DHVI - Duke Regional Biocontainment Laboratory

A Data Management Plan created using DMPTool

Creators: Gregory Sempowski, Heather Lynch

Affiliation: Duke University

Funder: Various Sponsors/Cost Recovery Cores

Template: Duke University Data Management Plan Template and Guidance

Project abstract:

This DMP applies to all data generated for Duke RBL sponsored research programs and shared resource Core Units.

Last modified: 04-16-2020

DHVI - Duke Regional Biocontainment Laboratory - Roles and Responsibilities

Roles and Responsibilities

Who will be implementing data management procedures and what are their individual responsibilities?

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 research labs and service cores of the Duke Regional Biocontainment Laboratory in the Duke Human Vaccine Institute.

- **Data Generators -** Students, trainees, laboratory technical staff, team leads/unit managers and PIs. Generators are to contact their assigned project team lead or unit manager with concerns.
- **Data Managers -** Team leads/unit managers and/or PIs. Managers are to contact their assigned PI or the DHVI RQM team with concerns.
- **Statisticians** - Various statistical resources are used by the RBL teams and are assigned/contracted based on expertise, ability and availability by the RBL Team lead/unit manager and/or PI. Statisticians are to contact the project PI or the DHVI RQM team with concerns.

All members of the RBL teams are responsible for maintaining a culture dedicated to the highest ideals of research integrity. This includes but is not limited to:

1. the selection of appropriate methodologies for interrogation of research questions,
2. understanding of the technical limitations of research methods selected,
3. unbiased collection of raw data,
4. proper storage of research data,
5. selection of appropriate methods for analysis and interpretation of results and
6. accurate presentation of analysis/interpretations in all written and verbal scientific communications.

Prior to the initiation of a new research or service project the PI, Unit/Lab Manager, and research support team should establish specific objectives of the project and appropriate methodologies for sample analysis and data procurement. Whenever possible and especially when observer bias may influence results, strategies such as sample ID blinding and randomization should be utilized to facilitate the collection of unbiased results. All lab members are encouraged to voice concerns regarding the collection of unbiased data and propose methods for minimizing bias in all research activities. All experiments and support assays should be designed with specific attention to appropriate controls to facilitate the interpretation of the biological relevance of experimental treatments and assay performance.

Any concerns regarding data integrity should be brought immediately to the attention of the Unit/Lab Manager, PI, RBL Director or DHVI Director.

It is important to note that some projects, grants, contracts, etc. may have specific requirements for data management that are equal to or better than those described here. In these situations, the data management policies of the sponsor, regulatory agency, etc. will supersede those described here.

How will new people be trained and ongoing communication be encouraged amongst all team members with respect to implementing best practices in data management?

Upon joining Duke RBL research teams and/or Core lab units all data generators, managers and statisticians will be trained in best practices in data management. The training process will include instruction on appropriate data management for their position and role in the Duke RBL. The PI and/or Unit/Lab Manager will meet with the new team member during their onboarding process to explain the expectations of data management and research integrity in that laboratory/unit/core. This will include, but is not limited to, a discussion of the specific practices within that lab for archiving and managing data in compliance with the guidelines as described in this DMP. All staff involved in data management activities will be required to acknowledge receipt and understanding of policies and procedures described in this guidance document in paper records or electronic tracking systems (eg. ELN, Q Pulse).

New team members will be asked to consider reviewing the Duke University Library's online data management guidelines and attending an in-person data management Responsible Conduct of Research (RCR) forum through the Library or the ASIST program.

Before departing, team members should organize and index the data and research records for which they have been responsible, and discuss with the team leads/unit managers and/or PIs whether it is permissible to retain any copies. In addition, at least one team lead/unit manager and/or PI will be designated to review files with the soon-to-be-departing person, to ensure there is a good understanding of the organization and location of the remaining data and research records.

DHVI - Duke Regional Biocontainment Laboratory - Research Methods and Data Description

Research Methods and Data Description

What type of research does this plan cover? (i.e., briefly summarize your project scope or research unit's focus area that will be covered under this DMP)?

The Duke Regional Biocontainment Laboratory (RBL) is an unique component of the Duke Human Vaccine Institute (DHVI) in that it includes both independent laboratories running directly funded, hypothesis driven research projects and collaborative research units (shared resources/core labs) performing cost-recovery assays and small animal models.   Prior to the initiation of a new project the PI, Unit/Lab Manager and research support team shall discuss the specific practices, based on the guidelines detailed here, that will be utilized for data management of that project.

How will the research, summarized above, be carried out? (i.e., provide a general, "big picture" overview for the methodologies in practice)

The Duke RBL laboratories and service cores generate a wide-range of data sets from Research Use Only (RUO) assays and animal models categorized as basic research, developed, optimized or qualified. Whenever possible and especially when observer bias may influence results, strategies such as sample ID blinding and randomization should be utilized to facilitate the collection of unbiased results. All lab members are encouraged to voice concerns regarding the collection of unbiased data and propose methods for minimizing bias in all research activities. All experiments and support assays should be designed with specific attention to appropriate controls to facilitate the interpretation of the biological relevance of experimental treatments and assay performance.

What types of data will your research team be generating/using including the format, size, and source?

Herein we define "data" as information directly generated by instruments and/or manually collected by visual observation of the scientific teams. Typical data sets are in the form of Excel tables. Analyses include but are not limited to any data tables and graphics resulting from interpretation of the underlying raw data (by specialized software such as FlowJo or statistical tools such as Prism, SAS, R).

DHVI - Duke Regional Biocontainment Laboratory - Storage and Organization Workflow

Storage and Organization Workflow

Where and how will the data, documentation, and resources be stored while you are actively using the data?

All primary source data including laboratory notebooks, assay record sheets, electronic spreadsheets etc. are the property of the University, not the investigator or issuer of a given notebook, and as such are not to be removed from Duke property. If it is necessary to work from written primary source documents off site a photocopy of the original document must be made. All electronic data can only be accessed via secure connection to the Duke network and should not be stored on any personal devices. All laboratory books and electronic files are open to members of the originating laboratory, PI and relevant RBL/DHVI leadership, unless restricted by regulatory bodies.

Each individual is responsible for their current working lab notebook which should be stored in the lab or at their desk/office and be made available upon request. All completed lab notebooks should be indexed and stored in the lab or with the PI or Unit/Lab manager. Written records and lab notebooks will be kept for at least 10 years. All electronic copies of data are stored on DHVI network servers and located on lab specific network drives, or in a defined archive location, for at least 10 years after it has been published. Data is backed up and managed by Duke Office of Information Technology (OIT).

How will the data, documentation, and resources be organized within and across file systems?

The Duke RBL laboratory and service core units have a comprehensive folder structure on the DHVI N Drive mentioned above. Folder access and privileges are set by DHVI IT team at the written request and approval of the team lead/unit manager and PI. Unit and project specific directories are established such that persons will know how and where to put new data sets and find older datasets and records.

File naming conventions including common/universal naming strategies such as avoiding spaces and special characters are established within each group, unless otherwise dictated by sponsor or regulatory bodies. Tools such as DukeBox and RedCAP are used when appropriate to maintain provenance/version control.

Raw data files created by scientific equipment should be preserved in their unaltered state:

- If an instrument produces a results file from an experimental run, the raw results file should be archived prior to any analysis/manipulation.
- If an instrument produces additional files that describe experimental details aside from the results, those files should also be archived.
- If the data is produced as a hard copy, a scanned PDF should be made and archived prior to analysis.
- If the data consists of observations that are recorded manually, the data should be entered directly into an electronic file (e.g. Word, Excel) that is archives, or should be recorded on paper that is scanned to a PDF and archived.
- All files should be consistently labeled and annotated with Study ID and nature of data set (i.e. raw, post analysis, final)

Analyses

The method used for analyzing data should be described either in a laboratory SOP or by appropriate notations in the laboratory notebook/study plan. Such notations should be sufficient so that someone proficient in the experimental technique being performed could replicate the analysis. It is recommended to retain all intermediate analytical steps, including analyses that are subsequently deemed incorrect.

Tables/Graphics

Tables and graphics produced from analyzed and/or raw data should be saved in the same location as the raw data from which they are derived. It is recommended that intermediate versions of the tables/graphics are also saved. T he software used to create the tables/graphics should be noted in the lab notebook, process sheets, etc. if not otherwise indicated.

DHVI - Duke Regional Biocontainment Laboratory - Documentation and Metadata

Documentation and Metadata

What documentation or metadata will be created during the project and in what format?

When appropriate and allowed, administrative, study-level and file level information will be documented stored with associated data sets, analyses and charts/tables. Each are described below.

- **Administrative documentation** explains the provenance of the data. It includes information on the authors, grant funding, roles and responsibilities of the research team, storage locations, file organization and naming conventions.
- **Study level documentation** includes information on your research methods including data collection methods, processing and analysis decisions to ensure accuracy, completeness and authenticity of research methods and results. Examples of this type of documentation include calculated variables, weighting, and other processes that have been applied to the data in the transformation from raw to analyzed data to reported results.
- **File level documentation** defines the contents of the actual data files; e.g., labeling variables, values, fields, measurement units, etc.

DHVI - Duke Regional Biocontainment Laboratory - Data Sharing and Archiving

Data Sharing and Archiving

What data will you share at the conclusion of your project? Where will you make it available and under what conditions?

Core Service Projects: All primary data, analyzed data, metadata and documentation will be made available to the service requestor and their project PI via Duke email or other mechanism such as DukeBox. The Core will retain these original files/documents as described in Storage and Organization Workflow.

Laboratory Research Projects: Sharing of data generated by sponsored research projects in Duke RBL labs will be dictated by individual program/center/project data sharing plans. In general, we will make our data available through peer-reviewed publications, presentations at scientific meetings and workshops, and through upload to journal or sponsor associated data repositories, such as Immport (NIH/NIAID). In the event a sponsor or journal is unable to host the data sets, then we will use the Duke Research Data Repository which supports data curation through a local, domain agnostic and can facilitate deposit with other Duke affiliated repositories (i.e., Vivli, Qualitative Data Repository, or ICPSR). Once deposited in a repository, embargos might be placed on data access for a certain period of time to honor the researcher's rights of first use (usually no more than 2 years).

If data are from human subjects, Data Managers will ensure consent for data sharing was obtained.

How will you prepare (i.e., "curate") your data to support future access and use?

When applicable, discipline-specific metadata or other forms of end-user documentation (i.e., README files, data dictionaries, instruments, etc.) will be prepared for future use. **Best Practices and Considerations:**

- Consider how you will organize your files for reuse and apply standardized descriptive file names.
- Apply discipline-specific metadata or generate other forms of end-user documentation (i.e., README files, data dictionaries, instruments, etc.).
- If data are restricted or sensitive, develop a de-identification plan that will ensure you follow ethical data sharing practices including compliance with relevant regulations such as the HIPAA Privacy Rule for Protected Health Information.
- Consider transforming data into open or non-proprietary file formats to enhance interoperability, portability, and preservation.
- Consider what other supplementary files should be included alongside the data or in a separate repository (i.e., containers, code files, software, other machine-readable metadata, or physical data resources).

Where will you archive your data at the conclusion of the project to ensure collaborators and Duke University research stakeholders can gain access?

If data are NOT deposited in a permanent repository or archive for broader sharing as discussed above, then data will continue to be secured and accessible during the established retention period stipulated by institutional policies and protocols or external sponsor requirements using DHVI network servers and/or cloud options approved by Duke (i.e., DukeBox, LabArchives, and OSF).

If appropriate, describe what data will be destroyed to comply with legal or policy requirements and how you will dispose of the data?

We do not anticipate that any data will be destroyed. If required to destroy data, then we will work with DHVI IT to do so consistent with Duke policies.