



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



**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)

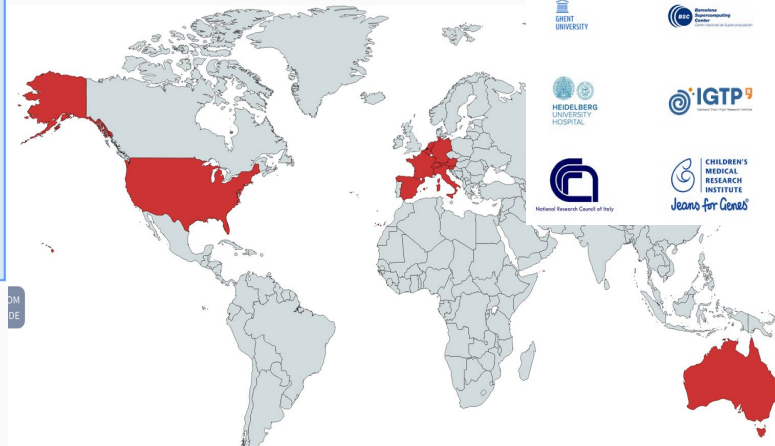
IPC Consortium - 21 Partners from 3 Continents

Project Management

Coordinator
TEC, Klaus-Michael Koch

Technical Lead
IBM, Maria Rodriguez Martinez

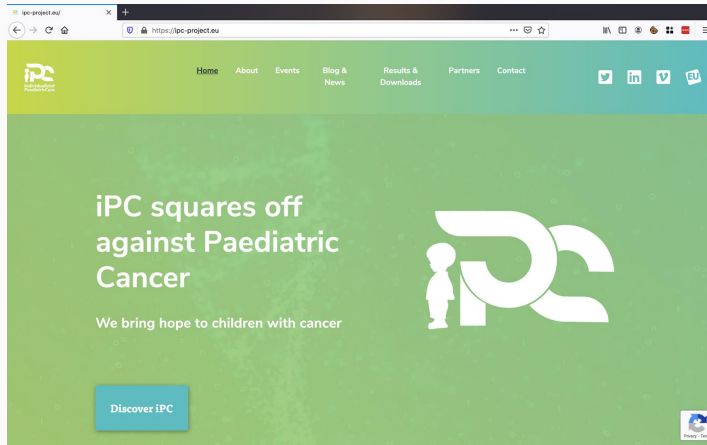
Scientific Lead
MPG, Hans Lehrach



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.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...)

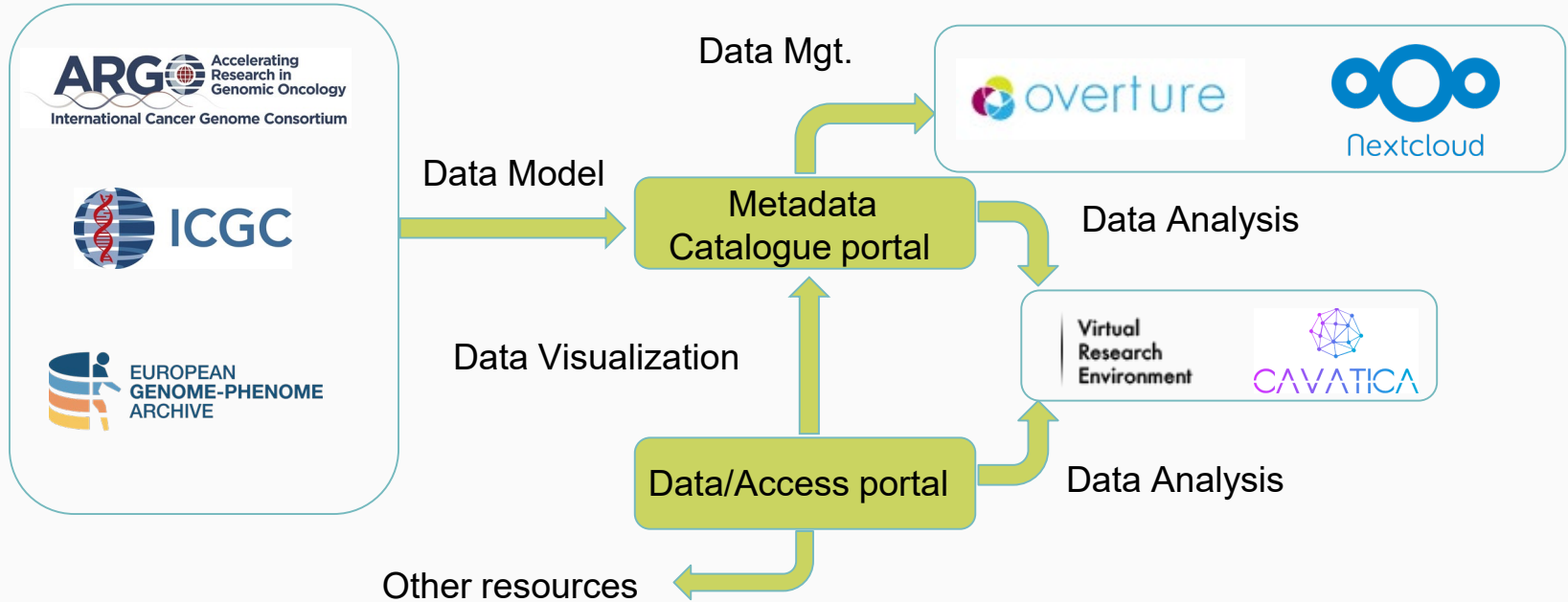
all this while...

Data Access procedures and compliance to Data protection laws are simplified.

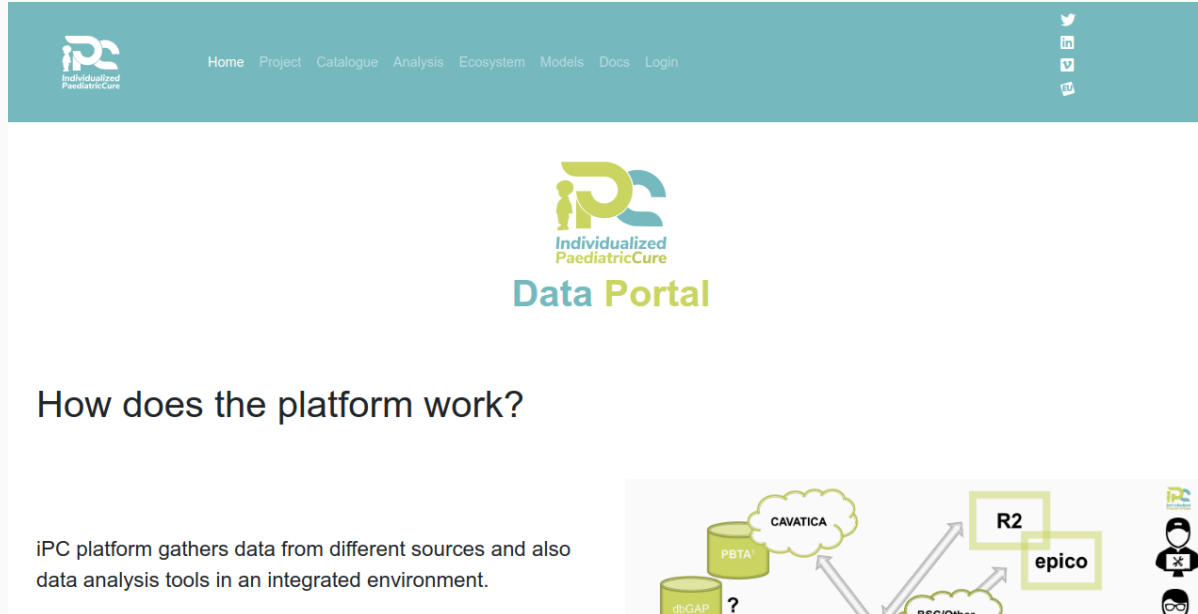
Which are our main objectives (so far) ?

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

iPC Platform architecture




iPC Access/Data portal



The screenshot shows the iPC Data Portal website. The header is teal with the iPC logo on the left, navigation links (Home, Project, Catalogue, Analysis, Ecosystem, Models, Docs, Login) in the center, and social media icons on the right. The main content area is white and features the iPC logo and the text "Data Portal". Below this, the question "How does the platform work?" is posed. A text box explains that the iPC platform gathers data from different sources and also data analysis tools in an integrated environment. To the right of this text is a diagram illustrating the data flow: data from sources like dbGAP, PBT, and CAVATICA is integrated into a central system (R2) which then feeds into analysis tools like epico. The diagram also shows a cloud labeled "BSC/Other" and a question mark, indicating ongoing or future data integration.

How does the platform work?

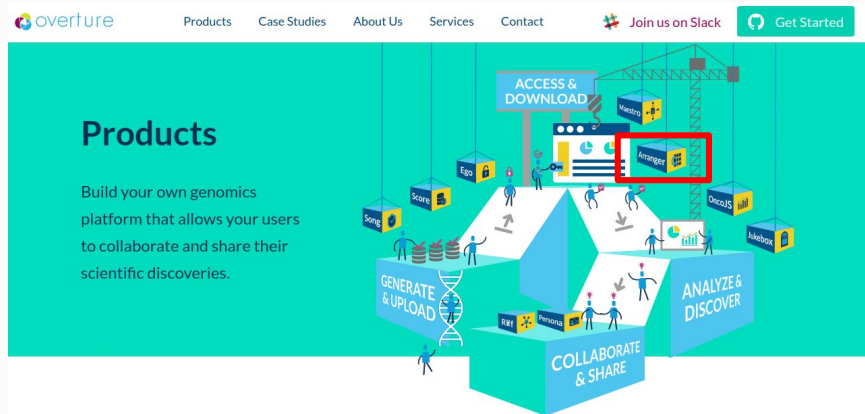
iPC platform gathers data from different sources and also data analysis tools in an integrated environment.



```
graph LR; dbGAP[(dbGAP)] --> CAVATICA((CAVATICA)); PBT[(PBT)] --> CAVATICA; CAVATICA --> R2[R2]; BSC((BSC/Other)) --> R2; R2 --> epico[epico];
```

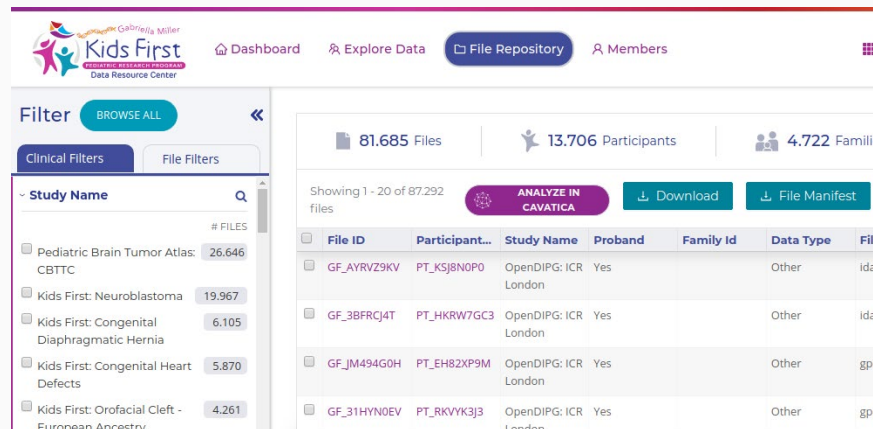
- Access platform services from one single place and one single login.

Implementation of a central platform for data/metadata storage.



➤ Overture's stack

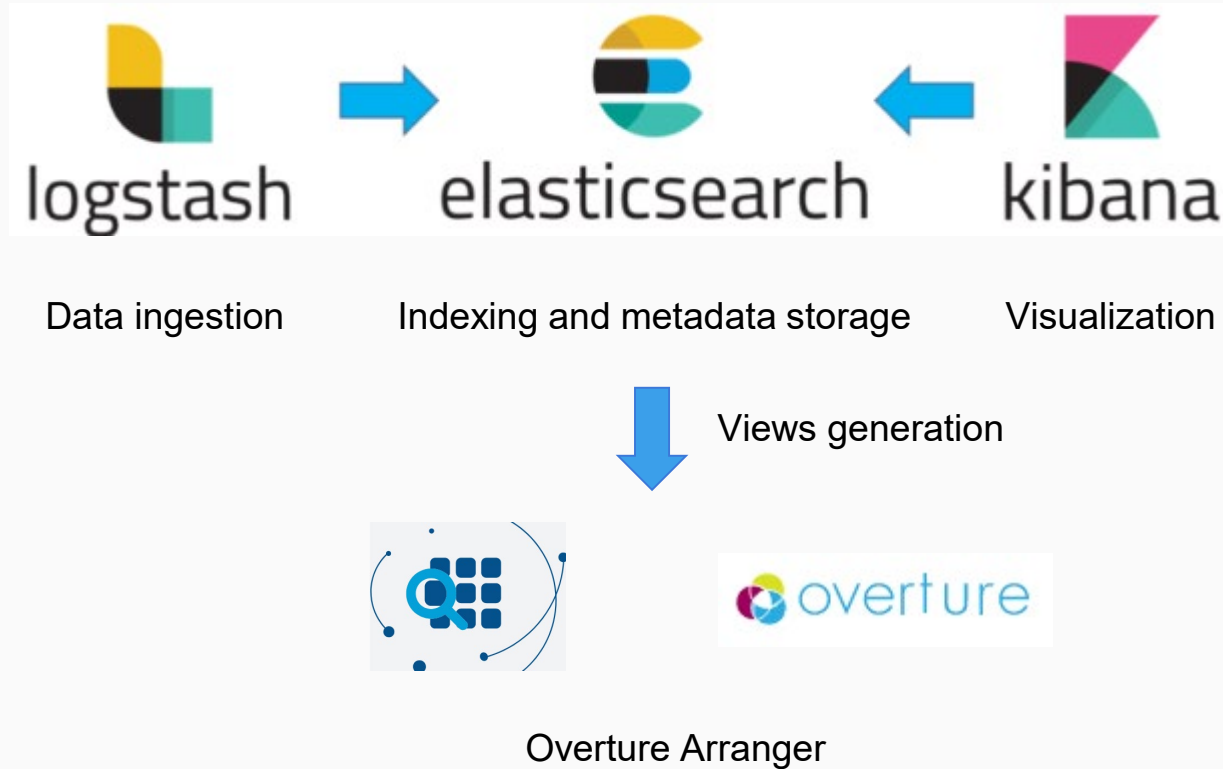
- ❑ Arranger: Data portals UI components.



➤ Arranger example

- ❑ Kids First Data Portal

Implementation of a central platform for data/metadata storage.



Implementation of a central platform for data/metadata storage.

Arranger

Save Project Cancel

Fields Aggs Panel Table Quick Search

Field filter

Type

Is active

Is array

Is primary key

Quicksearch enabled

Fields (157)

acl

availability

controlled_access

created_at

data_type

experiment_strategies

external_id

file_format

file_name

instrument_models

is_harmonized

is_paired_end

kf_id

latest_did

Field: kf_id

Display Name

Kf Id

Aggregation Type

id

☐ Active

☐ Quicksearch enabled

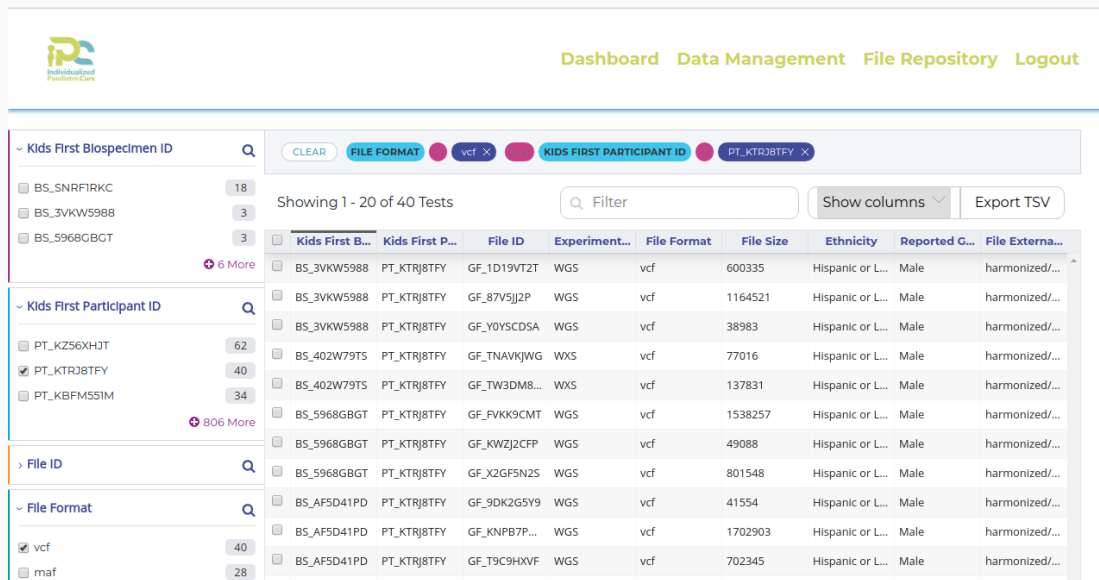
☐ Is primary key

☐ Is array

➤ Arranger server UI

📄 Generate views from an ElasticSearch index

iPC Catalogue - Data visualization based on Arranger



The screenshot displays the iPC Catalogue interface, which is a data visualization tool based on Arranger. The interface includes a sidebar with filters for 'Kids First Biospecimen ID', 'Kids First Participant ID', 'File ID', and 'File Format'. The main area shows a table of tests with columns for 'Kids First B...', 'Kids First P...', 'File ID', 'Experiment...', 'File Format', 'File Size', 'Ethnicity', 'Reported G...', and 'File Externa...'. The table is filtered to show 1-20 of 40 tests. The filters are set to 'FILE FORMAT' (vcf), 'KIDS FIRST PARTICIPANT ID' (PT_KTRJ8TFY), and 'PT_KTRJ8TFY'.

Kids First B...	Kids First P...	File ID	Experiment...	File Format	File Size	Ethnicity	Reported G...	File Externa...
BS_3VKW5988	PT_KTRJ8TFY	GF_1D19VT2T	WGS	vcf	600335	Hispanic or L...	Male	harmonized/...
BS_3VKW5988	PT_KTRJ8TFY	GF_87V5J2P	WGS	vcf	1164521	Hispanic or L...	Male	harmonized/...
BS_3VKW5988	PT_KTRJ8TFY	GF_Y0YSCDSA	WGS	vcf	38983	Hispanic or L...	Male	harmonized/...
BS_402W79TS	PT_KTRJ8TFY	GF_TNAVKJWG	WXS	vcf	77016	Hispanic or L...	Male	harmonized/...
BS_402W79TS	PT_KTRJ8TFY	GF_TW3DM8...	WXS	vcf	137831	Hispanic or L...	Male	harmonized/...
BS_5968GBGT	PT_KTRJ8TFY	GF_FVK9CMT	WGS	vcf	1538257	Hispanic or L...	Male	harmonized/...
BS_5968GBGT	PT_KTRJ8TFY	GF_KWJ2CFP	WGS	vcf	49088	Hispanic or L...	Male	harmonized/...
BS_5968GBGT	PT_KTRJ8TFY	GF_X2GF5N2S	WGS	vcf	801548	Hispanic or L...	Male	harmonized/...
BS_AF5D41PD	PT_KTRJ8TFY	GF_9DK2G5Y9	WGS	vcf	41554	Hispanic or L...	Male	harmonized/...
BS_AF5D41PD	PT_KTRJ8TFY	GF_KNPB7P...	WGS	vcf	1702903	Hispanic or L...	Male	harmonized/...
BS_AF5D41PD	PT_KTRJ8TFY	GF_T9C9HXVF	WGS	vcf	702345	Hispanic or L...	Male	harmonized/...



React

iPC Data Catalogue: Open Paediatric Brain Tumor Atlas

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 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.

sing

Proces

Participant

Biospecimen n

Biospecimen 2

Biospecimen 1

File n

File 2

File 1

File 1

File 2

...

C.

ion

Ingest



iPC Catalogue - Outbox API

Data model

Phenotypic descriptors: **Which?**

File descriptors:

Where?

 mongoDB.

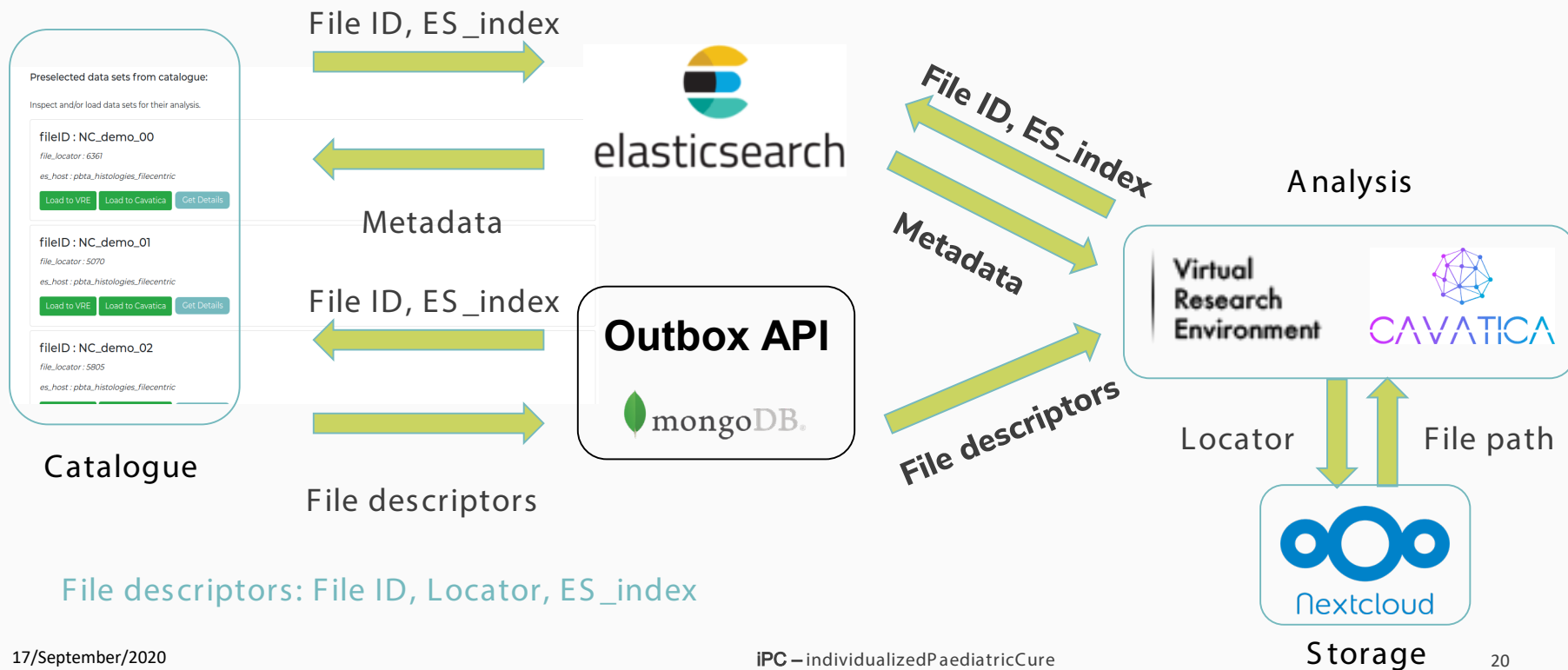
Outbox API

Catalogue

<input type="checkbox"/>	File Name	Kf Id
<input type="checkbox"/>	HKLYJCCXX-2.hgv.bam	GF_PA4QSRH7
<input type="checkbox"/>	HKMFHCCXX-7.hgv.bam	GF_JWWBRFFR
<input type="checkbox"/>	HKNFYCCXX-5.hgv.bam	GF_5KSEA2F2
<input type="checkbox"/>	HKM22CCXX-5.hgv.bam	GF_TKSAVY3Z

Show 20 rows

iPC Catalogue - Analysis - Data storage



Implementation of a central platform for data/metadata storage.

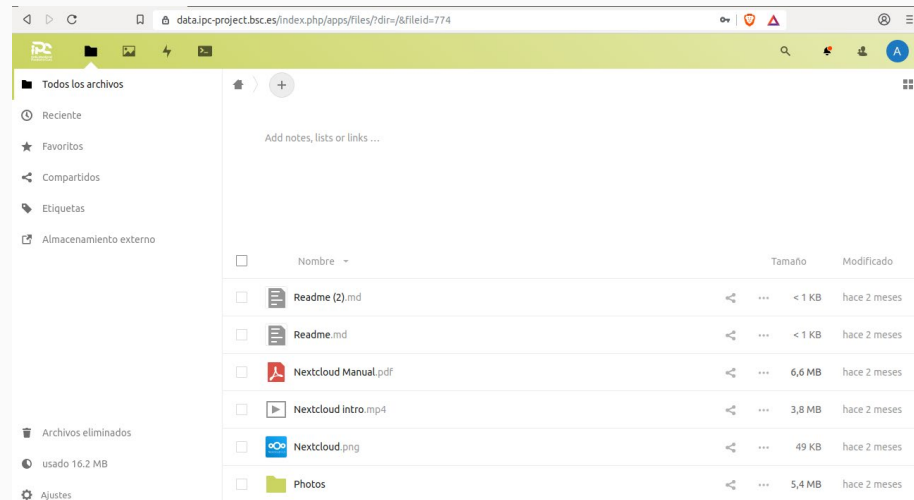
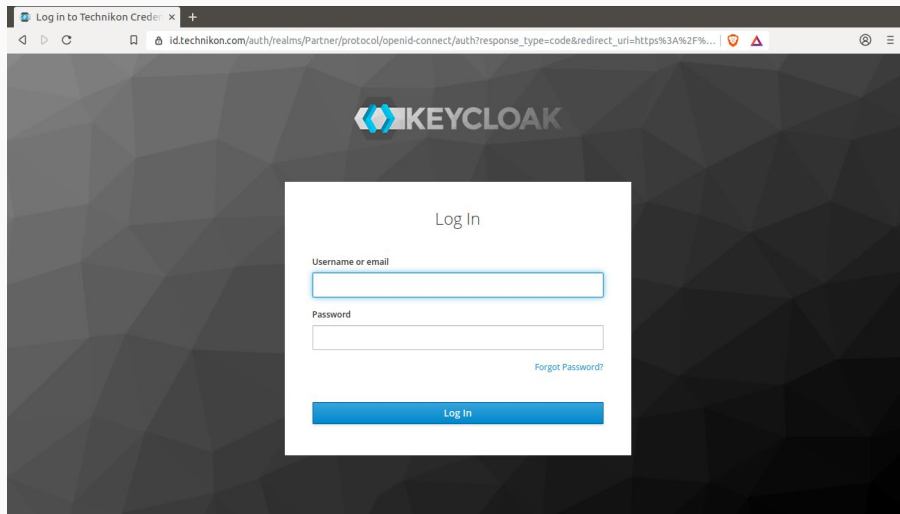


TECHNIKON

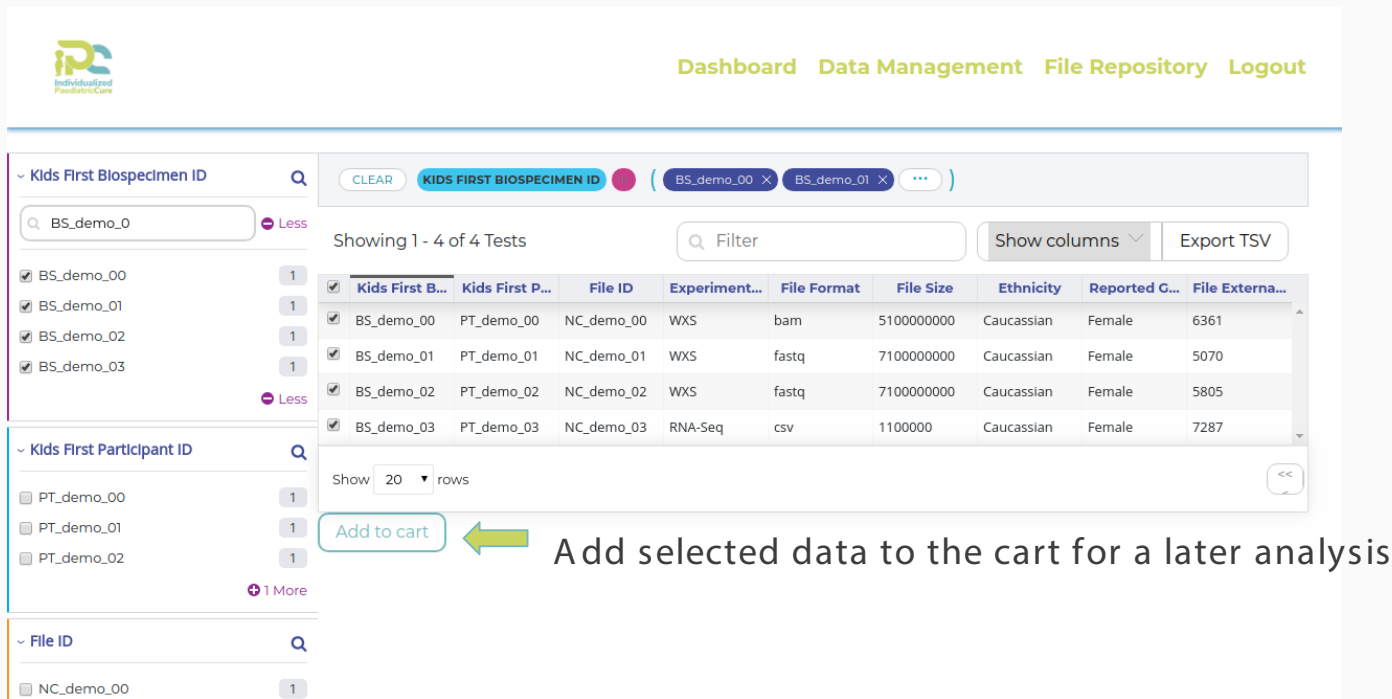


Authentication: Technikon's Keycloak

Nextcloud personal dashboard



iPC Catalogue - Select data from Arranger table



Dashboard Data Management File Repository Logout

~ Kids First Biospecimen ID

CLEAR KIDS FIRST BIOSPECIMEN ID (BS_demo_00 X BS_demo_01 X ...)

BS_demo_0 Less

Showing 1 - 4 of 4 Tests Filter Show columns Export TSV

<input checked="" type="checkbox"/>	Kids First B...	Kids First P...	File ID	Experiment...	File Format	File Size	Ethnicity	Reported G...	File Externa...
<input checked="" type="checkbox"/>	BS_demo_00	PT_demo_00	NC_demo_00	WXS	bam	5100000000	Caucassian	Female	6361
<input checked="" type="checkbox"/>	BS_demo_01	PT_demo_01	NC_demo_01	WXS	fastq	7100000000	Caucassian	Female	5070
<input checked="" type="checkbox"/>	BS_demo_02	PT_demo_02	NC_demo_02	WXS	fastq	7100000000	Caucassian	Female	5805
<input checked="" type="checkbox"/>	BS_demo_03	PT_demo_03	NC_demo_03	RNA-Seq	csv	1100000	Caucassian	Female	7287

Show 20 rows

Add to cart

~ Kids First Participant ID

PT_demo_00 1

PT_demo_01 1

PT_demo_02 1

+ 1 More

~ File ID

NC_demo_00 1

Add selected data to the cart for a later analysis

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.

fileID : NC_demo_03

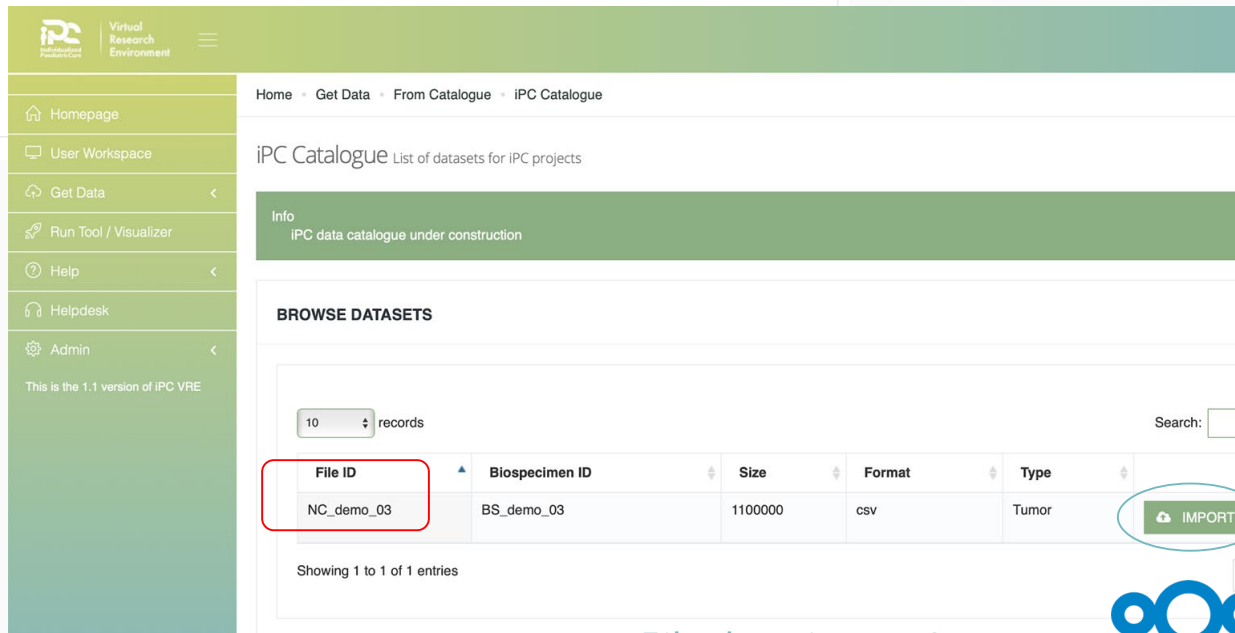
file_locator : 7287

es_host : pbta_histologies_filecentric

Unload data

Get Details

Go to iPC openVRE



Home » Get Data » From Catalogue » iPC Catalogue

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

Showing 1 to 1 of 1 entries

IMPORT



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 .

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author's view – the European Commission is not responsible for any use that may be made of the information it contains. The users use the information at their sole risk and liability.

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](mailto:coordination@IPC-project.eu)