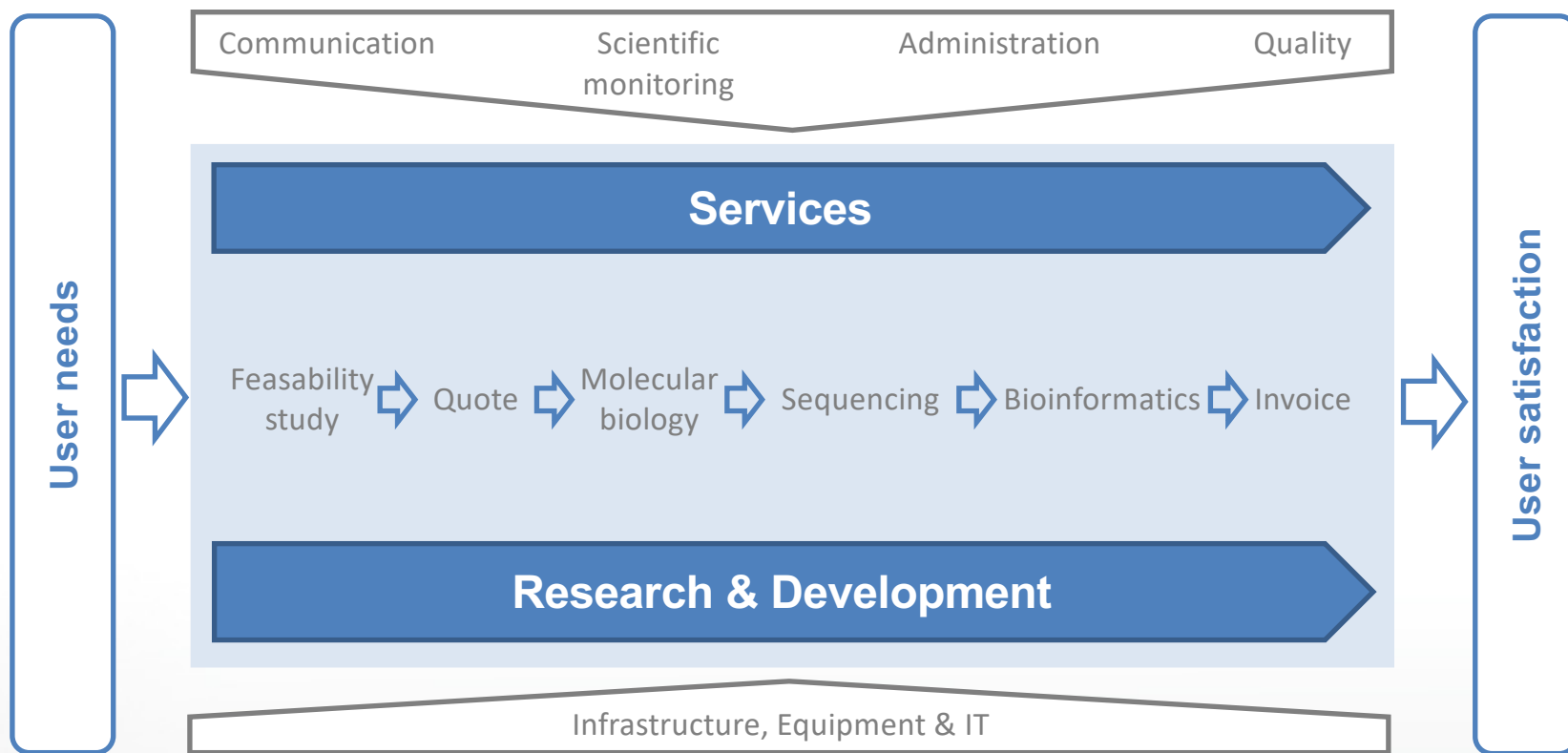- TE detection (ONT)
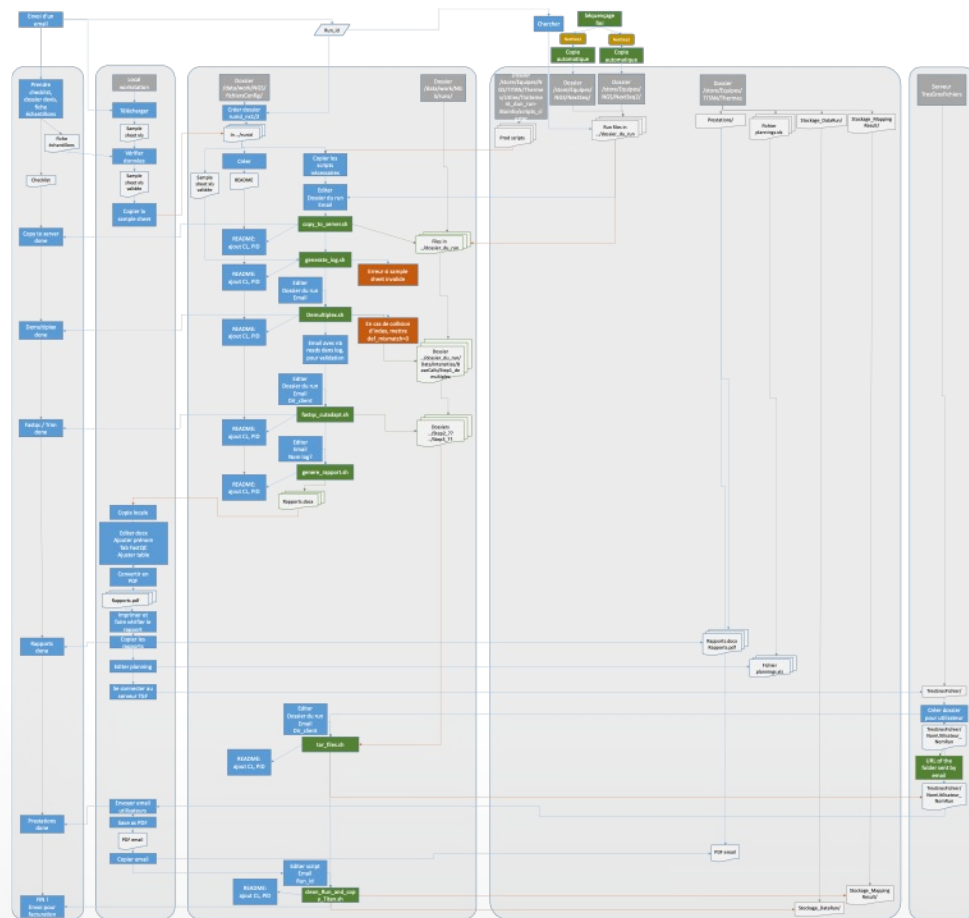
- **On-demand analyses**

I2BC cluster

GPU

« Theoretical » data flow

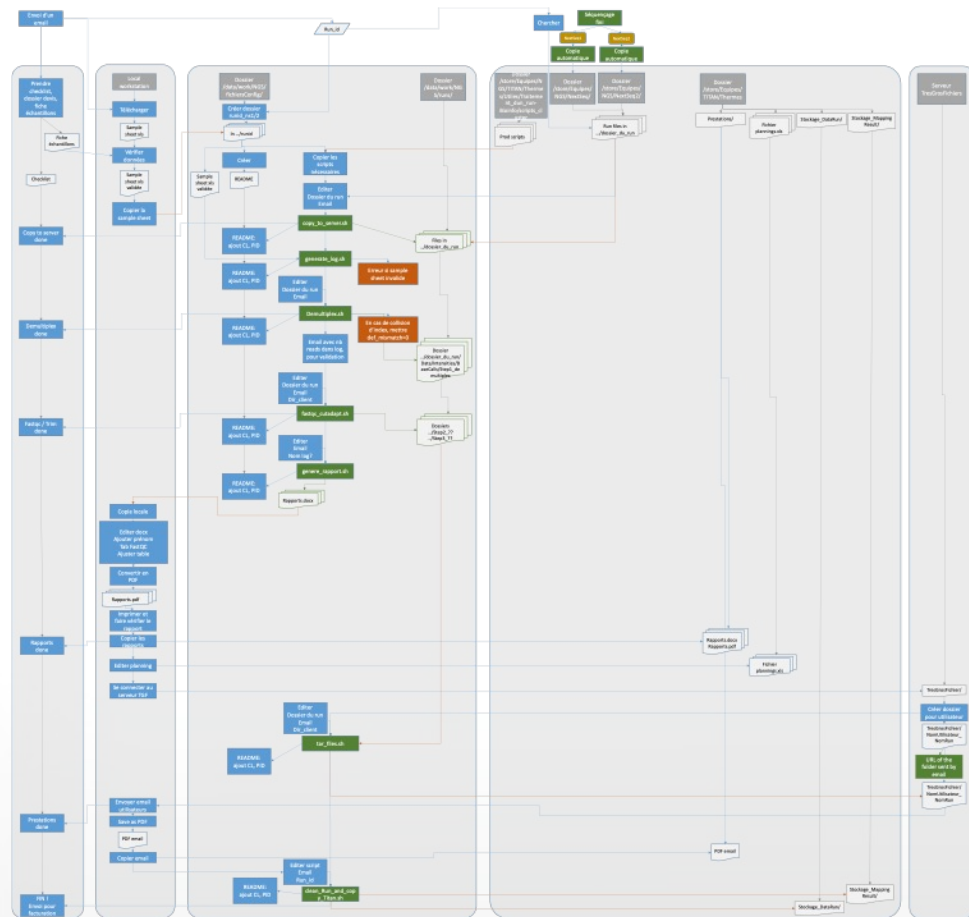# Real data flow

Real data flow

- (Who?)
- What?
- Where?
- When?
- How?

Who?

- Who is in charge of the current step?

What?

- What is being processed? What is the outcome?

Where?

- Where is the data, before/after the step?

When?

- When is that step performed?

How?

- How is the data processed?

User needs

"Produce new data!"

Feasability study

User satisfaction

User needs

"Produce new data!"

Metadata !

Feasability study

User metadata
- Request form
- Number of samples
- Species, …

User satisfaction

User needs

"Produce new data!"

Metadata !

User metadata
- Request form
- Number of samples
- Species, …

Feasability study

PF's metadata   => Traceability
- Date of arrival
- Person receiving the samples…

User satisfaction

User needs

"Produce new data!"

Metadata !

Feasability study

Molecular biology

User metadata
• Request form
• Number of samples
• Species, …

PF's metadata   => Traceability
• Date of arrival
• Person receiving the samples…

• Protocol, kit
• Person treating the samples at each step, date,…

User satisfaction

User needs

"Produce new data!"

User satisfaction

Metadata !

Feasability study

Molecular biology

Data production
(ultra raw data)

Secured storage => Risk management
(automatic, local, short/medium term)

User metadata
- Request form
- Number of samples
- Species, …

PF's metadata   => Traceability
- Date of arrival
- Person receiving the samples…

- Protocol, kit
- Person treating the samples at each step, date,…

- Sequencer, FC, person, date

User needs

"Produce new data!"

User satisfaction

Metadata !

Feasability study

Molecular biology

Data production
(ultra raw data)

Primary treatment

Raw data

User metadata
• Request form
• Number of samples
• Species, …

PF's metadata   => Traceability
• Date of arrival
• Person receiving the samples…

• Protocol, kit
• Person treating the samples at each step, date,…

• Sequencer, FC, person, date

Secured storage   => Risk management
(automatic, local, short/medium term)

• Software, version, CL => Reproducibility

Secured storage
(automatic, long term)

User needs

"Produce new data!"

Metadata !

User metadata
- Request form
- Number of samples
- Species, …

Feasability study

PF's metadata   => Traceability
- Date of arrival
- Person receiving the samples…

Molecular biology

- Protocol, kit
- Person treating the samples at each step, date,…

Data production
(ultra raw data)

- Sequencer, FC, person, date

Secured storage => Risk management
(automatic, local, short/medium term)

User satisfaction

Shared storage
(automatic, very short term)

Primary treatment

- Software, version, CL => Reproducibility

Metadata
(reports)

Raw data

Secured storage
(automatic, long term)

User needs

"Produce new data!"

Metadata !

User metadata
- Request form
- Number of samples
- Species, …

Feasability study

PF's metadata   => Traceability
- Date of arrival
- Person receiving the samples…

Molecular biology

- Protocol, kit
- Person treating the samples at each step, date,…

Data production
(ultra raw data)

- Sequencer, FC, person, date

Secured storage   => Risk management
(automatic, local, short/medium term)

User satisfaction

Shared storage
(automatic, very short term)

Primary treatment

- Software, version, CL => Reproducibility

Secured storage
(automatic, long term)

Metadata
(reports)

Raw data

Remote copy
(automatic, short term)

- Software, version, CL

Bioinformatic analysis

Analysis output data

Secured storage
(automatic, medium term)

User needs

"Produce new data!"

"Reuse!"

Deposit / share / publish

Structure data / metadata

Choose a repository

User satisfaction

Shared storage
(automatic, very short term)

=> Integrity

Metadata
(reports)

**Analysis
output data**

**Bioinformatic
analysis**

Raw data

**Primary
treatment**

Data production
(ultra raw data)

Molecular
biology

Feasability
study

Metadata !

User metadata
• Request form
• Number of samples
• Species, …

PF's metadata   => Traceability
• Date of arrival
• Person receiving the samples…

• Protocol, kit
• Person treating the samples at
  each step, date,…

• Sequencer, FC, person, date

Secured storage   => Risk management
(automatic, local, short/medium term)

• Software, version, CL => Reproducibility

Secured storage
(automatic, long term)

Remote copy
(automatic, short term)

• Software, version, CL

Secured storage
(automatic, medium term)

Difficulties

Support
* Paper vs numerical (meta)data

Storage
* Financial, ecological aspects
* Avoid redundancy (discussions with our institute)
* For how long ? (storage vs archive)
* Users still contact us for old data

Bioinformatic analysis
* Reports, integrity verification
* FAIR for bioinformatics: versioning system, Docker images…
* How to facilitate deposition on repositories?

Conclusions

(Meta)data producers

Conclusions

(Meta)data producers

Metadata collection…

- starts before data production
- starts before anything is done
- is (part of) our Quality system

Conclusions

(Meta)data producers

Metadata collection…

- starts before data production
- starts before anything is done
- is (part of) our Quality system

A data management plan is (part of) our Quality system

- Traceability, risk management, reproducibility…

Conclusions

(Meta)data producers

Metadata collection…
- starts before data production
- starts before anything is done
- is (part of) our Quality system

A data management plan is (part of) our Quality system
- Traceability, risk management, reproducibility…

Far from perfect, but improving
- FAIR at all levels (inside and outside the facility)
- Improve output reports and data structure

Conclusions

(Meta)data producers

Metadata collection…
- starts before data production
- starts before anything is done
- is (part of) our Quality system

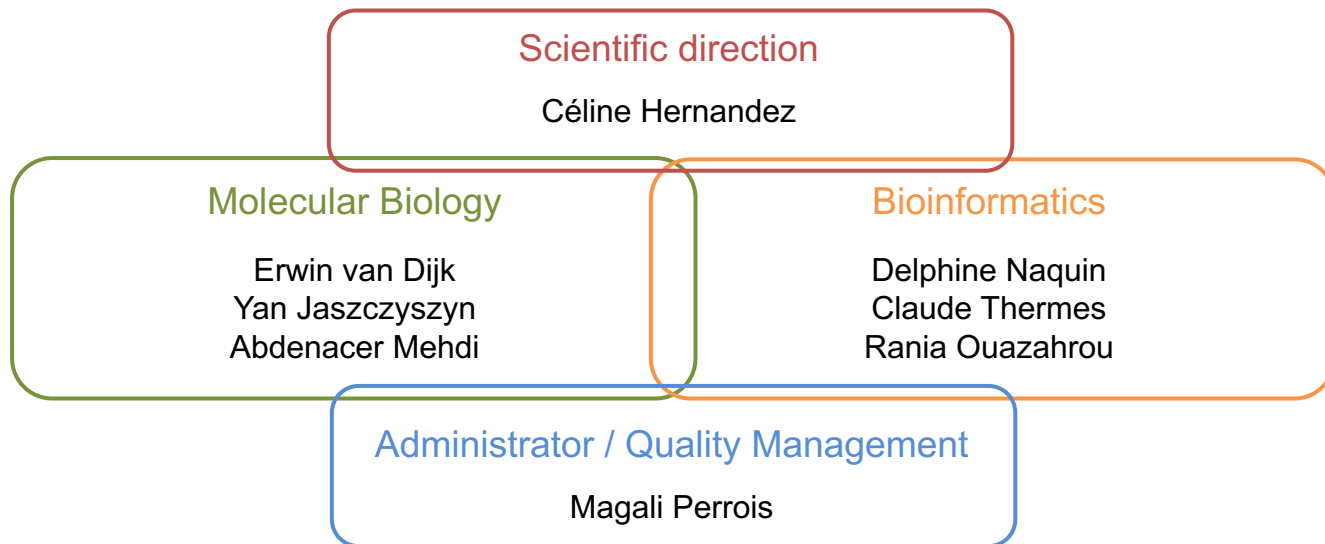A data management plan is (part of) our Quality system
- Traceability, risk management, reproducibility…

Far from perfect, but improving
- FAIR at all levels (inside and outside the facility)
- Improve output reports and data structure

"Data producer" or "Research output transformer"?

High Throughput Sequencing Facility

Scientific direction

Céline Hernandez

Molecular Biology

Erwin van Dijk
Yan Jaszczyszyn
Abdenacer Mehdi

Bioinformatics

Delphine Naquin
Claude Thermes
Rania Ouazahrou

Administrator / Quality Management

Magali Perrois

IBiSA • Infrastructures en Biologie Santé et Agronomie

LRQA CERTIFIED ISO 9001 · NFX 50-900

ISO 9001:2015
NFX 50-900:2016

FRANCE GENOMIQUE

I2BC-sequencage@i2bc.paris-saclay.fr