REGIONAL BIOCONTAINMENT LABORATORY DATA MANAGEMENT PLAN

DATA TYPE

TYPES AND AMOUNT OF SCIENTIFIC DATA EXPECTED TO BE GENERATED IN THE PROJECT: SUMMARIZE THE TYPES AND ESTIMATED AMOUNT OF SCIENTIFIC DATA EXPECTED TO BE GENERATED IN THE PROJECT.

This facility regularly produces ELISA, ELISPOT, Flow Cytometry, Gene Expression, Luminex/MBAA, qRT-PCR, RNA Seq, scRNA-seq (10x), MSD, HAI, Neutralizing Antibody Titer, Plaque Forming Unit, HAU, TCID50, and Animal Subject data. These data will be generated from real-time PCR instruments, plate readers, ELISpot readers, flow cytometers, luminex bead/MSD readers, and other scientific instruments**.**

Data will be collected from experiments performed in response to research requests submitted via CoreResearch or performed for laboratories external to Duke.. The following data files will be used or produced during the facility's work: thermocycler data files; ELISpot plate images; FCS flow cytometry files; RBX Luminex bead reader files; MSD reader files; Microsoft Excel files of various types including, but not limited to, well absorbances/fluorescence intensities, spot counts and ELISA metrics; flowjo analysis workspaces; Graphpad Prism files of analyzed data and Excel files of analyzed data.

All assays will be documented in an Electronic Laboratory Notebook (LabArchives) and all raw, transformed and analyzed data files will be attached to the relevant LabArchives record. The LabArchive record will be locked after the assay is complete so that no further edits can be performed. Raw data will be transformed using platform-specific software (CFX Maestro, Gen5, SpectraMax, Bio-Plex Manager, FlowJo, ForeCyte) and Microsoft Excel, and the subsequent processed datasets will be sent to the requestor at the completion of the assay.

All raw and intermediary data files will be available on request. All data will be provided in a format that is compatible with at least one public data repository upon request.

SCIENTIFIC DATA THAT WILL BE PRESERVED AND SHARED, AND THE RATIONALE FOR DOING SO: DESCRIBE WHICH SCIENTIFIC DATA FROM THE PROJECT WILL BE PRESERVED AND SHARED AND PROVIDE THE RATIONALE FORTHIS DECISION.

For data generated for experiments performed in response to a CoreResearch request or other request from investigators outside of Duke, the requestor will be responsible for sharing these data to the repository of their choice.

All data and associated electronic laboratory notebook records will be preserved internally for at least ten years after assay completion. Leftover samples will be preserved for 30 days and then returned to the requestor or destroyed depending on the needs of original requestor.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate interpretation of the data and help requestors comply with data sharing requirements, all metadata required for the Human Immunology Project Consortium (HIPC) data standards will be shared and associated with the relevant datasets. This can include, when appropriate, animal study design, animal subject information, bio-sample information, adverse events, interventions, inclusion criteria, protocols, animal subject assessments, standard curves, control sample information, sample treatments, and reagent metadata.

RELATED TOOLS, SOFTWARE, AND/OR CODE

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

All processed data will be made available in jpg, csv, excel, or txt format and will not require the use of specialized tools to be accessed or manipulated.

Some raw data files will require specialized software for access and manipulation. CFX files will require CFX Maestro software, which can be purchased from Bio-Rad corp. RBX files will require Bio-Plex manager software, which can be purchased from Bio-Rad corp. FCS files and flow cytometry workspace files will require FlowJo software, which can be purchased from Treestar/BD corp. MSD files require MSD Discovery Workbench, which can be purchased from Meso Scale Diagnostics corp.

STANDARDS

STATE WHAT COMMON DATA STANDARDS WILL BE APPLIED TO THE SCIENTIFIC DATA AND ASSOCIATED METADATA TO ENABLE INTEROPERABILITY OF DATASETS AND RESOURCES, AND PROVIDE THE NAME(S) OF THE DATA STANDARDS THAT WILL BE APPLIED AND DESCRIBE HOW THESE DATA STANDARDS WILL BE APPLIED TO THE SCIENTIFIC DATA GENERATED BY THE RESEARCH PROPOSED IN THIS PROJECT. IF APPLICABLE, INDICATE THAT NO CONSENSUS STANDARDS EXIST

The following data standards and ontologies will be used wherever possible to enhance the reusability and interoperability of scientific data: Cell Ontology (CO), Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM), Clinical Measurement Ontology (CMO), Disease Ontology (DO), Gene Ontology (GO), Human Immunology Project Consortium (HIPC), Human Phenotype Ontology (HPO), Immuno Polymorphism Database- ImmMunoGeneTics/Human Leukocyte Antigen Database (IPD-IMGT/HLA), Medical Dictionary for Regulatory Activities (MedDRA), National Cancer Institute Thesaurus (NCIT), National Center for Biotechnology Information (NCBI) Taxonomy, Ontology for Biomedical Investigators (OBI), Protein Ontology (PRO), Uber-Anatomy Ontology (Uberon),and Vaccine Ontology (VO).

DATA PRESERVATION, ACCESS, AND ASSOCIATED TIMELINES

REPOSITORY WHERE SCIENTIFIC DATA AND METADATA WILL BE ARCHIVED: PROVIDE THE NAME OF THE REPOSITORY(IES) WHERE SCIENTIFIC DATA AND METADATA ARISING FROM THE PROJECT WILL BE ARCHIVED;SEESELECTING A DATA REPOSITORY)

All datasets produced by the facility will be shared with the original requestor. It will be the responsibility of the requestor to share his/her data in a public repository.

All raw and internally analyzed data and associated electronic laboratory notebook records will be preserved internally for at least ten years after assay completion.

HOW SCIENTIFIC DATA WILL BE FINDABLE AND IDENTIFIABLE: DESCRIBE HOW THE SCIENTIFIC DATA WILL BE FINDABLE AND IDENTIFIABLE, I.E., VIA A PERSISTENT UNIQUE IDENTIFIER OR OTHER STANDARD INDEXING TOOLS.

The Immunology Database and Analysis Portal (ImmPort) provides metadata, persistent identifiers, and long-term access. The NIH supports this repository. The shared research and clinical data, as well as the analytical tools in ImmPort are available to any researcher after registration. The facility will format all data generated to fit the ImmPort data model to facilitate sharing by the requesting investigator.

WHEN AND HOW LONG THE SCIENTIFIC DATA WILL BE MADE AVAILABLE: DESCRIBE WHEN THE SCIENTIFIC DATA WILL BE MADE AVAILABLE TO OTHER USERS (I.E., NO LATER THAN TIME OF AN ASSOCIATED PUBLICATION OR END OF THE PERFORMANCE PERIOD, WHICHEVER COMES FIRST) AND FOR HOW LONG DATA WILL BE AVAILABLE.

Data will be sent to requestors as soon as it passes all internal quality control and approval processes. Data and associated electronic laboratory records will be stored internally for at least ten years after release.

It will be the duty of requestors to make their data publicly available (or not) to comply with data sharing policy.

ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

FACTORS AFFECTING SUBSEQUENT ACCESS, DISTRIBUTION, OR REUSE OF SCIENTIFIC DATA: NIH EXPECTS THAT IN DRAFTING PLANS, RESEARCHERS MAXIMIZE THE APPROPRIATE SHARING OF SCIENTIFIC DATA. DESCRIBE AND JUSTIFY ANY APPLICABLE FACTORS OR DATA USE LIMITATIONS AFFECTING SUBSEQUENT ACCESS, DISTRIBUTION, OR REUSE OF SCIENTIFIC DATA RELATED TO INFORMED CONSENT, PRIVACY AND CONFIDENTIALITY PROTECTIONS, AND ANY OTHER CONSIDERATIONS THAT MAY LIMIT THE EXTENT OF DATA SHARING. SEE FREQUENTLY ASKED QUESTIONS FOR EXAMPLES OF JUSTIFIABLE REASONS FOR LIMITING SHARING OF DATA.

It will be the responsibility of external requestors to determine if there are factors or limitations that will affect the access, distribution or reuse of the scientific data generated in the RBL.

WHETHER ACCESS TO SCIENTIFIC DATA WILL BE CONTROLLED: STATE WHETHER ACCESS TO THE SCIENTIFIC DATA WILL BE CONTROLLED (I.E., MADE AVAILABLE BY A DATA REPOSITORY ONLY AFTER APPROVAL).

Whether or not access to scientific data the facility produces will be controlled is up to the requestors who submit assay requests. The facility will not share data with anyone outside the Regional Biocontainment Laboratory other than the original requestor and individuals they designate unless required to do so by institutional, regulatory, or legal mandate.

PROTECTIONS FOR PRIVACY, RIGHTS, AND CONFIDENTIALITY OF HUMAN RESEARCH PARTICIPANTS: IF GENERATING SCIENTIFIC DATA DERIVED FROM HUMANS, DESCRIBE HOW THE PRIVACY, RIGHTS, AND CONFIDENTIALITY OF HUMAN RESEARCH PARTICIPANTS WILL BE PROTECTED (E.G., THROUGH DE-IDENTIFICATION, CERTIFICATES OF CONFIDENTIALITY, AND OTHER PROTECTIVE MEASURES).

The facility will not receive and will not request any protected health information, and any protected health information must be permanently removed from sample vials and sample documentation before submission to the facility.

All human samples submitted to the facility must have been collected and shared under an IRB-exempt or IRB-approved study. It is the requestors' responsibility to ensure this requirement is met.

OVERSIGHT OF DATA MANAGEMENT AND SHARING

DESCRIBE HOW COMPLIANCE WITH THIS PLAN WILL BE MONITORED AND MANAGED, FREQUENCY OF OVERSIGHT, AND BY WHOM AT YOUR INSTITUTION (E.G., TITLES, ROLES).

All students, postdoctoral/clinical fellows, research staff, resident trainees, visiting scholars, faculty, and principal investigator(s) should follow the guidelines within this data management plan as contributing members of their assigned role in the service of the Duke Regional Biocontainment Laboratory in the Duke Human Vaccine Institute. Each unit within the RBL shall appoint two individuals to serve as data managers for that unit. These data managers will be in charge of ensuring compliance within their unit.

The following individuals will be responsible for updating and revising the Data Management and Sharing Plan yearly and when necessary.

- Herman Ford Staats, Director Regional Biocontainment Laboratory and Professor of Pathology, Duke University, https://orcid.org/0000-0003-1039-1087, herman.staats@duke.edu
- Nathan L. Lee, Scientific Data Manager II, Duke University, nathan.l.lee@duke.edu

APPENDIX 1: DATA MODELS AND STANDARDS

- ImmPort data model:
    - https://www.immport.org/shared/dataModel

- HIPC Standards:
    - https://docs.immport.org/datasubmission/general/hipcstandards/
- ImmPort Submission Templates:
    - https://docs.immport.org/datasubmission/general/submissiontemplates/
- Ontologies:
    - https://docs.immport.org/home/ontologies/