

Uso de Aspera-RedIRIS para la transferencia de datos genómicos en el CNAG-CRG y RD-Connect.

Marcos Fernández Callejo
Jornadas Técnicas de RedIRIS 2019
Sevilla, 29 de Mayo de 2019

cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG^{IB}
Centre
for Genomic
Regulation



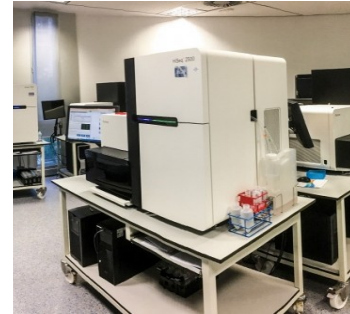
1. Introducción

Centro Nacional de Análisis Genómico (CNAG-CRG)

Misión: Desarrollar proyectos de análisis genómico con el objetivo de mejorar la salud y la calidad de vida de las personas, en colaboración con la comunidad médica e investigadora Catalana, Española, Europea e Internacional.

Infraestructura de laboratorio:

- 5 Instrumentos de secuenciación de lecturas cortas. (Illumina)
- 4 Instrumentos de secuenciación de lecturas largas. (Oxford Nanopore)



Infraestructura Computacional:

- 3.552 cores, 3.7 PB disco + 7.5 PB Cintas, 35.TB RAM, 200 Tflops

Certificaciones/Acreditaciones ISO

- Certificación ISO 9001 (2014)
- Acreditación 17025 (2016)

Entre los 5 primeros centros de secuenciación en Europa.



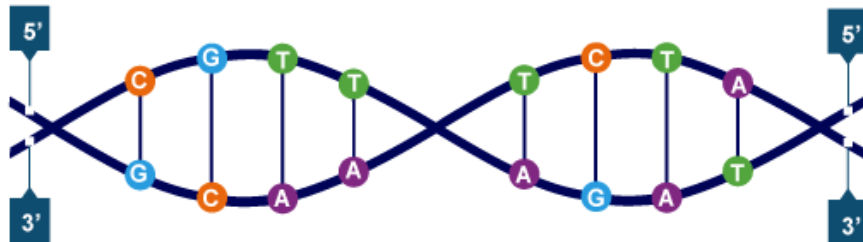
cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG[®]
Centre
for Genomic
Regulation

1. Introducción

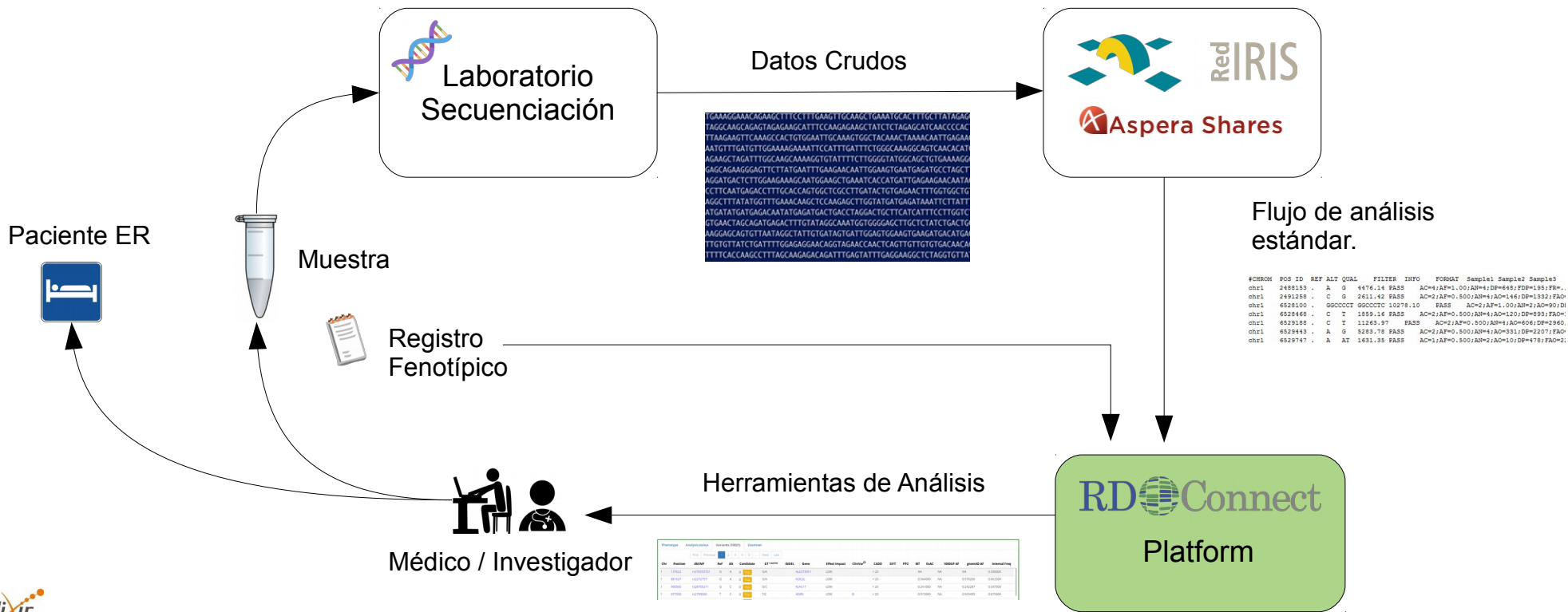
Enfermedades Raras y Genómica



Genoma Humano:

- 3,200,000,000 bp
- 2 x 23 cromosomas
- Cada uno de nosotros >3.000.000 variantes

2.RD-Connect Genome-Phenome Analysis Platform (GPAP)





3. Interfície RD-Connect GPAP

https://platform.rd-connect.eu

RD-Connect GENOMICS ABOUT WELCOME S.BELTRAN (PLATFORM V0.13.0, DATASET 20170426) FAQ LOGOUT

Filters ▾ PRESET FILTERS RESET SHARE ▶ RUN QUERY

Variant Type: coding high moderate **Population:** exac **SNV->MT:** A D **SNV->SIFT:** D **SNV->PP2:** D P

| Gene Name | Transcript ID | Effect Impact | Consequence | Feature Type | HGVS coding | Amino Acid change | Amino Acid length | Genotype Number | Exon Rank | CDS Position | Transcript BioType |
|-----------|----------------|---------------|------------------|--------------|-------------|-------------------|-------------------|-----------------|-----------|--------------|--------------------|
| MFAP2 | ENST0000037553 | MODERATE | missense_variant | transcript | c.313A>C | p.Thr105Pro | 183 | 1 | 7/9 | 313/552 | protein_coding |
| MFAP2 | ENST0000037553 | MODERATE | missense_variant | transcript | c.310A>C | p.Thr104Pro | 182 | 1 | 6/8 | 310/549 | protein_coding |
| MFAP2 | ENST0000043854 | MODERATE | missense_variant | transcript | c.310A>C | p.Thr104Pro | 182 | 1 | 7/9 | 310/549 | protein_coding |






Variants (11) Exomiser

First Previous 1 Next Last EXPORT ALL

| Chr | Pos | dbSNP | Ref | Alt | Candidate | GT ^{E000010} | GT ^{E000036} | GT ^{E000037} | INDEL | Gene Name | Effect Impact | ClinVar | CADD | SIFT | PP2 | MT | ExAC | 1000GP AF |
|-----|-----------|-------|-----|-----|-----------|-----------------------|-----------------------|-----------------------|-------|-----------|---------------|---------|------|------|-----|----|------|-----------|
| 1 | 17302199 | . | T | G | ADD | T/G | T/T | T/T | | MFAP2 | MODERATE | | 26.9 | D | D | D | NA | 0 |
| 11 | 93535027 | . | C | A | ADD | C/A | C/C | C/C | | MED17 | MODERATE | | 34 | D | D | D | NA | 0 |
| 13 | 77835365 | . | G | A | ADD | G/A | G/G | G/G | | MYCBP2 | MODERATE | | 25.6 | T | D | D | NA | 0 |
| 13 | 110437802 | . | A | C | ADD | A/C | A/A | A/A | | IRS2 | MODERATE | | 24.6 | D | D | D | NA | 0 |

4. Recepción Datos Genómicos

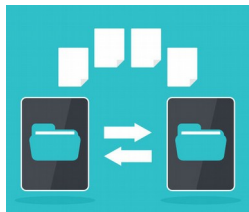
4.1 Información descriptiva de los datos (Metadatos)

- Pacientes 
- Fenotipo 
- Tipo de Secuenciación 
- Tejido 
- Tipo de Fichero 
- ...

4.2 Envío de Datos (Ficheros Genómicos)



- Herramienta para compartir contenido en forma de ficheros y directorios.
- Acceso seguro y transferencia rápida.
- Servidor **Aspera**. High Speed Transfer Server.



5. Envío de Metadatos

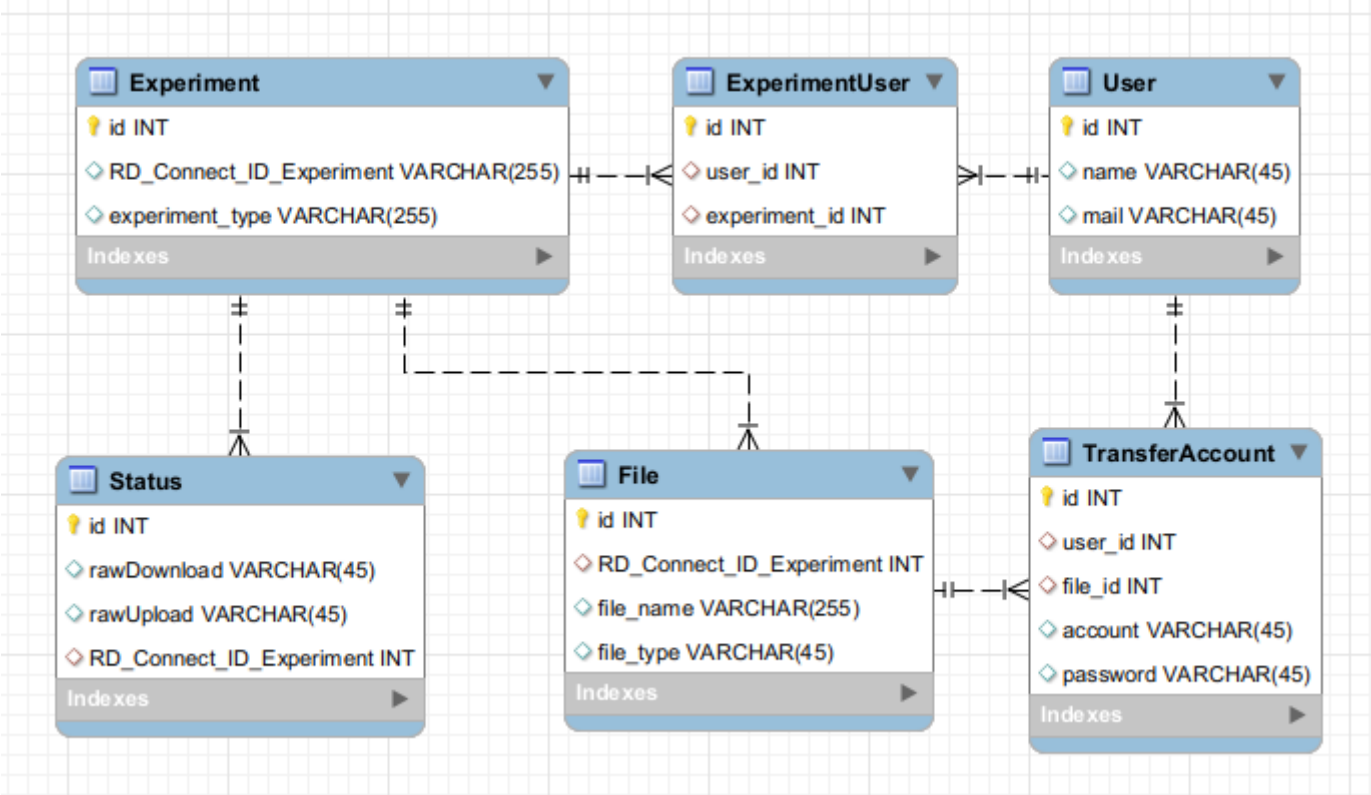
RD-Connect SUBMISSION PLATFORM

Submission Data to RD-Connect [See instructions](#)

- 1- Submit Participant Set
- 2- Submit Experiment Set
- 3- Upload data

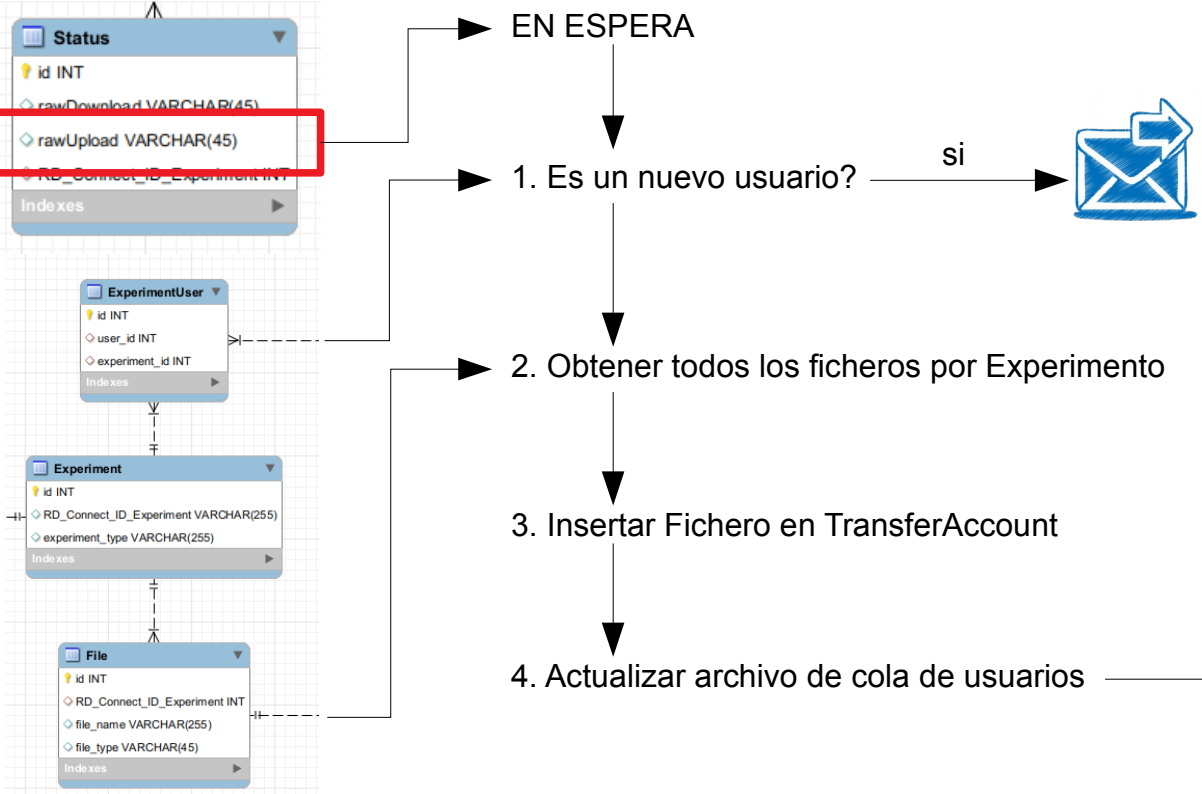
Manage your RD-Connect

- Participant
- Experiment
- Experiment Statistics
- Uploaded Files
- Experiment Analysis Status



6. Automatización Recepción de Datos.

6.1 Recepción metadatos nuevos pacientes

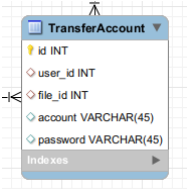


Archivo de cola de usuarios.

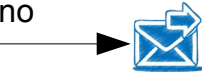
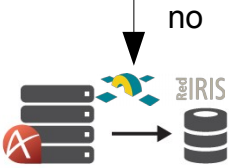
```
{
  "submission_user_1": {
    "user": "submission_user_1",
    "email": "submission_user_1@mail",
    "mail_wait": false,
    "files": [ "file_1.fastq.gz", ... ],
    "experiments": [ "experiment_1", ... ]
  },
  ...
  "submission_user_N": {
    "user": "submission_user_N",
    "email": "submission_user_N@mail",
    "mail_wait": false,
    "files": [ "file_1.fastq.gz", ... ],
    "experiments": [ "experiment_1", ... ]
  }
}
```


6. Automatización Recepción de Datos.

6.2 Asignación Cuentas Red Iris Aspera Shares



1. La cuenta fue ya asignada?



No hay cuentas disponibles

2. Generar Password

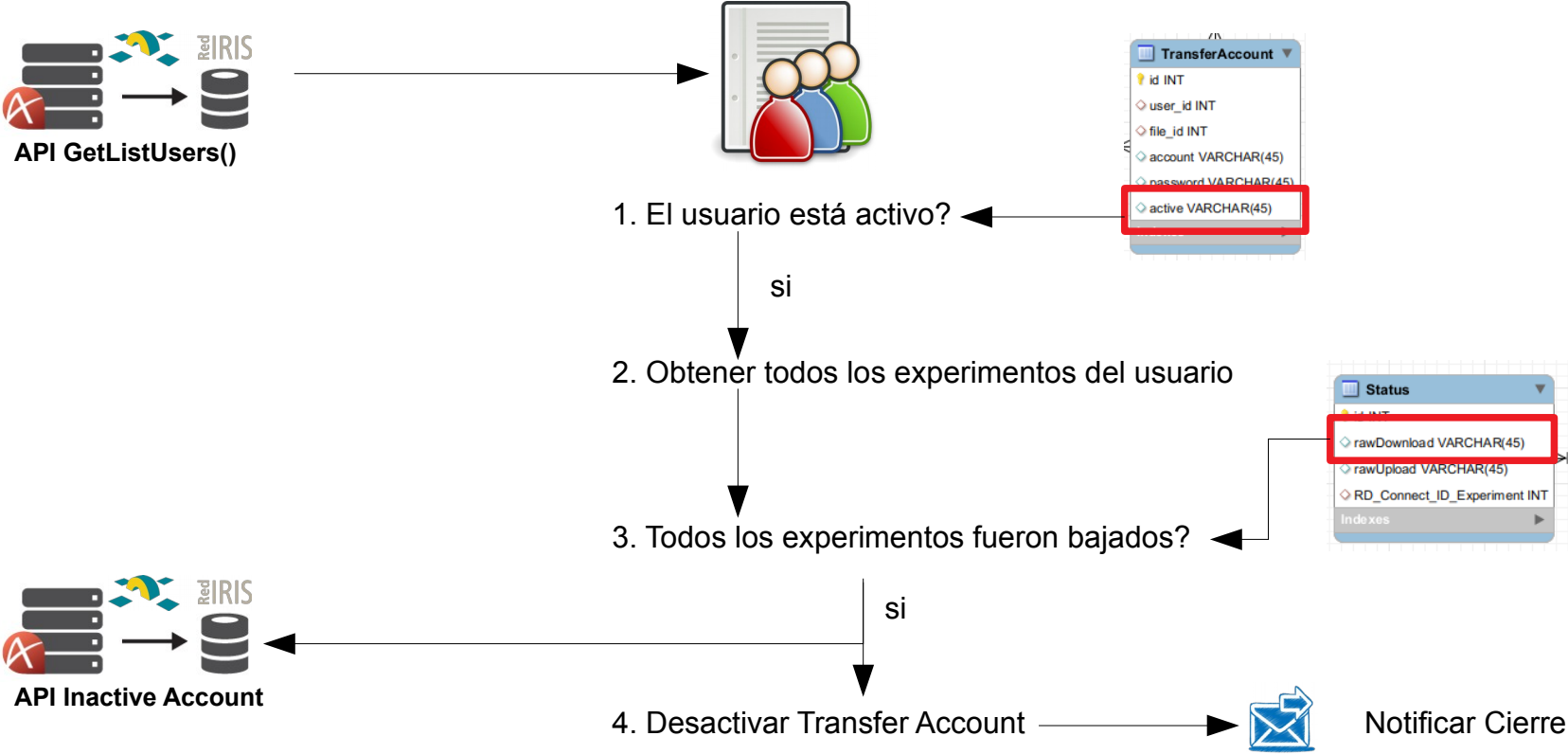


3. Actualizar tabla Transfer Account



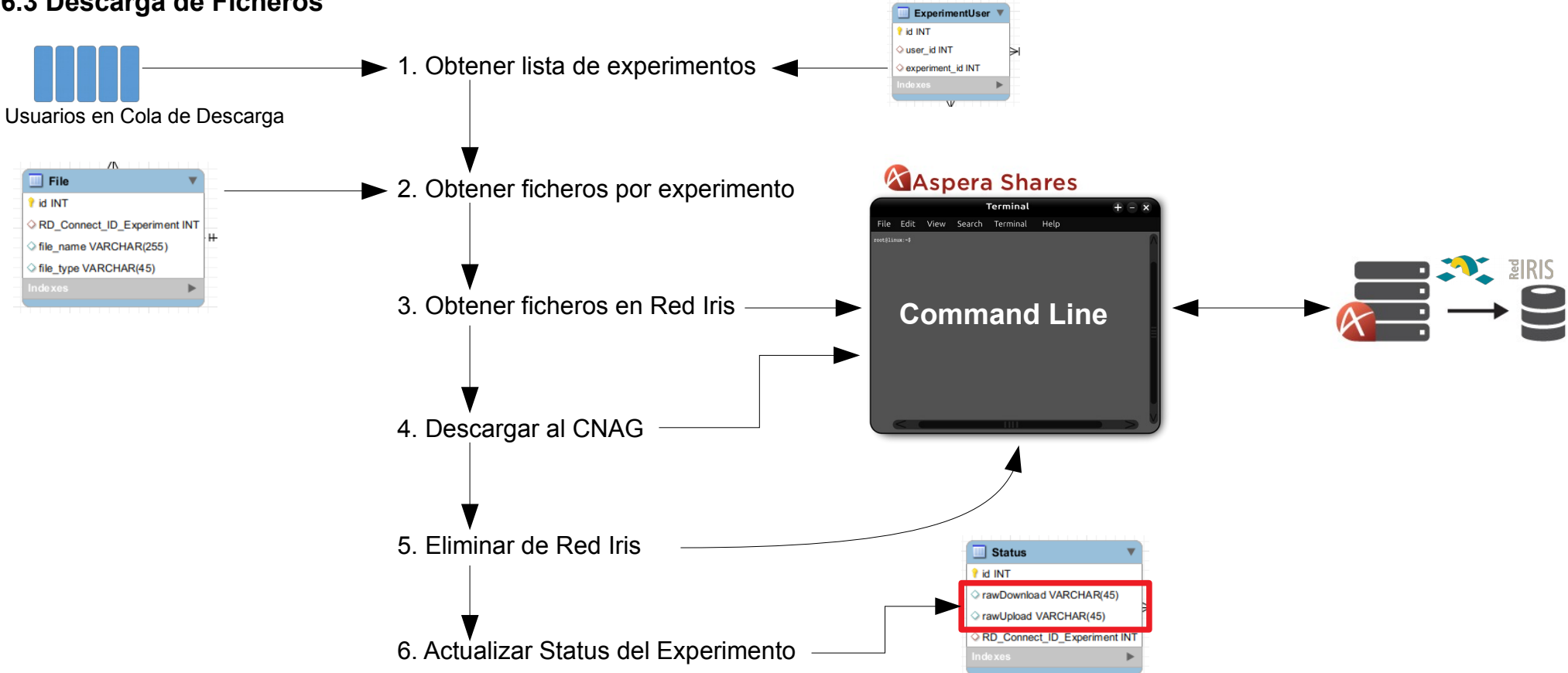
6. Automatización Recepción de Datos.

6.2 Desactivar Cuentas



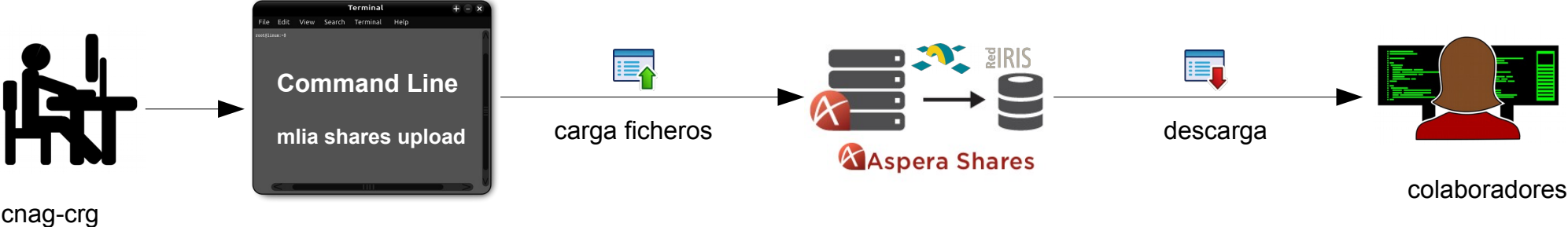
6. Automatización Recepción de Datos.

6.3 Descarga de Ficheros



7. Otros usos de Aspera Shares

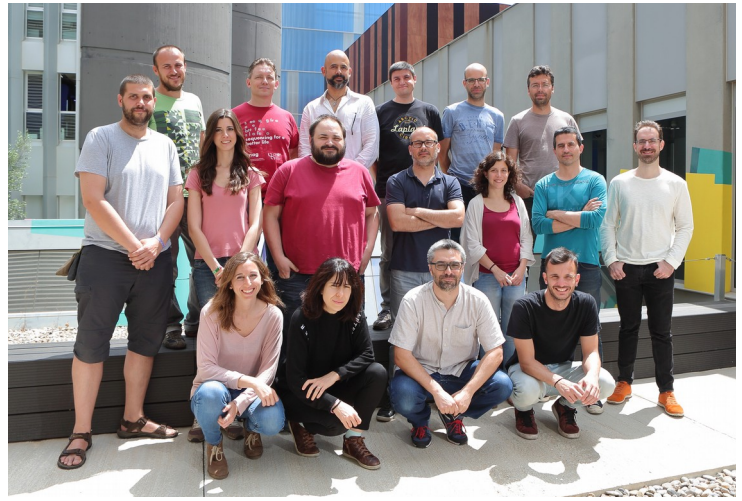
7.1 Transferir Análisis Genómicos a colaboradores del CNAG-CRG



7. Agradecimientos



Bioinformatics Analysis Unit



- Sergi Beltrán**
Davide Piscia
Lesley Matalonga
Steven Laurie
Daniel Picó
Anastasio Papakonstantinou
Cristina Luengo
Carles García
Marcos Fernández
Raul Tonda
Jean Remi Trotta
Genis Parra
Gemma Bullich
Ricky Joshi
Jordi Morata
Matt Ingham
Jordi Camps
Raul Alcántara
Eloi Casals

Antonio Fuentes Bermejo

