1. ­Where is the data located? ­

The Cancer Moonshot Patent Data and associated documentation are located on the [USPTO Developer Hub](https://developer.uspto.gov/product/cancer-moonshot-patent-data).

1. What other datasets (government or private) would you want to investigate/link together if you had the time?

Below is a non-exhaustive list of datasets or sources we suggest participants to investigate and link to the Cancer Moonshot Patent Data. We encourage participants to utilize additional datasets and resources beyond those listed.

[www.PatentsView.org](http://www.PatentsView.org) – A data visualization and analysis platform build on a database that longitudinally links inventors, their organizations, locations, and overall patenting activity. The data visualization tool, bulk data downloads, and flexible API enable a broad spectrum of users to explore the dynamics of patenting activity over time and space.

[www.Lens.org](http://www.Lens.org) – The Lens serves nearly all of the patent documents in the world as open, annotatable digital public goods that are integrated with scholarly and technical literature along with regulatory and business data. The [PatSeq](https://www.lens.org/lens/bio) facility provides multiple resources for exploring patent documents with biologics and sequence data.

The [Patent Examination Research Dataset](http://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair) (PatEx) contains detailed information on publicly viewable patent applications filed with the USPTO through December 2015 in flat .csv files. The data are sourced from the Public Patent Application Information Retrieval system (Public PAIR).

The National Institutes of Health (NIH) [RePORTER](https://projectreporter.nih.gov/reporter.cfm) (query and export) and NIH [ExPORTER](https://exporter.nih.gov/) (raw data download) are excellent resources for information on NIH federal grants. We already merged the Cancer Moonshot Patent Data with the NIH RePORTER patent information, providing the NIH Federal grant number for matches. These grant numbers can be used to retrieve additional data from NIH ExPORTER on the researchers, the project requirements, academic journal publications, and clinical studies.

Federal [RePORTER](https://projectreporter.nih.gov/reporter.cfm) (query and export) and Federal [ExPORTER](https://exporter.nih.gov/) (raw data download) provide detailed information on research projects funded by multiple federal government entities, including Agency for Healthcare Research and Quality (AHRQ), Center for Disease Control and Prevention (CDC), Food and Drug Administration (FDA), NIH, and U.S. Department of Veterans Affairs (VA). Please note that patent information is only available for projects funded by: AHRQ, CDC, FDA, NIH, and VA.

[FDA Orange Book](http://www.accessdata.fda.gov/scripts/cder/ob/): Approved Drug Products with Therapeutic Equivalence Evaluations – We merged the Cancer Moonshot Patent Data with the FDA Orange Book patent data, providing the FDA application number for matches. These application numbers can be used to extract additional information from FDA on approved drugs and their clinical trial and approval process.

1. Are there any restrictions/limitations on the kind of data which can be brought in to supplement the dataset you provided?

There are no restrictions or limitations with respect to data participants use to supplement the Cancer Moonshot Patent Data. We prefer participants utilize open data sources, but that is not a requirement.

1. ­Could I map the selected patents with the Institute of Electrical and Electronics Engineers (IEEE) databases? ­

Yes.

1. Is there a source for PTAB data in bulk format?

Yes, https://developer.uspto.gov/product/ptab-andor-prps-weekly

1. Is it possible to link the dataset to the PAIR data so that one can tell which technologies have not been abandoned?

The Cancer Moonshot Patent Data can be merged to the [Patent Examination Research Dataset](http://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair) (PatEx) via the patent application number to determine the status of pending applications as of December 31, 2015.

1. It would be great to indicate which issued patents were subject to 101 rejections.­

The USPTO has not publically released metadata capturing which patents have been subject to 101 rejections. Information about the 101 rejections are embedded within the text of the Office Action images, which can be downloaded individually from the Image File Wrapper of a specified case via Public PAIR.

1. ­Is there an indication which issued patents were subject to non-publication requests? ­­

The [WIPO Standard ST.16 codes](https://www.uspto.gov/learning-and-resources/support-centers/electronic-business-center/kind-codes-included-uspto-patent) (kind codes) include a letter, and in many cases a number, used to distinguish the kind of patent document (e.g., publication of an application for a utility patent, utility patent, plant patent, design patent, etc.) and the level of publication (e.g., first publication, second publication, or corrected publication). Kind codes are the last one or two digits of the “Patent\_or\_Publication\_ID” field in the Cancer Moonshot Patent Data. The kind code “B1” indicates a patent grant not previously published as a pre-grant publication. Note, however, that a “B1” kind code does not necessarily mean that the patent was subject to a non-publication request. It only means that the patent was not previously published, which may occur if an application is allowed and issued within 18 months of filing.

The Image File Wrapper of a select case in Public PAIR will include the code "PG.NONPUB.RQ" for the document requesting non-publication. It may include the code "RESC" for rescinding of a non-publication request.

1. ­What is the hopeful outcome of the Cancer Moonshot Challenge? What does the USPTO hope to get out of this challenge? ­

Data visualizations and stories to empower the federal government—as well as the medical, research, and data communities—to make more precise funding and policy decisions based on the commercialization lifecycle of the most promising treatments, while maximizing U.S. competitiveness in cancer investments.

1. Have the patent applications that have been granted been included in the dataset?

The Cancer Moonshot Patent Data includes both patent applications and granted patents. If a patent had previously been published as a pre-grant publication, the pre-grant publication was included in the data and will have the same “Application\_Number” value as the patent.

Some patents may not have a corresponding pre-grant publication because the patent application was filed before November 29, 2000, had been subject to a non-publication request, or had been granted before the 18-month publication date (and did not a have an early publication request).

1. Why have the claims not been searched and only title and abstracts? ­

We initially performed keyword searches against the full text of U.S. patent documents but found that queries resulted in a large number of false positives. This is because claims can include clauses describing a broad range of possible applications of the invention. By limiting our search to titles and abstracts, we ensure our results include cancer-specific innovations and excludes patent documents that only reference a cancer-relevant term in the detailed description of the invention. Nevertheless, participants are welcome to extend the dataset to produce comparison visualizations having more recall (fewer false negatives).

1. ­What sequence listing is included in data feed? ­

A Boolean flag in the Cancer Moonshot Patent Data indicates whether a patent contains a sequence.

For more information about the actual sequences contained in the patent document, see the [PatSeq](https://www.lens.org/lens/bio) feature of [www.Lens.org](http://www.Lens.org), the USPTO official search site, or the National Center for Biotechnology Information (NCBI).

NCBI hosts patented sequence information (but not sequences in Pre-Grant Publications). To search sequences in the NCBI website, see <http://www.ncbi.nlm.nih.gov/nuccore/?term=gbdiv_pat%5Bprop%5D>. For more information about patent searching on NCBI, see: <http://www.ncbi.nlm.nih.gov/education/patent_and_ip_faqs/>.