

## How results metadata files were obtained

For these search tasks, I obtained metadata files via the data repository user interfaces as described below. There are alternative ways to run queries and obtain files, such as through APIs, python and R packages, GraphQL query interfaces, and cloud software platforms; for pointers to information about these alternative ways of searching, see the Use Case Exercise 2 document.

### GDC metadata files:

The GDC patient case files were obtained via the GDC data exploration UI ([https://portal.gdc.cancer.gov/analysis\\_page?app=](https://portal.gdc.cancer.gov/analysis_page?app=)), by first selecting filters in the Cohort Builder, then going to the 'Cases' view, then selecting 'Table' view, and then clicking on 'tsv'.

The GDC sample metadata files were obtained by going to the 'Cases' view, then selecting 'Table' view, clicking on the 'Biospecimen' button, and then by clicking the download 'tsv' option below. This downloads a tar file; the sample metadata data file within this is named 'sample.tsv'.

The GDC files with file metadata were obtained via the GDC UI by first selecting cohort filters, then clicking on 'Repository' (getting to this to work sometimes required refreshing the page), then selecting filters for the data files, then clicking on 'Download Associated Data', and then choosing 'Sample Sheet'.

### PDC metadata files:

The PDC files were obtained using the PDC data exploration UI (<https://proteomic.datacommons.cancer.gov/pdc/browse>) by selecting filters. All metadata files were then obtained by selecting Export All Manifests csv (tsv is fine too). The patient case file is the Clinical one, the specimens file is the Biospecimens one, and the file metadata file is the Files one.

### HTAN metadata files:

The HTAN files were obtained using the Human Tumor Atlas Network (HTAN) exploration UI (<https://humantumoratlas.org/explore>) by selecting filters. All metadata files were obtained by selecting the appropriate data type tab ('Cases' for patient case data; 'Biospecimens' for specimen data; and 'Files' for file metadata) and then clicking Download Metadata.

The views, opinions, and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation. This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004.