RI-INBRE Bioinformatics Core

A Data Management Plan created using DMPTool

Creators: Laura Bellavia, Laura Bellavia

Affiliation: University of Rhode Island (URI)

Template: National Institutes of Health (NIH)

Last modified: 05-24-2019

Copyright information:

RI-INBRE Bioinformatics Core

Created using the DMPTool. Last modified 05-24-2019

1 of 3

RI-INBRE Bioinformatics Core

Data type: human genomic data and non-human genomic data

Data generated by the RI-INBRE Core Facilities or by individual INBRE investigators will be shared in full compliance with NIH protocols for sharing of large-scale human and non-human genomic data, including genome-wide association studies, metagenomic sequencing, functional genomics data (RNA-seq/microarray, ChIP-seq/epigenomics, proteomics, etc).   RI-INBRE will also require users and investigators to follow established public standards for sharing of small-scale bioinformatic data. This will include deposition of sequences into public databases, collection of detailed metadata for samples, and release of custom code or pipelines used for bioinformatics analysis.

The data expected to be generated or analyzed by INBRE researchers will include but is not limited to nucleotide sequences (including genomic and/or metagenomic data), protein sequences, functional genomics data (RNA-seq, ChIP-seq, etc.) and protein structure data. The Bioinformatics Core will dissemninate data sharing protocols to all researchers to ensure compliance with NIH, INBRE and journal requirements.

It is anticipated that most data will be from non-human organisms, in particularly mouse and microbes. If data is generated or analyzed from human sources, the NIH Genomic Data Sharing Policy for human sequence data will be followed. This will include deposition of experimental/clinical data to the proper repositories and anonymization of the data to comply with the HHS Regulations for the Protection of Human Subjects and the Health
Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

In addition to standard bioinformatics data, the Bioinformatics Core is also compiling a student tracking database to allow for rapid and accurate assessment of outcomes for students working on INBRE projects. This database will be compliant with the The Family Educational Rights and Privacy Act (FERPA) to protect student privacy. The full database limits the type of personal data collected on the students (no SSN, BoD, address, etc.) and will only be available fully to RI-INBRE administrative staff and selectively to RI-INBRE researchers. If the data is shared with the public or other INBRE states, the data will be anonymized to remove identifying information about the students, including removal of non-critical meta data that could be used to infer a student's identity.

Data repositories

Non-Human Sequence Data

All nucleotide sequence data generated by the Bioinformatics Core, analyzed by the Bioinformatics Core, or generated by RI-INBRE researchers will be deposited in the appropriate NIH-compliant databases. Small-scale (e.g. genetic) sequencing or assembled sequences will be deposited in GenBank. Raw sequence data from large-scale projects will be deposited in the Sequence Read Archive. Protein sequence data will be deposited in the Protein database and protein structure data will be deposited in the RCSB protein data bank. Experimental data for functional genomics analyses including transcriptomics (microarray/RNA-seq) and epigenomics (ChIP-seq) will be deposited in the Gene Expression Omnibus database. Researchers may also deposit their data in additional repositories. In order to promote reproducibility of research, all code generated by the Bioinformatic Core or RI-INBRE researchers for analysis of data will be open source and made available to the public through standard sources (e.g. GitHub, SourceForge, etc.) within nine months of validation.

Human Sequence Data

Any sequence data generated from humans will comply with the above requirements. In addition, any data regarding individual human subjects (GWAS, SNP data, precision medicine, etc.) must be anonymized prior to deposition to ensure patient privacy. Researchers should consider the eighteen data elements defined by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) in such cases. Data should be deposited in a controlled access database such as the database of Genotypes and Phenotypes (dbGaP) within 45 days of data generation.

Data submission expectations and timeline

RI-INBRE will require all researchers publishing data using RI-INBRE resources to submit their data as described above prior to publication regardless of the individual data sharing policies of the journals publishing the data. Data should be deposited no later than six months after data submission is initiated.

The Bioinformatics Core and RI-INBRE researchers will also comply with the MIxS (Minimum Information about any (x) Sequence) standards implemented by the Genome Standards Consortium. These protocols include the MIGS (genomes), MIMS (metagenomes) and MIMARKS (marker genes) standards and govern the quality of the sequencing data, methods used to generate and/or assemble the data, and collection of relevant metadata (e.g. environment, biogeographical data, etc.).

Informed consent and institutional certification

The Bioinformatics Core and RI-INBRE researchers will comply with all standards of informed consent for data generated for human subjects as outlined by the NIH and the RI-INBRE institutions and their institutional review boards (IRBs) or equivalent bodies.

Exceptions to data submission expectations

Question not answered.

Intellectual Property

Question not answered.

Created using the DMPTool. Last modified 05-24-2019

3 of 3