

FH Data Science Lab DMSP Template Text

For latest versions of docs <https://hutchdatascience.org/projects/nihdatasharing/>
To be notified of relevant updates sign up: <https://fhdata.substack.com/>

Disclaimer: These materials have been developed at Fred Hutch for the purpose of assisting our faculty scientists with the NIH data sharing policy requirements enacted in 2023. While we are confident in the content of these materials, we make no guarantee that they will fully comply with NIH's policy. We will revise these materials on an on-going basis, as appropriate.

This document is intended to be an evolving warehouse of text that can serve as a template for a variety of data types and conditions that can be a basis for developing your NIH DMS Plan. Simply remove the text that doesn't apply to your situation and then edit the text to match your plans. If you want to read more about a section, the links will take you to the bottom of this document where we have curated additional advice and details about what is required in each section. This material has been collected and consolidated from many groups at Fred Hutch including Shared Resources, OSR, and other partners.

Template Data Management and Sharing Plan (copy/paste/delete!)**Element 1: Data Type****A. Types and amount of scientific data expected to be generated in the project**

Our proposal will generate raw data of the following types and sizes:

Technology: We will generate data using [insert technology description]. **File Type:** Data for this study will generate [insert raw data file description]. The amount of data generated per sample is [insert average file size] **Number of files:** We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

Our proposal will generate processed data of the following types and sizes:

Technology: We process the [insert technology type corresponding to raw data type above] using [insert brief pipeline description here]. **File Type:** The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size] **Number of files:** We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

Shared Resources Examples:**ANTIBODY TECHNOLOGY CORE:**

We will collect raw data describing the production and characterization of custom antibodies in tabular format using Excel files. The total amount of data total amount of data describing antibody production is <10MB.

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

CELLULAR IMAGING CORE:

We will collect raw data using [select applicable technologies: confocal microscopy, widefield microscopy, TIRF microscopy, light-sheet microscopy, super-resolution microscopy, gel and blot scanning]. Data for this study will generate image files in [select applicable file format: .tif, .czi, .lsm, .nd2, .ims, .lif] format. The amount of data generated per sample is [insert average file size]. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

collected] for a total data volume of [multiply file size and total number of files].

ELECTRON MICROSCOPY CORE:

We will collect TEM images using ThermoFisher Talos L120C TEM instrumentation with ThermoFisher TEM Imaging and Analysis software (TIA) (version 5.3 SP1) or Leginon (version 3.6). Data for this study will generate high-resolution images in .ser, .emi, TIFF formats (TIA software), and .mrc (Leginon). The average amount of image data generated per sample is 1.6 GB (TIA), or 50 MB/image (Leginon). We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

We will collect SEM images using JEOL JSM 6610LV SEM instrumentation and JEOL SEM Control User Interface (version 3.11). Data for this study will generate high resolution images in TIFF format. The average amount of image data generated per sample is 480 MB. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

We will collect cryoEM images using ThermoFisher Glacios Cryo-TEM instrumentation with SerialEM (version 4.0). Data for this study will generate high-resolution images in .tif format. The amount of data generated per movie is 900 MB. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the .tif files using WARP or cryoSPARC. The data processing will result in .mrc. The amount of data generated per image is 100-500 MB. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

EXPERIMENTAL HISTOPATHOLOGY CORE:

We will collect [select: 20X, 40X] [select: brightfield, fluorescent] images using the Aperio platform. Data for this study will generate .svc or .svn files. The amount of data generated per sample is 1-4GB per image.

-or-

We will collect/produce MOTIF images (imaging and spectral unmixing) using the Polaris automated quantitative pathology imaging system. Data for this study will generate .qptiff or .tiff files. The amount of data generated per sample is approximately 18GB per image.

-or-

We will collect/produce [20X 5x10 mm area/40X 3x5 mm area] Vectra images using [insert technology]. Data for this study will generate .qptiff, .tiff, or .im3 files. The amount of data generated per sample is approximately 27.5GB per image.

-or-

We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

FLOW CYTOMETRY CORE:

We will collect raw data using flow cytometers from BD Biosciences and/or Sony Biotechnology. Data for this study will generate Flow Cytometry Standard files in FCS format, widely adopted and maintained by the International Society for Advancement of Cytometry (ISAC). The amount of data generated per sample is approximately 10MB. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the raw FCS using software such as FloJo to yield summary tables of comparable cell groupings. The data processing will result in tabular data in CSV/TSV format. The amount of analyzed data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

GENOMICS & BIOINFORMATICS CORES:

We will collect raw data using sequencing instrumentation such as the Illumina NextSeq or PacBio Sequel IIe. Primary data for this study will be generated in FASTQ or BAM format. The amount of data generated per sample ranges from 500MB to 20GB depending on the assay type and yield. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We will process the raw FASTQ/BAM data using bioinformatics software to be described in all associated publications. The data processing will result in structured data in tabular (CSV/TSV) or standardized genomic file formats (e.g. FASTA, BAM, VCF, HDF5). The amount of derived data generated per sample ranges from 100MB to 1GB depending on the specific analysis. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

IMMUNE MONITORING CORE:

We will collect assay results for clinical trials using [insert technology description]. Data for this study will generate Excel

files and [insert raw data file description]. The amount of data generated per sample is [insert average file size]. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

LARGE ANIMAL FACILITY CORE:

We will collect medical records and clinical pathology data from veterinary research involving large animals. Data for this study will generate text documentation as well as tabular data. The amount of data generated per sample is less than 1MB, with the total volume of data collected not to exceed 10MB in aggregate file size.

PRECLINICAL IMAGING CORE (IVIS):

We will collect *in vivo* imaging data using Perkin Elmer IVIS Spectrum instrument. Data for this study will generate image data in TIFF, PNG, and text file formats. The amount of data generated per sample is typically 6MB. We anticipate collecting data from [insert total number of samples/files to be collected – typically 12 files per session] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

PRECLINICAL IMAGING CORE (MicroCT):

We will collect raw data using Micro-Computed Tomography (micro-CT) on the Quantum GX2 platform. Data for this study will generate images in the Digital Imaging and Communications in Medicine (DICOM) format, and TIFF images. The amount of data generated per sample is 277 or 140 MB for DICOM or TIFF files, respectively. We anticipate collecting data from [insert total number of samples/files to be collected – typically 30 files per session] for a total data volume of [multiply file size and total number of files]. We process the raw DICOM/VOX files using microCT viewing software to generate static images from slices of the data. The data processing will result in high resolution images in JPEG/PNG/BMP format. The amount of data generated per timepoint is 277 MB (DICOM export) * [number of mice imaged per time point] or 140 MB (tiff export) * [number of mice imaged per time point]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

PRECLINICAL IMAGING CORE (MRI):

We will collect data using a 7T Dry Magnet Magnetic Resonance Imaging (MRI) instrument, generating files in the Digital Imaging and Communications in Medicine (DICOM) format. The amount of data generated per sample is 15MB. We anticipate collecting data from [insert total number of samples/files to be collected – typically 20 files per session] for a total data volume of [multiply file size and total number of files], per timepoint. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

PRECLINICAL MODELING CORE:

We will collect raw data using [insert technology description]. Data for this study will generate [insert raw data file description]. The amount of data generated per sample is [insert average file size]. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the [insert raw data type] using [insert brief pipeline/technology description]. The data processing will result in [insert processed data file description]. The amount of data generated per sample is [insert average file size]. We anticipate generating [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

PROTEOMICS CORE:

We will collect raw proteomics data using tandem (MS/MS) mass spectrometry. Data for this study will generate .raw files with file sizes that range from 0.5 to 2 GB. We anticipate collecting data from [insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files]. We process the raw mass spectrometry files using Thermo Scientific Proteome Discoverer v2.5 and provide analyzed results using the Proteome Discoverer viewer. The data processing will result in pdResult, Excel, and image files generating 1-30 GB. We anticipate generating

[insert total number of samples/files to be collected] for a total data volume of [multiply file size and total number of files].

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

SMALL ANIMAL FACILITY CORE:

We will collect raw data using *in vivo* imaging technology, hematology analyzers, and spreadsheets (Excel). Data for this study will generate both bioluminescence images (in TIFF format) as well as tabular summaries of hematology results and breeding records (in DOC and XLS format). The amount of data generated per sample is ~10MB for a typical image. We anticipate collecting data from ~150 sessions annually for a total data volume of 1.5 GB.

THERAPEUTIC PRODUCTS CORE:

We will collect raw data using automated cell counters and [insert additional technologies]. Data for this study will generate tabular data in Excel format quantifying the yield and purity of cellular products which have been generated. The amount of data generated per sample is no more than 1MB, and we anticipate that the total aggregate file size will be no more than 10MB.

-or-

Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.

B. Scientific data that will be preserved and shared, and the rationale for doing so

Our proposal will preserve and share the following types of data

Raw data: [insert technologies from list above to be shared] to facilitate re-analysis and re-use of the data by other investigators.

Processed data: [insert processed data to be shared] to facilitate re-analysis and re-use of the data by other investigators.

Our proposal will not preserve and share the following data types

Raw data: [insert technologies from list above] will not be shared because they are Level 0 data (<https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/data-submission-and-release-expectations>) and can only be opened in a limited number of non-open source software programs.

Raw data: [insert technologies from list above] will not be shared because they are Level 0 data (<https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/data-submission-and-release-expectations>) and can only be easily opened in proprietary, licensed viewing softwares.

Raw data: [insert technologies from list above] will not be shared because they are Level 0 data (<https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/data-submission-and-release-expectations>); are very large and are only needed for advanced data processing or to reconstruct the scientifically accepted raw data type [insert raw data type here].

Raw data: [insert technologies from list above] will not be shared because they are Level 1 data (<https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/data-submission-and-release-expectations>); are very large and are only needed for advanced data processing or to reconstruct the scientifically accepted raw data type [insert raw data type here].

Raw data/processed data: [insert technologies from list above] will not be shared because the IRB for this protocol does not include consent for public data sharing.

Raw data/processed data: [insert technologies from list above] are not suitable to be shared in identified form due to IRB restrictions. However, de-identified data with randomly generated participant or sample

IDs will be applied to the de-identified data.

Raw data/processed data: [insert technologies from list above] are not suitable to be shared due to sovereignty restrictions related to individuals from the population sampled.

Shared Resources Examples:

ANTIBODY TECHNOLOGY CORE:

Excel files describing antibody production will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

CELLULAR IMAGING CORE:

Raw and processed images will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

ELECTRON MICROSCOPY CORE:

Raw and processed EM images will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

EXPERIMENTAL HISTOPATHOLOGY CORE:

[Aperio brightfield images/Aperio fluorescent images/MOTIF images/Vectra images] will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

FLOW CYTOMETRY CORE:

Raw FCS and processed CSV/TSV data will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

GENOMICS & BIOINFORMATICS CORES:

Raw nucleotide sequence data in FASTQ format and key derived results, including alignments in BAM format or variant calls in VCF format for example, will be preserved and shared through appropriate public repositories such as GEO or dbGaP. De-identified metadata describing the specimens will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

IMMUNE MONITORING CORE:

[Insert raw and processed data types] will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

LARGE ANIMAL FACILITY CORE:

Medical records and clinical pathology data will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

PRECLINICAL IMAGING CORE (IVIS):

Imaging data of IVIS bioluminescence/biofluorescence will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

PRECLINICAL IMAGING CORE (MicroCT):

Processed MicroCT images in DICOM or TIFF format will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators. DICOM or TIFF (select one) files will be shared due to ease of access and the ability to open these files in many different, commonly available software applications.

The raw data (VOX files) are not shared as the data is very large and unusable in most commonly used software applications.

PRECLINICAL IMAGING CORE (MRI):

DICOM files will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators, due to the accessibility of viewing DICOM images using many commonly available image viewing applications.

PRECLINICAL MODELING CORE:

[Insert raw and processed data types] will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

PROTEOMICS CORE:

Both raw results in the form of spectra collected by the mass spectrometer as well as analyzed results in the form of detected peptides and proteins will be preserved and shared. By sharing raw data, we provide other researchers with the ability to re-search spectra as new peptide matches become available in protein database repositories. By sharing the peptides and proteins

detected in our analysis of those spectra, we provide other researchers with the processed data which was used to motivate the conclusions drawn from our study.

SMALL ANIMAL FACILITY CORE:

Bioluminescence images as well as hematology results and breeding records will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

THERAPEUTIC PRODUCTS CORE:

[Insert raw and processed data types] will be preserved and shared to facilitate re-analysis and re-use of the data by other investigators.

C. Metadata, other relevant data, and associated documentation

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata here] and will be submitted in accordance with FAIR data principles according to [if it exists insert FAIR standards for data type from here <https://www.nature.com/articles/sdata201618>]

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata here] and will be released in accordance with FAIR data principles in the form of a spreadsheet with consistent sample labels, dates in ISO 8601 format (YYYY-MM-DD), without empty cells, with one data item per cell, organized as a single rectangle (with subjects as rows and variables as columns, and with a single header row), with a corresponding data dictionary. Metadata will be released in raw form without calculations on the raw data files, font color or highlighting as data, with human and machine readable variable names, links to raw data urls for [insert raw data type], saved as a plain text file and uploaded to [insert location where metadata will be deposited]. For more information on data formatting see (<https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989>).

- Information for human data about de-identification
<https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard>

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata] and will be submitted in accordance with FAIR data principles.

-or-

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata] and will be submitted in accordance with FAIR data principles in the form of a spreadsheet with consistent sample labels, dates in ISO 8601 format, without empty cells, with one data item per cell, organized as a single rectangle (with subjects as rows and variables as columns, and with a single header row), with a corresponding data dictionary. Metadata will be released in raw form without calculations on the raw data files, font color or highlighting as data, with human- and machine-readable variable names, links to raw data URLs for [insert raw data type], saved as a plain text file and uploaded to [insert location where metadata will be deposited].

LARGE ANIMAL FACILITY CORE:

Metadata on medical record information will be collected via report features in our core database and will be submitted in accordance with FAIR data principles.

PRECLINICAL IMAGING CORE (IVIS):

Metadata on IVIS imaging will be collected via [insert process for collecting metadata – please note that it is important that whoever acquires the images on the IVIS machine input *all* important sample information into the Living Image software's ACQUIRE window to ensure this data is preserved at the time of acquisition] and will be submitted in accordance with FAIR data principles.

-or-

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata] and will be submitted in accordance with FAIR data principles in the form of a spreadsheet with consistent sample labels, dates in ISO 8601 format, without empty cells, with one data item per cell, organized as a single rectangle (with subjects as rows and variables as columns, and with a single header row), with a corresponding data dictionary. Metadata will be released in raw form without calculations on the raw data files, font color or highlighting as data, with human- and machine-readable variable names, links to raw data URLs for [insert raw data type], saved as a plain text file and uploaded to [insert location where metadata will be deposited].

PRECLINICAL IMAGING CORE (MicroCT):

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata] and will be submitted in accordance with FAIR data principles.

-or-

Metadata on [insert metadata descriptors] will be collected via [insert process for collecting metadata] and will be submitted in accordance with FAIR data principles in the form of a spreadsheet with consistent sample labels, dates in ISO 8601 format, without empty cells, with one data item per cell, organized as a single rectangle (with subjects as rows and variables as columns, and with a single header row), with a corresponding data dictionary. Metadata will be released in raw form without calculations on the raw data files, font color or highlighting as data, with human- and machine-readable variable names, links to raw data URLs for [insert raw data type], saved as a plain text file and uploaded to [insert location where metadata will be deposited].

PRECLINICAL IMAGING CORE (MRI):

Metadata on MRI data collection will be collected including hardware (field strength, manufacturer, model, RF coils), acquisition (pulse sequence, volume of interest locations, repetition time, total number of acquisitions), data analysis methods (analysis software, processing steps, output measure, quantification references) and data quality (reported variables, data exclusion criteria, and quality measures of postprocessing model fitting). Those metadata and associated documentation will be submitted in accordance with FAIR data principles.

PROTEOMICS CORE:

Metadata on biological samples, processing, technical replicates, fractionation, instrumentation, and configuration will be collected using the MAGE-TAB-Proteomics standard (<https://psidev.info/magetab>) and will be submitted in accordance with FAIR data principles.

SMALL ANIMAL FACILITY CORE:

Metadata on animal breeding and health will be collected via the ARRIVE guidelines (Animal Research: Reporting In Vivo Experiments) for describing *in vivo* experiments and will be submitted to public repositories in accordance with FAIR data principles. (<https://doi.org/10.1371%2Fjournal.pbio.3000411>)

Element 2: Related Tools, Software and/or Code

All software developed as a part of this proposal, will be developed in the open on [pick one of GitHub/GitLab/BitBucket], and released as a collection of open source [pick one or more of R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more of Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL]

All software developed as a part of this proposal, will be deposited after publication on [pick one of GitHub/GitLab/BitBucket], and released as a collection of open source [pick one or more of R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more of Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL]

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step by step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

The raw and processed data generated by this study can be accessed and manipulated using standard office suite spreadsheet software.

-or-

All software developed as a part of this proposal will be [developed in the open/deposited after publication] on [pick one: GitHub/GitLab/BitBucket] and released as a collection of open source [pick one or more: R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more: Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL].

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps

to support reproducibility of results.

CELLULAR IMAGING CORE:

The raw and processed data generated by this study can be accessed and manipulated using standard image viewing and processing software such as Fiji (<https://imagej.net/software/fiji/>) or CellProfiler (<https://cellprofiler.org/>).

EXPERIMENTAL HISTOPATHOLOGY CORE:

The raw and processed data generated by this study can be accessed and manipulated using HALO Image Analysis software available at <https://indicalab.com/halo/>.

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing it publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

LARGE ANIMAL FACILITY CORE:

The raw and processed data generated by this study can be accessed and manipulated using standard office suite applications for viewing and manipulating text files and tabular data (e.g. Microsoft Word and Excel).

PRECLINICAL IMAGING CORE (IVIS):

The IVIS data generated by this study can be accessed and manipulated using the open-source Aura software (available at <https://spectralin vivo.com/software/>) or with Living Image (available with a license, more information available at <https://www.perkinelmer.com/lab-products-and-services/resources/in-vivo-imaging-software-downloads.html>):

-or-

All software developed as a part of this proposal will be [developed in the open/deposited after publication] on [pick one: GitHub/GitLab/BitBucket] and released as a collection of open source [pick one or more: R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more: Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL].

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

PRECLINICAL IMAGING CORE (MicroCT):

The raw and processed data generated by this study can be accessed and manipulated using the Quantum GX2 Viewer and Analysis software, as well as Slicer (available at <https://www.slicer.org/>), Analyze 14.0 (available at <https://analyzedirect.com/analyze14/>) or other custom software specific for research (e.g. ImageJ (<https://imagej.nih.gov/ij/download.html>) or microDICOM (<https://www.microdicom.com/>)).

-or-

All software developed as a part of this proposal will be [developed in the open/deposited after publication] on [pick one: GitHub/GitLab/BitBucket] and released as a collection of open source [pick one or more: R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more: Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL].

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

PRECLINICAL IMAGING CORE (MRI):

The data generated by this study can be accessed and manipulated using the microDICOM software (available at <https://www.microdicom.com/>) or ImageJ software (available at <https://imagej.nih.gov/ij/download.html>).

-or-

All software developed as a part of this proposal will be [developed in the open/deposited after publication] on [pick one: GitHub/GitLab/BitBucket] and released as a collection of open source [pick one or more: R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software packages will be formally released via [pick one or more: Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL].

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

PROTEOMICS CORE:

The raw data generated by this study can be accessed with the latest software version of ThermoScientific Freestyle, and processed data can be accessed and manipulated using the desktop software ThermoScientific Proteome Discoverer v2.5 (available at <https://thermo.flexnetoperations.com/control/thmo/login>), as well as the web-based proteomics analysis functionality provided by MassIVE (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>).

-or-

All software developed as a part of this proposal will be [developed in the open/deposited after publication] on [pick one: GitHub/GitLab/BitBucket] and released as a collection of open source [pick one or more: R scripts/R packages/Python Libraries/Jupyter Notebooks/WDL Workflows/Galaxy Workflows/NextFlow workflows] with an MIT license. Software

packages will be formally released via [pick one or more: Bioconductor/CRAN/conda/PyPi/Galaxy Workbench/AnVIL].

-or-

Software used on this proposal is proprietary and there is no mechanism for sharing these software publicly. We will provide step-by-step analysis instructions including complete descriptions of all software versions, parameter settings, intermediate analysis steps, and intermediate calculations. Where possible, we will include screenshots of analysis steps to support reproducibility of results.

SMALL ANIMAL FACILITY CORE:

The raw and processed data generated by this study can be accessed and manipulated using standard desktop software for viewing images and spreadsheets (e.g., Microsoft Excel).

Element 3: Standards

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

The data produced in this project will be collected using the standards established by [insert organization and description]. [Insert additional details as appropriate].

FLOW CYTOMETRY CORE:

The data produced in this project will be collected using the Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) standard established by the International Society for Analytical Cytology Data Standards Task Force (ref: <https://doi.org/10.1002%2Fcyto.a.20623>).

LARGE ANIMAL FACILITY CORE:

The data produced in this project will be collected using the standards established by the Comparative Medicine Shared Resource and ARRIVE Guidelines. Animal Research: Reporting in Vivo Experiments).

PRECLINICAL IMAGING CORE (MicroCT):

The data produced in this project will be collected using the consensus recommendations established by the Minimum Reporting Standards for *in vivo* Magnetic Resonance Spectroscopy (MRSinMRS) (see <https://doi.org/10.1002%2Fnbm.4484>).

PRECLINICAL IMAGING CORE (MRI):

The data produced in this project will be collected using the consensus recommendations established by the Minimum Reporting Standards for *in vivo* Magnetic Resonance Spectroscopy (MRSinMRS) (see <https://doi.org/10.1002%2Fnbm.4484>).

PROTEOMICS CORE:

The proteomics data produced in this project will be collected using the standards established by the ProteomeXchange (<http://www.proteomexchange.org/>), a global consortium supporting open access to proteomics datasets since 2012. Each dataset defined by the ProteomeXchange is comprised of the data described by a single manuscript using the same data workflow (e.g. DDA). All such datasets are assigned a ProteomeXchange PXD identifier (PXD + a six figure integer, for additional details see: <http://www.ebi.ac.uk/miriam/main/collections/MIR:00000513>). Each datasets consists of (1) mass spectrometry output files (raw spectra as well as peak lists), (2) experimental and technical metadata (in the established PX XML format), (3) processed results (peptide/protein identification results at a minimum), and (4) any other associated files.

SMALL ANIMAL FACILITY CORE:

The data produced in this project will be collected using the standards established by the ARRIVE guidelines (Animal Research: Reporting In Vivo Experiments) for describing *in vivo* experiments.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

Sequence level data will be deposited in dbGap within 3 months of data generation and preserved for the duration of the grant funding.

Processed gene-level summaries will be deposited within 3 months of data generation and preserved for the duration of the grant funding.

Sequence level data will be deposited in dbGap at the time of publication and preserved for the duration of the grant funding.

Sequence level data will be deposited in dbGap at the time of publication and preserved according to SRA preservation standards.

Immunohistochemistry image data, as well as other types, can be deposited in HuMAP.
<https://hubmapconsortium.org/> and the portal url is: <https://portal.hubmapconsortium.org/>

- <https://software.docs.hubmapconsortium.org/faq> for how to actually work through submitting data to it.
- There is a DUA to be signed when submitted data to HuMap, so read here to see if those restrictions can be complied with with your datasets:
<https://hubmapconsortium.org/policies/external-data-sharing-policy/> and
https://hubmapconsortium.org/wp-content/uploads/2020/06/DUA_FINAL_2020_02_03_for_Signature.pdf

Metabolomics workbench:

<https://www.metabolomicsworkbench.org/data/DRCCDataDeposit.php>

There is not an existing public repository that complies with FAIR data principles that is suitable for [the data type; quantity; structure or degree of access control appropriate for sensitive data types] that will be generated in this project. Thus, due to restrictions on budgeting for funding data sharing after the termination of our grant, we will leverage our institution's standard data retention policies and storage resources to store our raw and processed datasets generated and/or used during the course of this project. We intend to share data with other investigators via direct contact with the PI due to the lack of publicly available repositories or tools to support FAIR data principles for our datasets.

Shared Resources Examples:

ANTIBODY TECHNOLOGY CORE:

Primary repositories for Excel files will be [insert repositories] for data that can be made publicly available and [insert repositories] for data that require access controls.

CELLULAR IMAGING CORE:

Primary repositories for cellular imaging data will be the Image Data Resource (<https://idr.openmicroscopy.org/>).

ELECTRON MICROSCOPY CORE:

Primary repositories for TEM/SEM data will be the Electron Microscopy Public Image Archive (EMPIAR; <https://www.ebi.ac.uk/empair/>), and CryoEM will be deposited in Electron Microscopy Data Bank (EMDB; <https://www.ebi.ac.uk/emdb/>).

EXPERIMENTAL HISTOPATHOLOGY CORE:

Primary repositories for Aperio brightfield, Aperio fluorescent, MOTIF, and Vectra images will be [insert repositories] for data that can be made publicly available and [insert repositories] for data that require access controls.

FLOW CYTOMETRY CORE:

The primary repository for FCS data will be the FlowRepository (<https://flowrepository.org/>).

GENOMICS & BIOINFORMATICS CORES:

Primary repositories for raw sequence data will be the Gene Expression Omnibus for data that can be made publicly available and dbGaP for data that require access controls. Both repositories are backed by the Sequence Read Archive (SRA) for storage of raw sequence data, typically in FASTQ format, although uBAM files may be submitted instead. In addition to raw data, both GEO and dbGaP may accept derived results (e.g., gene-barcode matrices, CRISPR sgRNA counts, etc.). Key derived results will accompany and be linked to corresponding raw data using standard SRA metadata spreadsheets. These data sharing repositories are managed and supported by the National Library of Medicine, and there is no current charge for submission or data hosting.

IMMUNE MONITORING CORE:

Primary repositories for [insert data type] will be [insert repositories] for data that can be made publicly available and [insert repositories] for data that require access controls.

LARGE ANIMAL FACILITY CORE:

Primary repositories for medical records will be web-based software and internal server backups for raw data that require access controls.

PRECLINICAL IMAGING CORE (IVIS):

Primary repositories for IVIS datasets and associated metadata will be The Cancer Imaging Archive (TCIA; <https://www.cancerimagingarchive.net/>), as well as the Imaging Data Commons (IDC; <https://datacommons.cancer.gov/repository/imaging-data-commons>), both maintained by the NIH/NCI.

PRECLINICAL IMAGING CORE (MicroCT):

Primary repositories for MRI datasets and associated metadata will be The Cancer Imaging Archive (TCIA; <https://www.cancerimagingarchive.net/>), as well as the Imaging Data Commons (IDC; <https://datacommons.cancer.gov/repository/imaging-data-commons>), both maintained by the NIH/NCI.

PRECLINICAL IMAGING CORE (MRI):

Primary repositories for MRI datasets and associated metadata will be The Cancer Imaging Archive (TCIA; <https://www.cancerimagingarchive.net/>), as well as the Imaging Data Commons (IDC; <https://datacommons.cancer.gov/repository/imaging-data-commons>), both maintained by the NIH/NCI.

PRECLINICAL MODELING CORE:

Primary repositories for [insert data type] will be [insert repositories] for data that can be made publicly available and [insert repositories] for data that require access controls.

PROTEOMICS CORE:

Primary repositories for raw data will be MassIVE (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>) and/or PRIDE (<https://www.ebi.ac.uk/pride/archive/>). MassIVE is a community resource developed by the NIH-funded Center for Computational Mass Spectrometry to promote the global, free exchange of mass spectrometry data. MassIVE datasets can be assigned ProteomeXchange accessions to satisfy publication requirements. The PRoteomics IDentifications (PRIDE) Archive database is a centralized, standards compliant, public data repository for mass spectrometry proteomics data, including protein and peptide identifications and the corresponding expression values, post-translational modifications and supporting mass spectra evidence (both as raw data and peak list files). PRIDE is a core member in the ProteomeXchange (PX) consortium, which provides a standardized way for submitting mass spectrometry-based proteomics data to public-domain repositories. Datasets are submitted to ProteomeXchange via PRIDE and are handled by expert bio-curators. All PRIDE public datasets can also be searched in ProteomeCentral, the portal for all ProteomeXchange datasets.

SMALL ANIMAL FACILITY CORE:

The primary repository for breeding records and hematology results will be internal shared databases, IVIS data can be accessed using the open-source Aura software (available at <https://spectralin vivo.com/software/>) or with Living Image (available with a license, more information available at <https://www.perkinelmer.com/lab-products-and-services/resources/in-vivo-imaging-software-downloads.html>

THERAPEUTIC PRODUCTS CORE:

Primary repositories for [insert data type] will be [insert repositories] for data that can be made publicly available and [insert repositories] for data that require access controls.

B. How scientific data will be findable and identifiable:

Shared Resources Examples:

ANTIBODY TECHNOLOGY CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

CELLULAR IMAGING CORE:

The Image Data Resource (IDR) provides stable IDs to individual studies. Primary references would be to a IDR study accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

ELECTRON MICROSCOPY CORE:

EMPIAR and EMDB provide stable IDs to EM datasets and derived molecular structures. Primary references would be to an EMD or EMPIAR dataset accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

EXPERIMENTAL HISTOPATHOLOGY CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

FLOW CYTOMETRY CORE:

The FlowRepository provides stable IDs to data collections associated with individual experiments or publications. Primary references would be to a Repository ID. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

GENOMICS & BIOINFORMATICS CORES:

The SRA, GEO, and dbGaP repositories provide stable IDs to experiments (BioProject), sequencing data (SRR), and biological specimens (BioSample). Primary references would be to a GEO series accession or BioProject experiment accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

IMMUNE MONITORING CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

LARGE ANIMAL FACILITY CORE:

Summaries of relevant data will be published. The medical record datasets submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

PRECLINICAL IMAGING CORE (IVIS):

Both TCIA and the IDC provide stable IDs to dataset collections representing the complete data collected for a particular publication. Primary references would be to a TCIA/IDC collection accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

PRECLINICAL IMAGING CORE (MicroCT):

Both TCIA and the IDC provide stable IDs to dataset collections representing the complete data collected for a particular publication. Primary references would be to a TCIA/IDC collection accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

PRECLINICAL IMAGING CORE (MRI):

Both TCIA and the IDC provide stable IDs to dataset collections representing the complete data collected for a particular publication. Primary references would be to a TCIA/IDC collection accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

PRECLINICAL MODELING CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

PROTEOMICS CORE:

MassIVE and PRIDE provide stable IDs to experiment collections, which are associated in turn with metadata describing the experiment as well as the raw and processed data which was submitted. Any publications associated with an experiment can be listed directly, making the scientific data discoverable by searching for the associated publication. Primary references would be to a dataset record available in MassIVE (<https://massive.ucsd.edu/ProteoSAFe/datasets.jsp>) or PRIDE (<https://www.ebi.ac.uk/pride/archive>). The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

SMALL ANIMAL FACILITY CORE:

Summaries of relevant data will be published. The breeding and hematology datasets submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project. Both TCIA and the IDC provide stable IDs to dataset collections representing the complete data collected for a particular publication. Primary references would be to a TCIA/IDC collection accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

THERAPEUTIC PRODUCTS CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

C. When and how long the scientific data will be made available:

Shared Resources Examples:

ANTIBODY TECHNOLOGY CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

CELLULAR IMAGING CORE:

Cellular imaging data will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

ELECTRON MICROSCOPY CORE:

TEM/SEM/CryoEM datasets will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

EXPERIMENTAL HISTOPATHOLOGY CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

FLOW CYTOMETRY CORE:

FCS data will be deposited in the FlowRepository [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

GENOMICS & BIOINFORMATICS CORES:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

IMMUNE MONITORING CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

LARGE ANIMAL FACILITY CORE:

Data will be made available at the time of publication.

PRECLINICAL IMAGING CORE (IVIS):

IVIS will be deposited in TCIA/IDC [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

PRECLINICAL IMAGING CORE (MicroCT):

MicroCT data will be deposited in TCIA/IDC [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

PRECLINICAL IMAGING CORE (MRI):

MRI data will be deposited in TCIA/IDC [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

PRECLINICAL MODELING CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

PROTEOMICS CORE:

Raw spectra as well as processed peptide and protein data will be deposited in MassIVE or PRIDE [at the time of publication] and preserved for the duration of the grant funding.

SMALL ANIMAL FACILITY CORE:

Breeding records and hematology results will be made available at the time of publication and preserved for the duration of the grant funding. IVIS will be deposited in TCIA/IDC and preserved indefinitely.

THERAPEUTIC PRODUCTS CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

[Insert raw data type] will not be shared because they are [Level 0/Level 1] data and [can only be opened in a limited number of open-source software programs/proprietary, licensed viewing software].

- or -

[Insert raw data type] will not be shared because they are [Level 0/Level 1] data, are very large, and are only needed for advanced data processing or to reconstruct the scientifically accepted raw data type [insert raw data type here].

- or -

[Insert raw/processed data type] will not be shared because the IRB for this protocol does not include consent for public data sharing.

- or -

[Insert raw/processed data type] are not suitable to be shared in identified form due to IRB restrictions. However, de-identified data with randomly generated participant or sample IDs will be shared.

- or -

[Insert raw/processed data type] are not suitable to be shared due to sovereignty restrictions related to individuals from the population sampled.

LARGE ANIMAL FACILITY CORE:

Raw data will not be shared because they can only be opened in a limited number of open-source software programs/proprietary, licensed viewing software.

SMALL ANIMAL FACILITY CORE:

We do not anticipate any significant factors limiting the access, distribution, or reuse of IVIS scientific data generated for this project. Breeding and hematology records will be retained in internal databases. Relevant information will be published.

B. Whether access to scientific data will be controlled:

Restricted data that is determined to be potentially identifying through indirect or deductive disclosure will only be provided under restricted data use agreements to users who demonstrate a valid research need and meet conditions of use. Restricted data will only be available for x years after the primary study completion date in compliance with the Fred Hutch policies.

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

All requests for the [raw/processed data described above] that is stored in [insert repository] will be [submitted to and processed by the NIH-designated data repository under their "controlled access" process].

PROTEOMICS CORE:

All proteomics datasets stored in MassIVE or PRIDE will be available to requesting researchers without the need for approval.

SMALL ANIMAL FACILITY CORE:

Scientific data will be made openly accessibly upon submission to a public data repository, without any subsequent requirement for access approval by requesting researchers.

THERAPEUTIC PRODUCTS CORE:

[Insert repository names] provide stable IDs to [insert levels such as project accession, SRA read accession, sequencing platform, etc.]. Primary references would be to a GEO series accession or SRA run accession. The dataset records submitted for this project will additionally be listed in the Data Availability sections of all manuscripts published as part of this project.

C. Protections for privacy, rights, and confidentiality of human research participants:

Access will be controlled based on compliance with patient consent and IRB protocol requirements. All deidentified study data will be made available as public use data to the research community via clinicaltrials.gov and peer-reviewed publications.

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

In order to achieve our goal of data sharing with the research community while not violating assurances and rights of study participants, we will create a dataset for sharing that (1) excludes participants whose consent forms specifically state that their data will not be shared outside of the study team, and (2) incorporates standard blurring or masking techniques for demographic, phenotypic, and descriptive variables so as to reduce risks of identifiability and/or confidentiality violation.

SMALL ANIMAL FACILITY CORE:

Data will not be collected from human research participants.

THERAPEUTIC PRODUCTS CORE:

[Insert data type] will be deposited in [insert repository] [insert timeframe: within X months of data generation or at the time of publication] and preserved for [insert time frame, such as the duration of the grant funding].

Element 6: Oversight of Data Management and Sharing

Oversight for, management of and compliance with this data management and sharing plan will be performed by [insert name], the [insert role] for this project. We will review our plan and identify any datasets that need to be shared, processed for sharing or updated [twice annually]. Additionally, we will implement aspects of our plan [when publications are accepted, at the termination of the grant].

Shared Resources Examples:

STANDARD FOR ALL CORES NOT SPECIFICALLY LISTED:

Laboratory staff [insert laboratory staff role here] will review data generated by this project on a [insert timeframe, e.g. monthly] basis, identifying any datasets which should be shared and updating a lab-based repository of scientific datasets. That staff member will also maintain a database of experimental metadata associated with the underlying datasets.

-or-

Datasets generated by this project will be monitored and managed by [insert collaborator here], part of the [x] group at [institution]. They will review and update all experimental metadata on a [insert timeframe] basis and maintain a database of experimental metadata associated with the underlying datasets.

LARGE ANIMAL FACILITY CORE:

Datasets generated by this project will be monitored and managed by Comparative Medicine veterinary staff, part of the Shared Resource group at Fred Hutch. They will review and update all experimental metadata at completion of studies and upon request by internal researchers.

SMALL ANIMAL FACILITY CORE:

Laboratory staff will review data generated by this project during and at end of study and update lab-based repository of scientific datasets. That staff member will also maintain a database of experimental metadata associated with the underlying datasets.