Individualized PaediatricCure

iPC Data Catalogue
Hands-On Training Session

iPC Data Catalogue : *Hands-On Training Session*

If data is broadly available, experts can create effective personalised therapies for kids with cancer.

Let's join forces

in creating a massive knowledge base for a joint fight against paediatric cancer.

Welcome to the Webinar!

Housekeeping

- **The webinar is being recorded.**

- **The session and material will be available on the iPC website.**

- **Raise hand / use chat.**

Agenda

- About iPC
  Lisa Burgstaller
  Technikon

- iPC Data Catalogue
  Alejandro Canosa
  Barcelona Supercomputing Center (BSC)

- Moving Forward
  Jolanda Modic
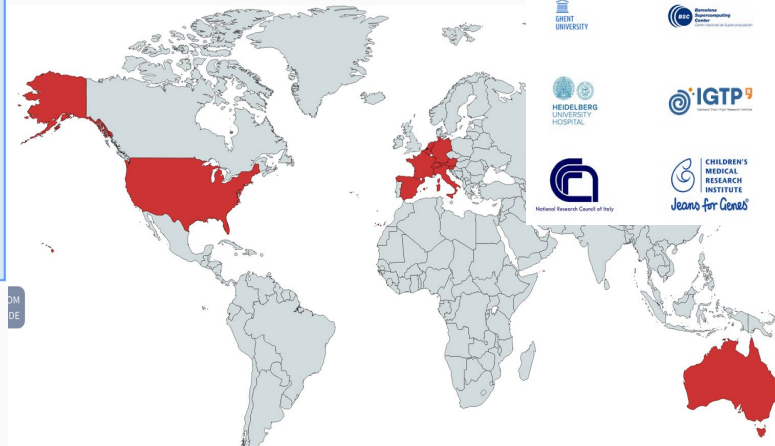 **XLAB**

About iPC

Lisa Burgstaller
(Technikon)

iPC Consortium - 21 Partners from 3 Continents

iPC in a Nutshell

- **iPC assembles, standardizes and harmonizes high-quality** multi-disciplinary data **on paediatric cancers that are made accessible to researchers, clinicians and stakeholders through a cloud-based platform.**

- **iPC leverages collected data to construct** predictive models **for paediatric cancer patients, tumours and treatments.**

- **iPC verifies the quality of predictions using dedicated data from clinical trials and preclinical models.**

- **iPC builds a** cloud-based service **with an HPC backend to produce** personalized treatment recommendations **for paediatric cancer patients.**

Discover iPC

https://ipc-project.eu/
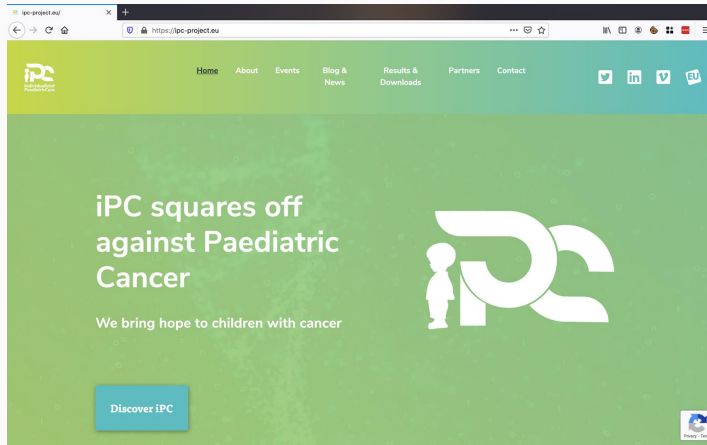
🐦 **@iPC_H2020**

💼 **/company/ipcproject/**

vimeo **/vimeo.com/technikon**

The Structure / Purpose of the Core Presentation

- Understanding the background
  What it is, how it works, how it fits into the iPC Platform?

- Seeing it in practice
  Live demo: How you can search, filter, select, and use data?

- Opportunities for collaboration
  How you can help us move forward?

iPC is here to help YOU advance YOUR research in paediatric cancer.
Let us know if you see opportunities for improvements.

iPC Data Catalogue

Alejandro Canosa
(Barcelona Supercomputing Center - BSC)

iPC Catalogue: Brief introduction

We aim to create a platform where scientists can visualize, search, filter, select, and analyze data related to paediatric cancer.

iPC Catalogue will host:
- ❖ **Newly produced data.**
- ❖ **Data from existing repositories hosting paediatric data (Kids-First, R2, dbGaP...)**
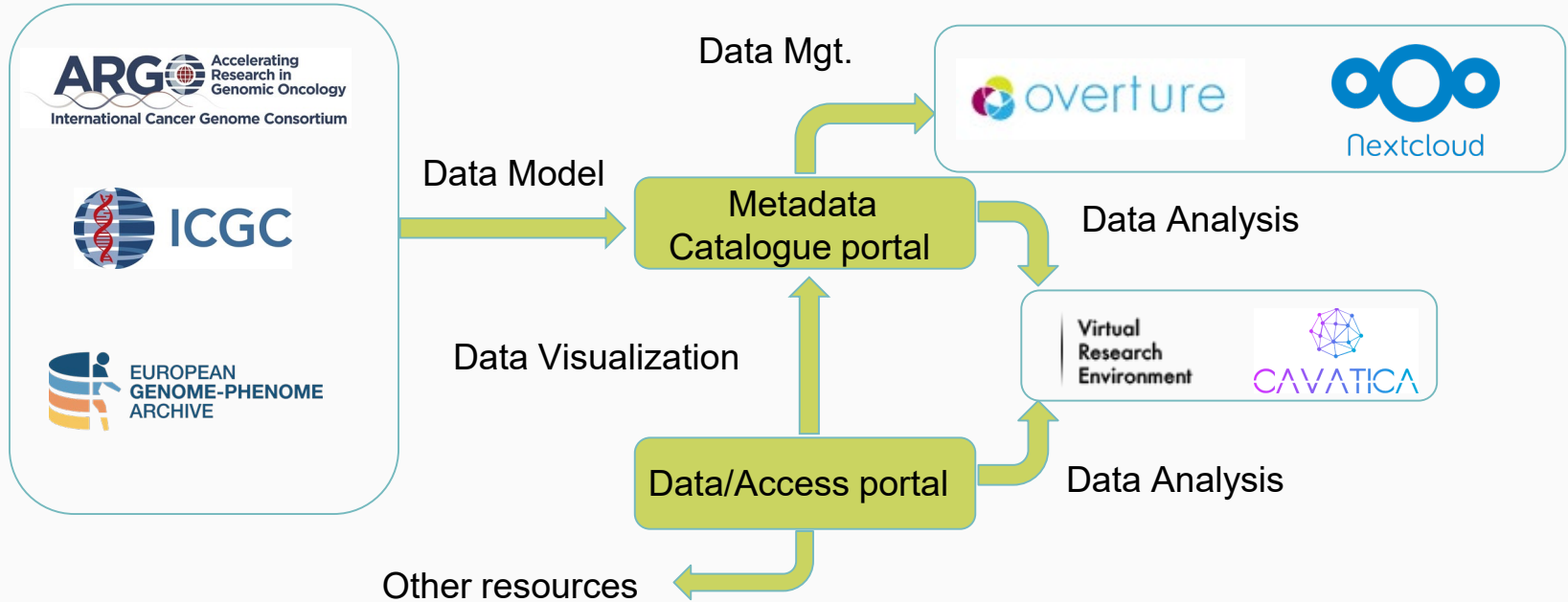
all this while…
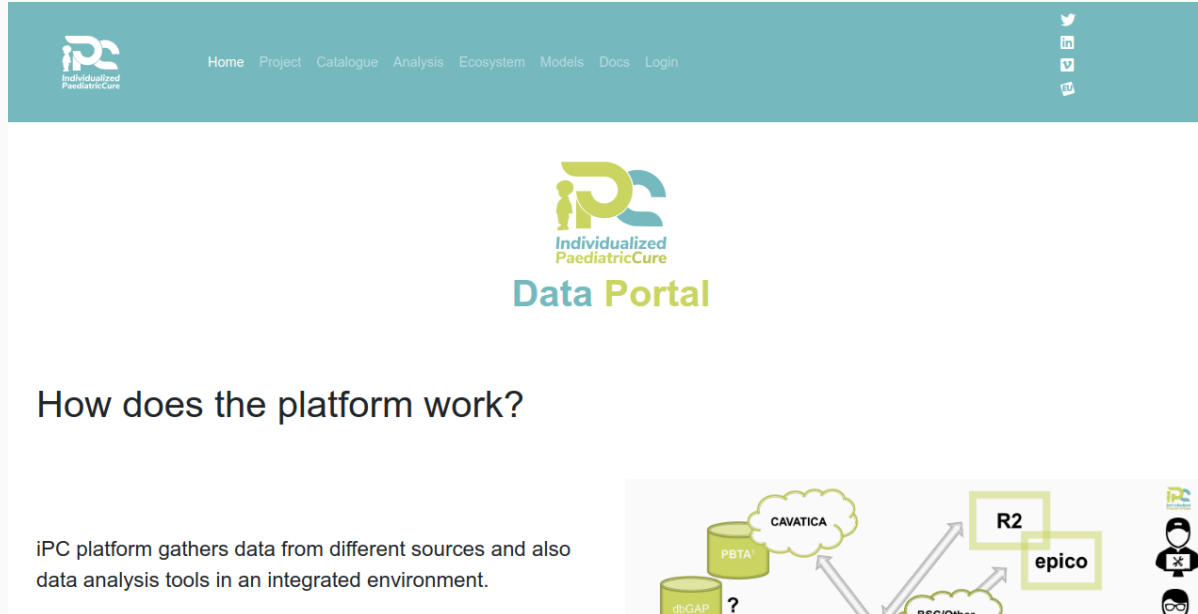**Data Access procedures and compliance to Data protection laws are simplified.**

Which are our main objectives (so far) ?

- ● **Development of a computational framework.**

- ● **Development of a central platform for data/metadata storage.**

iPC Platform architecture

iPC – individualizedPaediatricCure

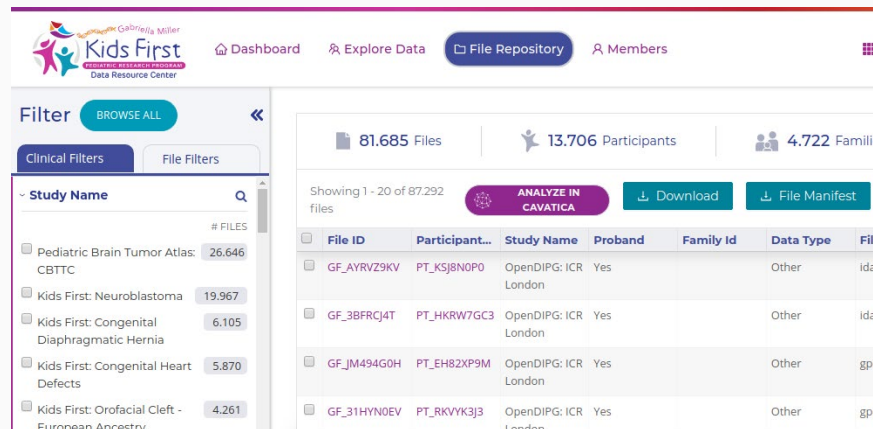# iPC Access/Data portal

❑ Access platform services from one single place and one single login.

Implementation of a central platform for data/metadata storage.
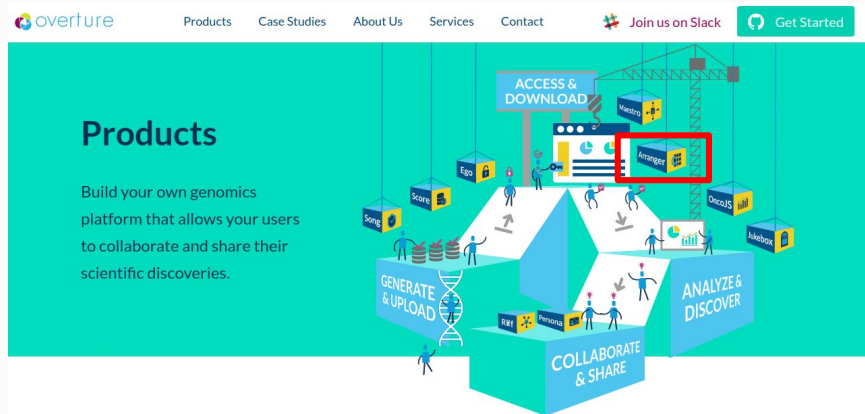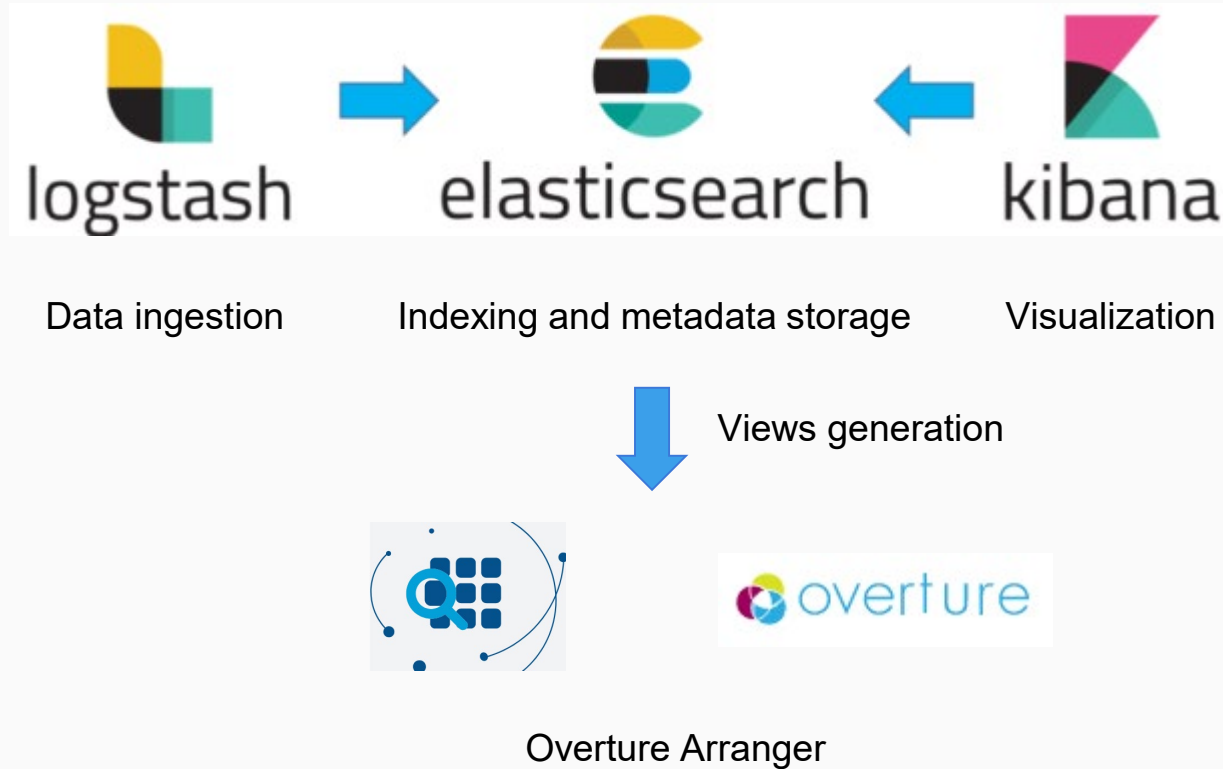

➢ **Overture's stack**

❑ Arranger: Data portals UI components.

➢ **Arranger example**

❑ Kids First Data Portal

Implementation of a central platform for data/metadata storage.

Data ingestion   Indexing and metadata storage   Visualization

Views generation

Overture Arranger

Implementation of a central platform for data/metadata storage.
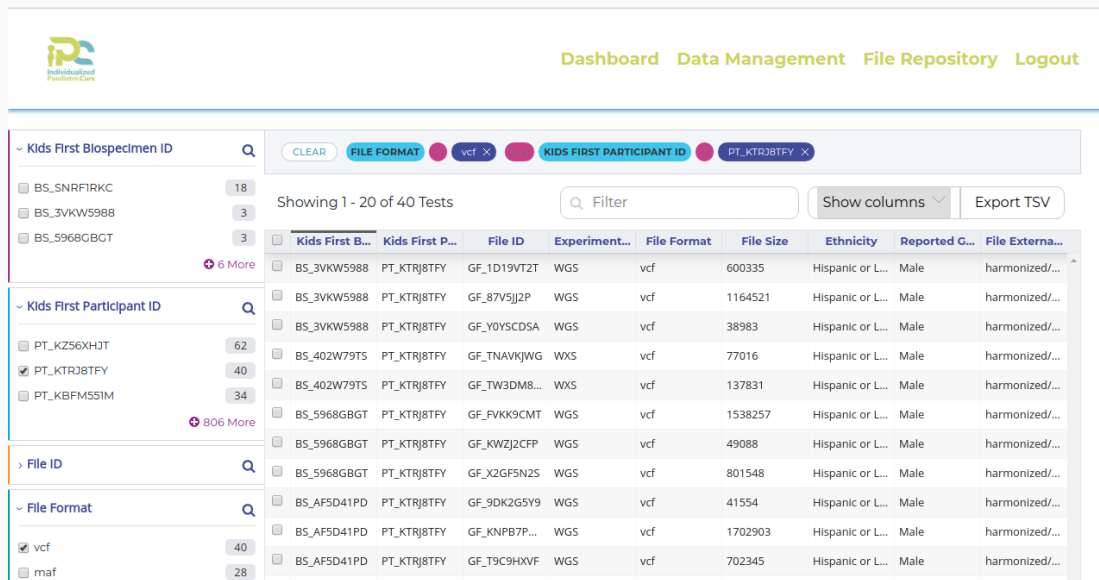
➢ Arranger server UI

❑ Generate views from an ElasticSearch index

iPC – individualizedPaediatricCure

iPC Catalogue - Data visualization based on Arranger

iPC Data Catalogue: Open Paediatric Brain Tumor Atlas

React

iPC Catalogue - Open Pediatric Brain Tumor Atlas.

➢ **Open Pediatric Brain Tumor Atlas (PBTA)**

❏ Genomic dataset (whole genome sequencing, whole exome sequencing, RNA sequencing, proteomic, and clinical data) for nearly 1,000 tumors.

❏ Metadata publicly available at Kids First Data Portal.

Experimental Strategy	Normal	Tumor
Targeted DNA Panel	1	1
RNA-Seq	0	1028
WGS	801	940
WXS	31	31

iPC — individualizedPaediatricCure

iPC Catalogue - Open Pediatric Brain Tumor Atlas.

➢ **Data handling:**

Phenotypic and file descriptors

A. Extraction

Participants list and associated metadata from Pediatric Brain Tumor Atlas

B. Processing

Participant

- → Biospecimen n
  - → File n
  - → File 2
  - → File 1
- → Biospecimen 2
- → Biospecimen 1
  - → File 1
  - → File 2
  - → ...

C. Ingestion

iPC Catalogue - Outbox API

Data model

Phenotypic descriptors: **Which?**

File descriptors: **Where?**

Outbox API

Catalogue

File Name	Kf Id
HKLYJCCXX-2.hgv.bam	GF_PA4QSRH7
HKMFHCCXX-7.hgv.bam	GF_JWWBRFFR
HKNFYCCXX-5.hgv.bam	GF_5KSEA2F2
HKM22CCXX-5.hgv.bam	GF_TKSAVY3Z

Show 20 ▾ rows

<<

iPC Catalogue - Analysis - Data storage

File ID, ES_index

Metadata

File ID, ES_index

File descriptors

Catalogue

Preselected data sets from catalogue:

Inspect and/or load data sets for their analysis.

fileID : NC_demo_00
file_locator : 6361
es_host : pbta_histologies_filecentric
Load to VRE Load to Cavatica Get Details

fileID : NC_demo_01
file_locator : 5070
es_host : pbta_histologies_filecentric
Load to VRE Load to Cavatica Get Details

fileID : NC_demo_02
file_locator : 5805
es_host : pbta_histologies_filecentric

Outbox API

File ID, ES_index

Metadata

File descriptors

Analysis

Virtual Research Environment

CAVATICA

Locator

File path

Storage

Nextcloud

File descriptors: File ID, Locator, ES_index

Implementation of a central platform for data/metadata storage.
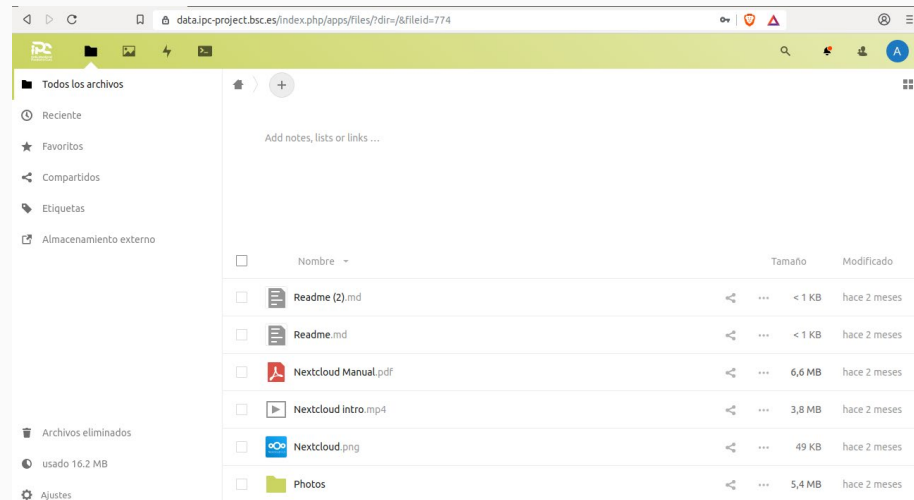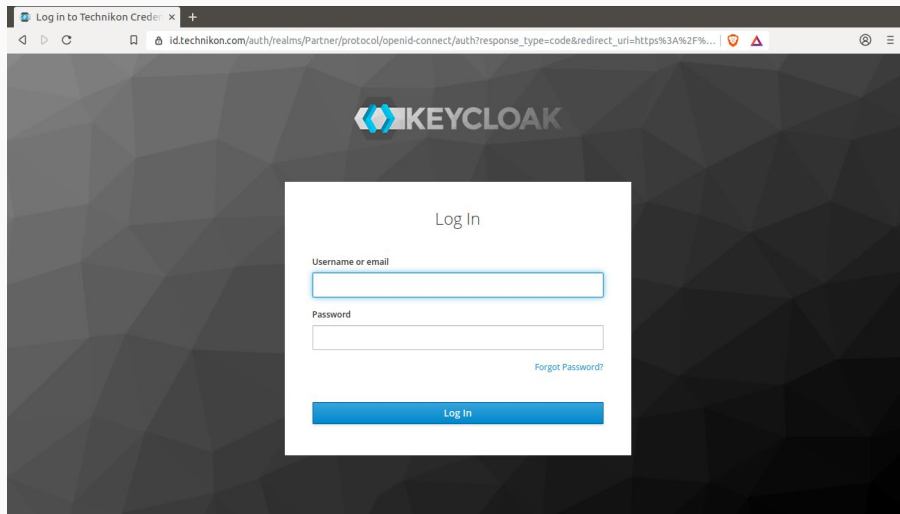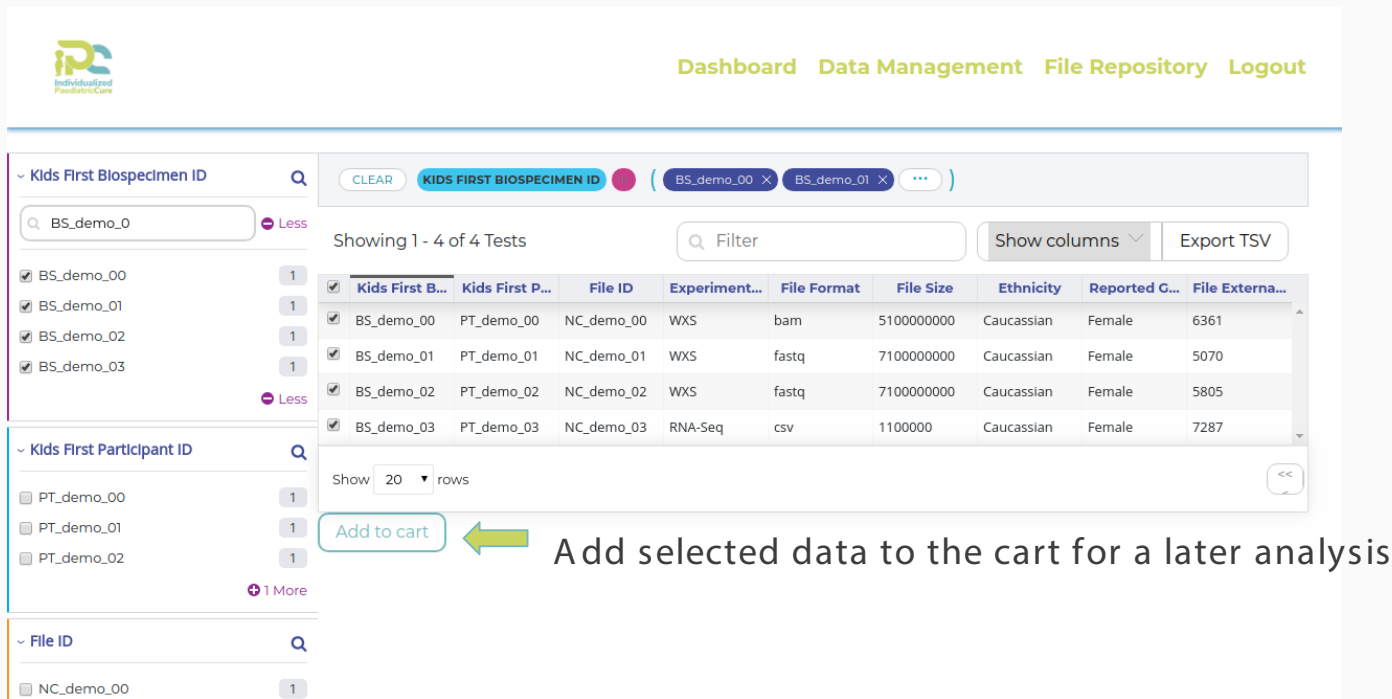

Authentication: Technikon's Keycloak

Nextcloud personal dashboard

iPC Catalogue - Select data from Arranger table

Add selected data to the cart for a later analysis

iPC – individualizedPaediatricCure

iPC Catalogue - Data Management section

Preselected data sets from catalogue:

Inspect and/or load data sets for their analysis.

fileID : NC_demo_00

file_locator : 6361

es_host : pbta_histologies_filecentric

Load to VRE   Load to Cavatica   Get Details

fileID : NC_demo_01

file_locator : 5070

es_host : pbta_histologies_filecentric

Load to VRE   Load to Cavatica   Get Details

fileID : NC_demo_02

file_locator : 5805

es_host : pbta_histologies_filecentric

Preselected data sets from catalogue:

Inspect and/or load data sets for their analysis.

fileID : REQUEST_DEMO_ID

file_locator : REQUEST_DEMO_LOCATOR

es_host : pbta_histologies_filecentric

Load to VRE   Load to Cavatica   Request Access

❑ Expose selected metadata to openVRE (Outbox API) or Cavatica (not yet).

❑ Data Access Requests ¿?

iPC Catalogue - Analysis (open VRE)

.

Data sets available to the iPC openVRE:

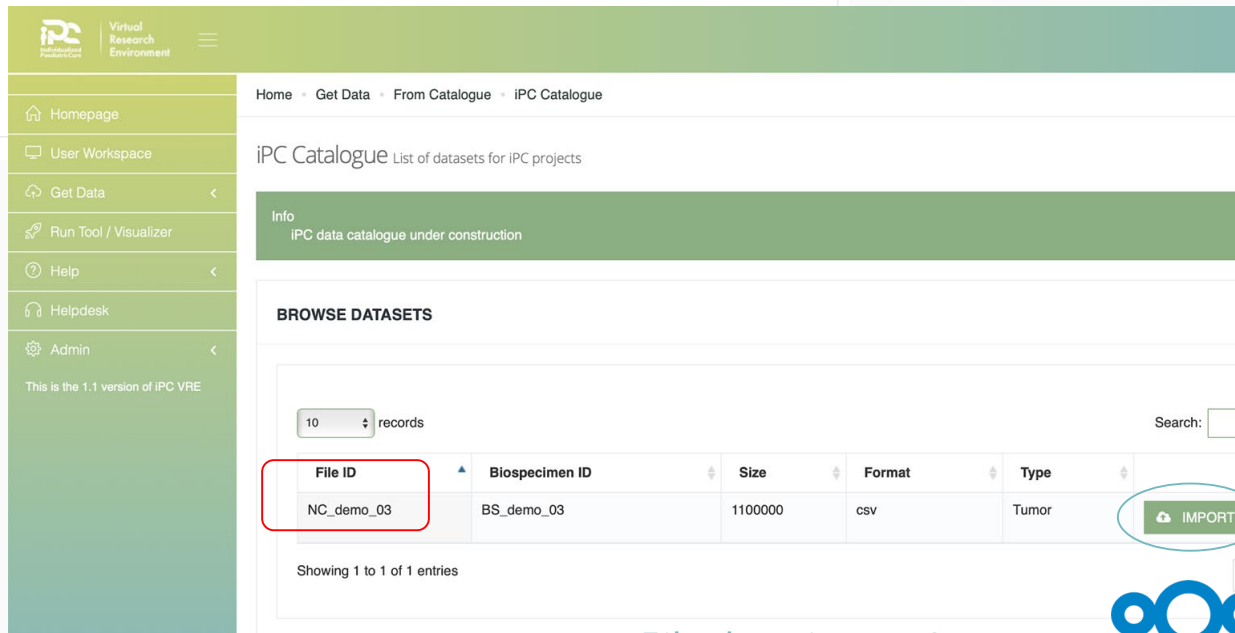Inspect and/or remove already loaded data sets into VRE.

Go to iPC openVRE

fileID : NC_demo_03

file_locator : 7287

es_host : pbta_histologies_filecentric

Unload data   Get Details

iPC · Virtual Research Environment

Home · Get Data · From Catalogue · iPC Catalogue

🏠 Homepage

🖥 User Workspace

☁ Get Data

🚀 Run Tool / Visualizer

❓ Help

🎧 Helpdesk

⚙ Admin

This is the 1.1 version of iPC VRE

iPC Catalogue List of datasets for iPC projects

Info
iPC data catalogue under construction

BROWSE DATASETS

10   records   Search:

File ID	Biospecimen ID	Size	Format	Type	
NC_demo_03	BS_demo_03	1100000	csv	Tumor	☁ IMPORT

Showing 1 to 1 of 1 entries

File descriptors: Locator

17/September/2020

Moving Forward

Jolanda Modic
(XLAB)

Moving Forward

iPC is here to help YOU advance YOUR research in paediatric cancer.

- Provide metadata
  Help us grow the data model, make the most of the available, public and private, data sources.

- Provide data
  Help us grow the data catalogue, make the most of iPC resources.

- We'd love to hear from you
  Reach out to alejandro.canosa@bsc.es and jose.m.fernandez@bsc.es

Thank you!

iPC Grant Agreement No. 826121

The **iPC** project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement
No 826121.

Coordinator

Technikon Forschungs- und Planungsgesellschaft mbH
Burgplatz 3a,
9500 Villach, AUSTRIA
Phone: +43 4242 233 55
Fax: +43 4242 233 55 77
Mail: **coordination@IPC -project.eu**