

iPC connects paediatric cancer experts with key data sources

Blog post by Jolanda Modic ([XLAB](#)) and Alejandro Canosa ([BSC](#))

Unlike the popular perspective and as defended by many (and pointed out by the North American [National Cancer Institute](#), NCI), cancer is not one but rather a collection of related diseases, usually known as types of cancer. They share the fact that, when sick, “[some of the body’s cells begin to divide without stopping and spread into surrounding tissues.](#)” All cancer types negatively affect the quality and the lifestyle of the patient and most are still very lethal. In 2018 alone, around [9.6 million people around the globe died from cancer](#) - 1 person every 3.3 seconds. In Europe, the cancer toll in 2018 was almost [2 million people](#). Cancer is also a great source of research problems in the biomedical domain, driving some of the biggest achievements in health during the XXI century.

Cancer was and remains one of the leading causes of death.

Apart from having a very high mortality rate, cancer also creates a significant burden on our healthcare systems and global economy. [One in five men and one in six women](#) worldwide develop cancer during their lifetime. Each of these will require a diagnosis, will undergo some sort of treatment, and will contribute to productivity loss. In 2018, the total cost of cancer in Europe was [€199 billion](#). In the US, the economic burden of cancer is around [1.8% of the Gross Domestic Product \(GDP\)](#).

The economic impact of cancer across the globe is substantial and increasing.

Cancer affects everyone. People from different backgrounds, of different races, of both genders, and of all ages - even our youngest. Luckily, cancer is relatively rare among kids. Nevertheless, it is still a [major cause of death](#) (in some regions even the leading cause of death) in children worldwide. Each year, across the globe, [approximately 200.000 children from 0 to 14 years old are diagnosed with cancer and around 75.000 children with cancer die.](#)



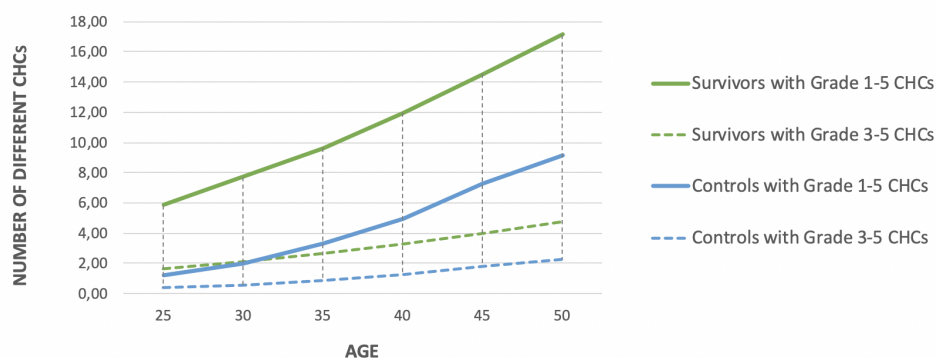
The majority of [paediatric cancers are different from cancer in adults](#). Children's bodies work in a unique way and so the type of the developed cancer, how far and how fast it spreads throughout the body, and how it responds to treatment is often different than in adults. Another differentiating factor is the cause. In adults, cancer is often a consequence of some environmental or lifestyle factor, like poor diet or smoking. But this is not the case with kids. Many times, the cause of a childhood cancer is not known. So studying the disease and preventing it is very challenging.

In adults, cancer is typically diagnosed in early stages. [With children it is quite different](#). When they are infants, they are unable to truly explain how they feel and they are unable to properly communicate their symptoms. When kids grow a bit older, the diagnosis is usually delayed because they are more resilient than adults - kids stay active, they play and run around regardless of the symptoms that would disable many adults from functioning normally. So, very often ([in 80% of children with cancer](#)), by the time cancer is discovered, the disease has already spread to other organs or tissues.

Paediatric cancer generally cannot be prevented or screened. The best strategy to fight the disease and improve outcomes relies on finding effective treatments.

When a child is diagnosed with cancer, due to the late discovery, the need for a treatment is typically urgent. But treating kids with cancers, again, differs from treating adults. Kids' bodies respond better to cancer treatments than adults. But the side effects of chemotherapy, radiotherapy, and immunotherapy are far more severe for children because they cause developmental disorders and are longer lasting, inducing secondary, life-threatening diseases. According to a recent [retrospective analysis of childhood cancer survivors](#), as a result of the treatments they had as kids, by the time they're 50 years old, more than 99% of paediatric cancer survivors will have various chronic health problems. In fact, they will experience, on average, 17 different chronic health conditions, of which almost 4 will be severe, disabling, or fatal. Even if cancer itself isn't fatal for a child, it will surely create a high burden of health complications during his/her adulthood and significantly affect his/her physical and mental quality of life.

Number of Chronic Health Conditions (CHCs) in paediatric cancer survivors compared to individuals without a history of a paediatric cancer



Over the past few decades, treatments for paediatric cancer have dramatically evolved. Survival rates have significantly increased, [from 5-year survival rates around 0% in the 1940s to modern survival rates of over 80%](#). Recently, a lot of efforts have been put towards developing personalised treatments for paediatric cancers that are more effective and, at the same time, less invasive and with fewer side effects, causing lesser lifelong damage to young bodies and brains that are still developing.

THE FUTURE OF PAEDIATRIC CANCER TREATMENT IS IN



PERSONALISATION

Personalised cancer care is promising, but there are several challenges to overcome. Such treatments can be very expensive and are currently only accessible through clinical trials and specific programmes. They are not (yet) available for all cancer types and are still being explored and developed. Advances in technology and data science greatly contributed to the development of personalised cancer treatments. But while rapid digitalisation and wide connectivity substantially improved cancer research, the basic ingredient - DATA - is not always easily accessible.

We are struggling to effectively gather, use, and share paediatric cancer data.

Paediatric cancer is rare. That's great! But it also means that no one hospital and no one research center in the world has enough data on a specific cancer type or subtype to perform meaningful research. To really understand the disease, draw the right conclusions, and win the cancer fight, researchers and clinicians need access to as much data from as many patients in as many institutions as possible.

The lack of data availability is tightly connected to a couple of other issues. Technology and regulation. Different organisations use different information systems and different standards for managing their data. So, whenever institutions want to share data sources, they usually encounter some [interoperability problem](#). The other problem they usually bump into are [legal barriers](#). Increasingly stricter privacy and data protection laws are raising barriers to collecting and sharing patients' data. Additionally, due to complex procedures and various policies that are in place in hospitals and organisations across different countries, sharing datasets usually comes with an incredible amount of bureaucracy.



*iPC is connecting different data sources
and creating a massive knowledge base
for a joint fight against paediatric cancer*

iPC is addressing the need to gather, harmonize, and share high-quality, multi-disciplinary paediatric cancer data - all this while simplifying data access procedures and at the same time ensuring compliance to various security policies and data protection laws. Making these resources easily findable and broadly available will support a wide community of cancer research experts in creating effective personalised therapies for kids with cancer. To this end, the consortium is building a cloud-based platform for storing, accessing, analysing, and sharing data and disseminating results relevant for paediatric cancer research. The platform will host newly produced datasets and will also link to other similar repositories and platforms hosting paediatric cancer data and offering data analytics and data visualisation tools (such as [EGA](#), [ELIXIR](#), [Kids First](#), [BBMRI-ERIC](#), [dbGaP](#), [R2](#), [PeCan](#)). This connected data infrastructure will simplify sharing of resources from multiple locations and will make data work better for researchers and clinicians around the globe.

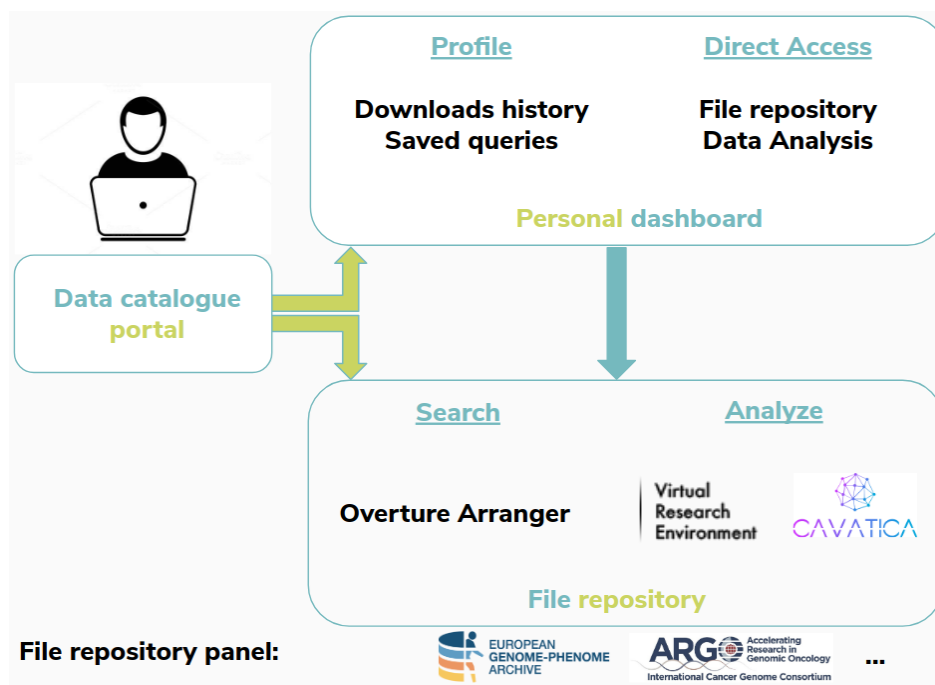
Extracting knowledge and creating value from data depends not only on the data availability, but also on data usability. Data needs to be Findable, Accessible, Interoperable, and Reusable ([FAIR](#)). Data and the associated information about the data (i.e. metadata) should be easy to find and, once the data is found, the user needs to know how to access the data easily and efficiently, but securely. The data should be in a format that supports integration with other data sources and with existing workflows and applications. Finally, data and metadata should be well-described so they can be easily reused in different settings.

Looking separately at each of the existing repositories and platforms hosting cancer research data, each of them already follows these principles. The problem appears when one wants to easily combine and utilize these resources. Different platforms store different types of data (e.g. different categories of genomic data, different forms of clinical data) describing different cancers (e.g. brain tumors, liver cancer, leukemia) in different file formats (as textual data, as images, or in specific [“genomic” formats](#) like BAM, FASTQ, TSV or others). Consequently and naturally, each platform adjusts and accommodates their search, access, combination, analysis, and visualisation functionalities according to the data sources that they have and according to the audience their specific resources are intended for.

*iPC will collect, integrate, and harmonize a wide list of existing data sources
and unstructured knowledge relevant for paediatric cancers.*

To this end, the iPC consortium will create a data catalogue portal (based on [Arranger](#), an open-source data search framework) and offer it through the iPC platform to a wide community of cancer researchers and clinicians for free. This catalogue will list existing datasets available in other repositories and will also enable users to upload their own data.

We will simplify access to available data sources without compromising their confidentiality, and will also offer users different models, tools, and computational resources for their paediatric cancer research or clinical practice. The overview of the data catalogue portal is illustrated below.



We will follow community-based standards (e.g. those from the [GA4GH](#) alliance) to harmonise metadata and thereby make datasets easily findable, regardless of their origin, type, format, the disease they describe or the profile of the patient they relate to. The visualisation and search functionalities offered by the portal are illustrated below. The visualisation and search functionalities offered by the portal are illustrated below.

The screenshot shows the data catalogue portal interface. The top navigation bar includes links for **Dashboard**, **Explore Data**, **File Repository**, **Members**, **Resources**, and **Logout**. The main content area displays a table of file demos with columns for **External Id**, **File Format**, **Kf Id**, and **Size**. The table is filtered by **Controlled Access** (Any, Yes 9, No 2) and **Created At** (Start Date, End Date). The **File Format** filter shows **bam** (6) and **vcf** (5). The **Participants Gender** filter shows **Female** (6) and **Male** (5). The **Size** filter shows a range from 47,620,160,721 to 87,620,100,743. The table shows 11 rows of data, with the first row being **s3://kf-seq-data-bcm/chung/HKLYCCXX-2...** in **bam** format, with **Kf Id** **GF_PA4Q5RH7** and **Size** **65776072338**.

External Id	File Format	Kf Id	Size
s3://kf-seq-data-bcm/chung/HKLYCCXX-2...	bam	GF_PA4Q5RH7	65776072338
s3://kf-seq-data-bcm/chung/HKMFHCCXX...	bam	GF_JWWBRFFR	62991171434
s3://kf-seq-data-bcm/chung/HKNFYCCXX...	bam	GF_5KSEA2F2	70709429717
s3://kf-seq-data-bcm/chung/HKM22CCXX...	bam	GF_TKSAVY3Z	65620100743
s3://kf-seq-data-bcm/chung/HRR32CCXX...	vcf	GF_RZSDVY3Z	87620100743
s3://kf-seq-data-bcm/chung/SDF32CCXX-4...	vcf	GF_MGFGZK4Z	57620160741
s3://kf-seq-data-bcm/chung/ALK32SYXT-4...	bam	GF_MTTGZK9D	47620160721
s3://kf-seq-data-bcm/chung/RTD02SYXT-3...	bam	GF_JH5GZK7L	64320160721
s3://kf-seq-data-bcm/chung/ZZZ42SYXT-4...	vcf	GF_RRGZL7L	64320160721
s3://kf-seq-data-bcm/chung/YHE42SYXT-7...	vcf	GF_QQDGL7C	54320160721
s3://kf-seq-data-bcm/chung/ASD56SYXF-7...	vcf	GF_ERCGZE8C	75520160781

This catalogue will provide a critical foundation for sharing valuable data sources, for the development of intelligent models and algorithms that enable simulation of virtual patients and prediction of optimal, individualized treatments for them.

The data catalogue will be available later this year, before the fall. When the catalogue is deployed, we will share a detailed guideline for its use and offer online training (sign up for the webinar [here](#)) to show how clinicians and scientists can start contributing to the catalogue and using our resources.

In the meantime, check out other related material on our [website](#), follow us on social media, and help us raise awareness among clinicians and scientists worldwide about this opportunity to join forces in the fight against paediatric cancer.