CRUK Data Hub
A Sneaky Peek
(or how to get more for your cancer buck)

Sarah Wooller[1], Ayoola Olojede[1], Andrew Levy[1], Sanika Raut[2], Eytan Kovler[2], Jamie Byrne[3], Loki Sinclair[3], Peter Harrison[3], Joseph Day[4], Iliada Eleftheriou[5], Andrew Blake[6], Frances Pearl[1]

1. University of Sussex, 2. CRUK, 3. HDR UK, 4.CRH, 5. University of Manchester, 6. University of Oxford.

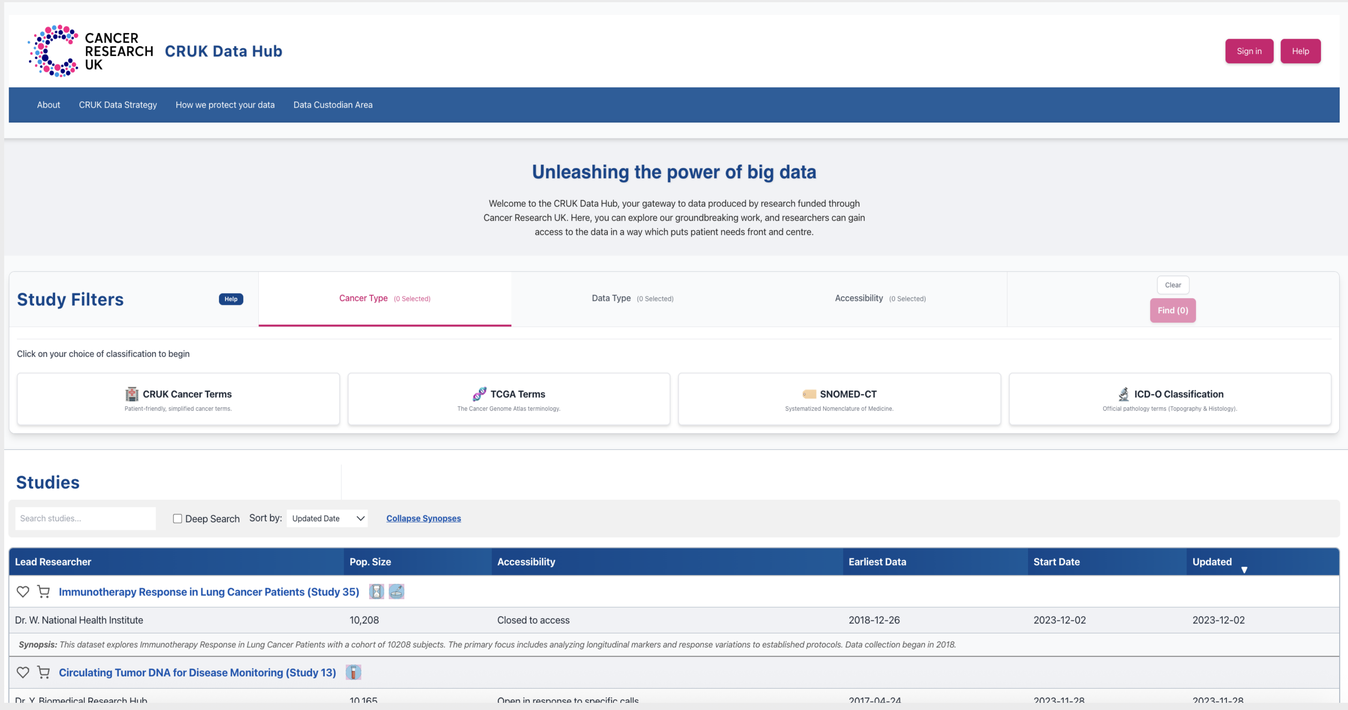
Contact: f.pearl@sussex.ac.uk

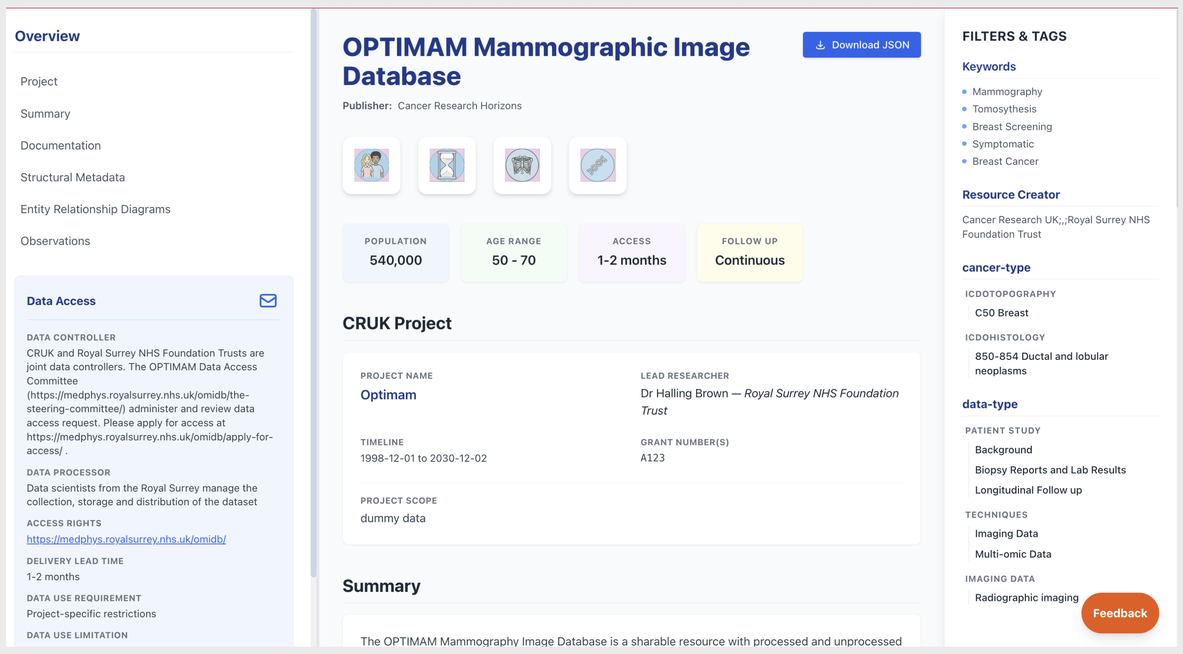
CRUK has funded the generation of thousands of cancer datasets, but most are only known to researchers directly involved in their generation or their established collaborators. It is very difficult for other researchers to know what data has already been generated and how to access theses datasets or samples for re-analysis. The process taken to access these data is often significantly time-consuming.

The CRUK Data Hub is a joint venture between the University of Sussex - Pearl Bioinformatics Lab, CRUK, HDR UK, CRH and the group advisors of the CRUK Data Science Community.

Using the power of the HDR UK Gateway we aim to make CRUK cancer datasets more findable and accessible. As well as inheriting HDRUK's elastic search the hub will enable datasets to be filtered by cancer type, data type and accessibility. Here we present our designs for the search, meta data and upload pages.

The Search Page:
Find the Data you need: We have included cancer filters for ICD O, TCGA, CRUK with more to come, as well as filters for different types of data to narrow down the search quickly and easily.

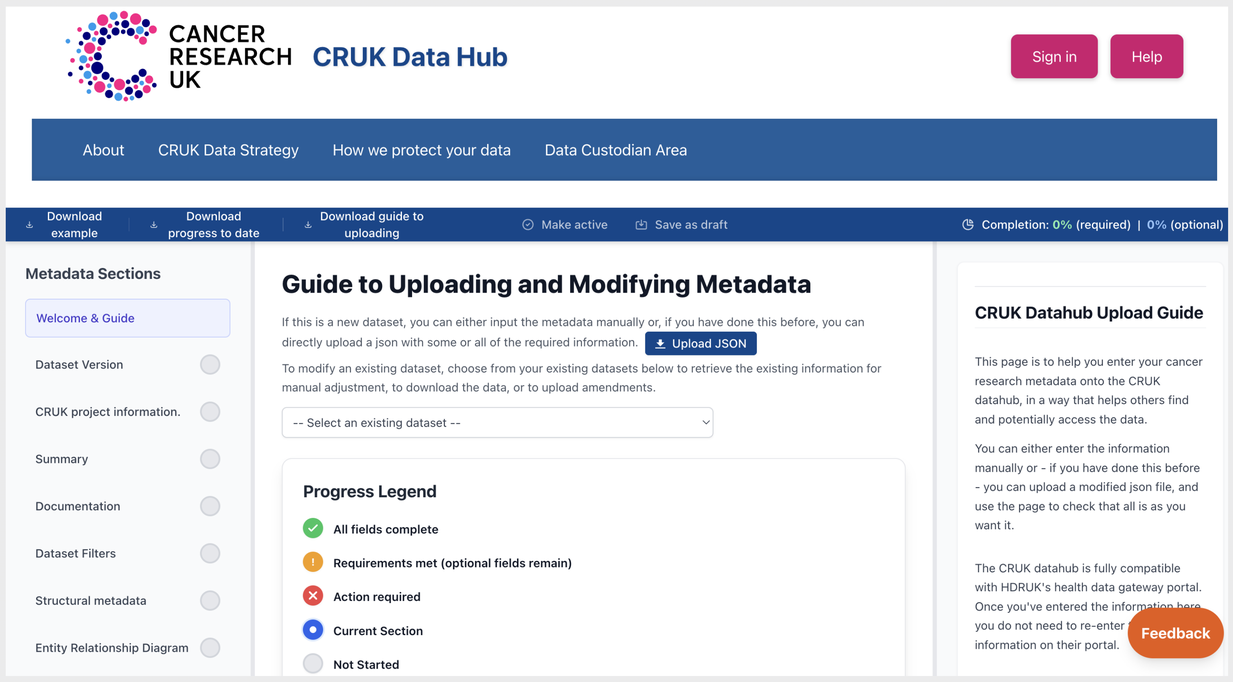




The Metadata Pages: Cancer metadata needs to be rich. We have enabled options for table structures, observations, filters, Entity Relationship Diagrams and more. They will also provide contact details for data access committees.
Our aim is to reduce the need for unnecessary queries and uncertainty.

The Upload Page. During this, the pilot stage the hub will be populated with high quality data resources from Cancer Research Horizons as well as linked data sets from the Cancer 4 Young People pilot studies.

Our aim is that inclusion in the data hub will be the standard for all datasets with CRUK funding.



Be part of our story. You can find our current designs and project news at https://sarahwooller.github.io/CRUK_datahub_landing_page

We welcome your feedback – either through the website or direct to skw24@sussex.ac.uk