

CRUK Datahub Upload Guide

This page is to help you enter your cancer research metadata onto the CRUK datahub, in a way that helps others find and potentially access the data. You can either enter the information manually or - if you have done this before - you can upload a modified json file, and use the page to check that all is as you want it.

The CRUK datahub is fully compatible with HDRUK's health data gateway portal. Once you've entered the information here you do not need to re-enter the information on their portal.

Before you start

You will need the following information.

Grant Documentation: CRUK grant number (and optionally a list of any other funding bodies).

Dataset Basics: Information describing your dataset, including a unique title, a concise abstract (up to 255 characters), and longer description of the dataset. This is an opportunity to include high quality information about the technical standards and methods used. If you have a website for your project or dataset you will able to link it.

Technical Specs: ICD-O descriptors for the cancers covered by the dataset are required. There is a search bar to help you find the right terms. These will automatically be translated into TCGA and CRUK cancer terms. It is also helpful to have a list of the data formats (CSV, DICOM, etc.) and terminologies (ICD-10, SNOMED CT) used.

Structural Assets: We ask for descriptions of the tables and the columns within, so that researchers can easily find out whether you have the information they need without needing to put in an enquiry. In order to give an idea of the completeness of the dataset we also ask you to include for each table how many complete rows there are. Please make descriptions understandable. And, if you have a complex relational database it is helpful to include an **Entity Relationship Diagram image file** (PNG/JPG/SVG < 5MB) to show how the different tables link together.

Demographics: Summarized counts for Age and Ethnicity.

Observations: If you have summary statistics about your dataset (like 5500 Matched tumour-normal samples) it is an excellent way to show your dataset to its best advantage.

Contacts: You will need the name of your Data Custodian organisation and a functional email for data access requests.

Navigation & Workflow

Sidebar: Use status icons to track progress (Red "X" for missing mandatory fields; Amber "!" for missing optional fields).

Dataset Filters: Tag your data with at least one Topography, Histology, Data Type, and Access Type. Access to data normally changes as a project and resulting dataesets mature.

Auto-Expansion: In those fields where you can add more than one entry typing in the last row automatically creates a new row.

Privacy: Apply standard low-number suppression in Demographic and Observation sections.