Reanalysis of Rat Data from Spacelab Life Sciences 2 (SLS-2) to Reveal Research Gaps in Spaceflight Data

E. Soler[1], C.R. Fowler[1], A. French[2], J. Galazka[3]

[1]Space Life Sciences Training Program, Moffett Field, CA; [2]Bionetics, Moffett Field, CA; and [3]Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA

Abstract

Using and analyzing the legacy data obtained in space life sciences missions has the potential to provide researchers a complete picture of the molecular changes associated with space without further experimentation. This project's objective is to extract, filter, organize, and analyze all *Rattus norvegicus* data and metadata obtained from Columbia's Spacelab Life Sciences 2 (SLS-2, STS-58) mission to explore the ways that we can compile information from model organisms, in our case rats, to create a reliable model to understand biological mechanisms in response to space flight changes. By reusing rare space legacy data coupled with data analysis techniques, we can combine individual preexisting datasets with current ones to gain new, comprehensive insights about the effects of spaceflight on our bodies. Our methods can also lead to the creation of a standardized pipeline that could be applied to other space life science datasets for analysis. In this review, every biological experiment conducted on rats in the SLS-2 Mission was studied with our pipeline to create a new biological library and model that could be used by scientists from around the world to discover novel discoveries and develop new hypotheses from this priceless information without the limitation of the costs of spaceflight experimentation.

Background and Introduction

Purpose

The purpose of this study was to:

1. Find and digitize all data collected during the SLS-2 mission.
2. Organize SLS-2 to allow arbitrary statistical analyses.
3. Create a whole-rat model using all SLS-2 data.

Challenges

1. Multiple datasets appear in different sources
2. Parameters, operationalized variables, standardized way of seeing data sets
3. Not all documents are programmable and standardized
4. Filtering relevant data

Significance

By digitizing, organizing and analyzing all data from SLS-2 we will generate new biological insights and demonstrate the power of reusing legacy data.

Materials and Methods

Data Sources

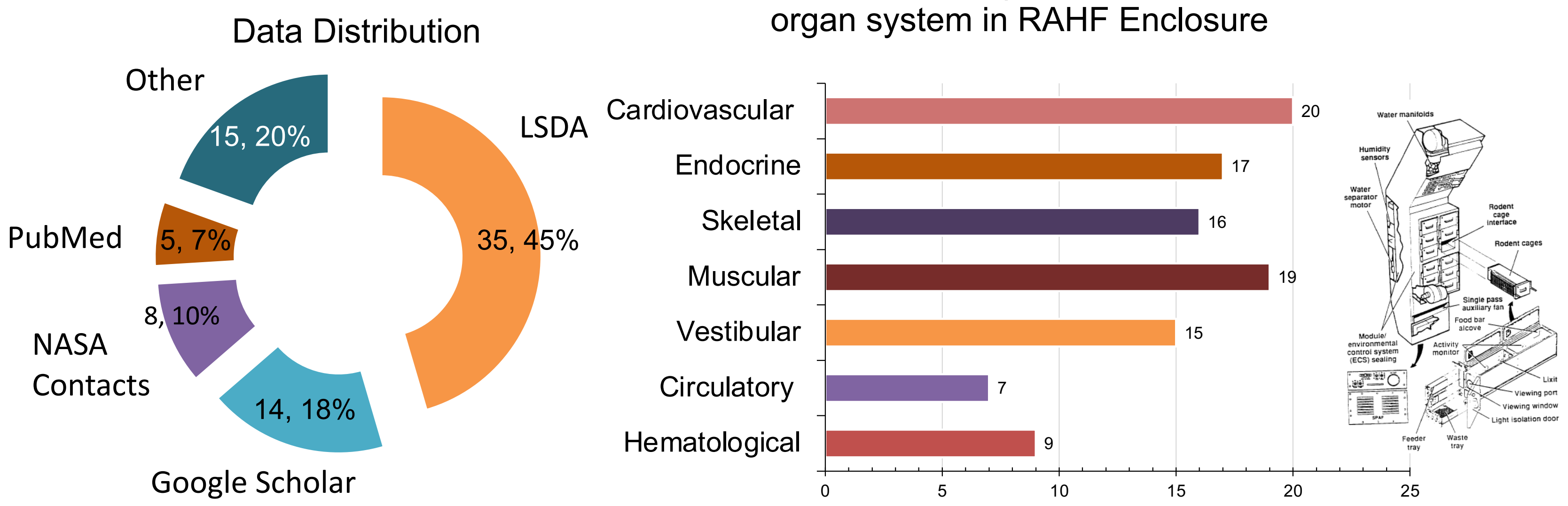
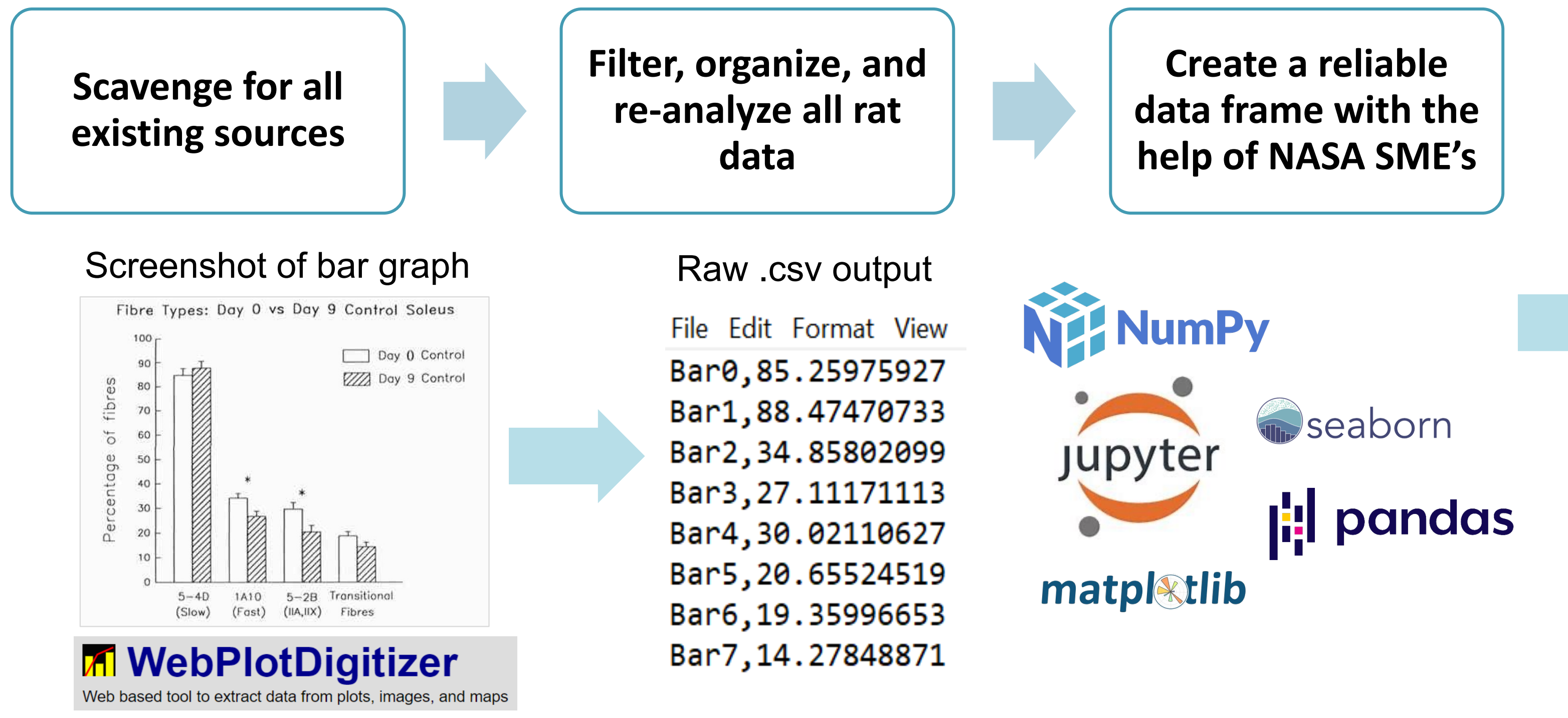


Figure 1: The dataset distribution of acquired data of the SLS-2 mission. Most of the data was from LSDA, then NASA contacts, and then external resources (Google Scholar, PubMed).



```
Bar0,85.25975927
Bar1,88.47470733
Bar2,34.85802099
Bar3,27.11171113
Bar4,30.02110627
Bar5,20.65524519
Bar6,19.35996653
Bar7,14.27848871
```

Figure 2: Workflow for gathering and organizing data before reanalysis.

Results

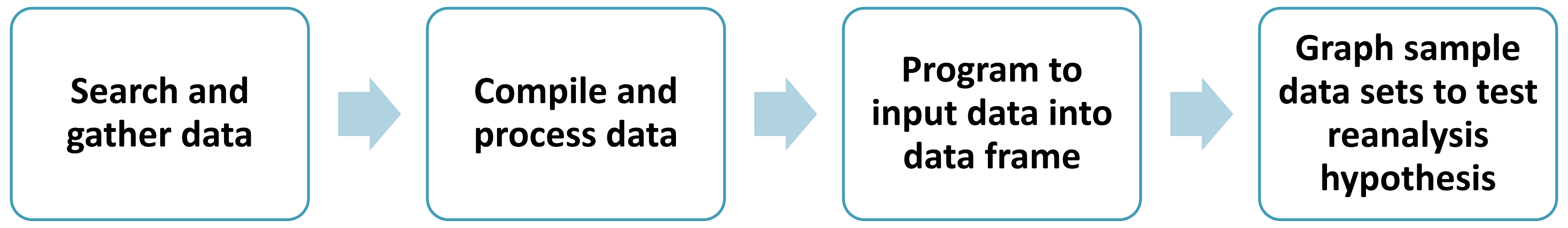


Figure 3: Pipeline that was followed to gather and process data to create reliable dataset to test our reanalysis hypothesis.

SLS-2: Possible changes in protein production in response to spaceflight

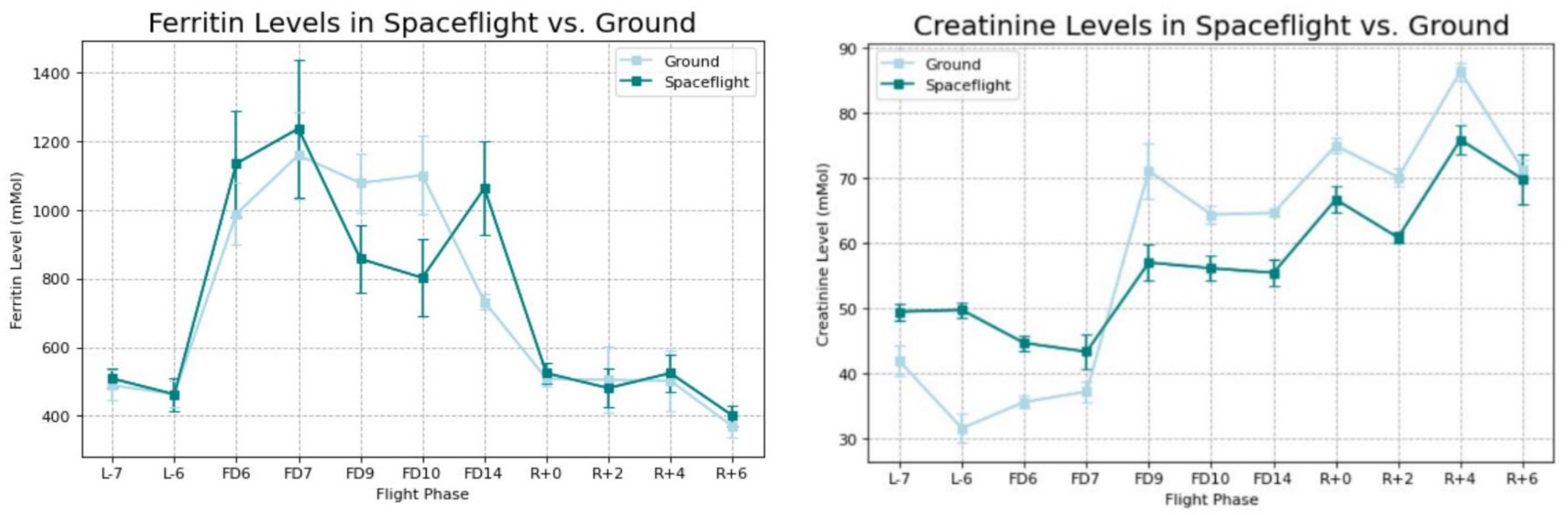


Figure 4: Creatinine, a muscle protein, levels were recorded for pre-launch (L-), spaceflight (FD), and recovery (R+). In spaceflight and the recovery stage, less of the protein is produced when comparing the control and flight. The error bars were calculated with a SEM.

SLS-2: Possible hematoligical changes in response to spaceflight

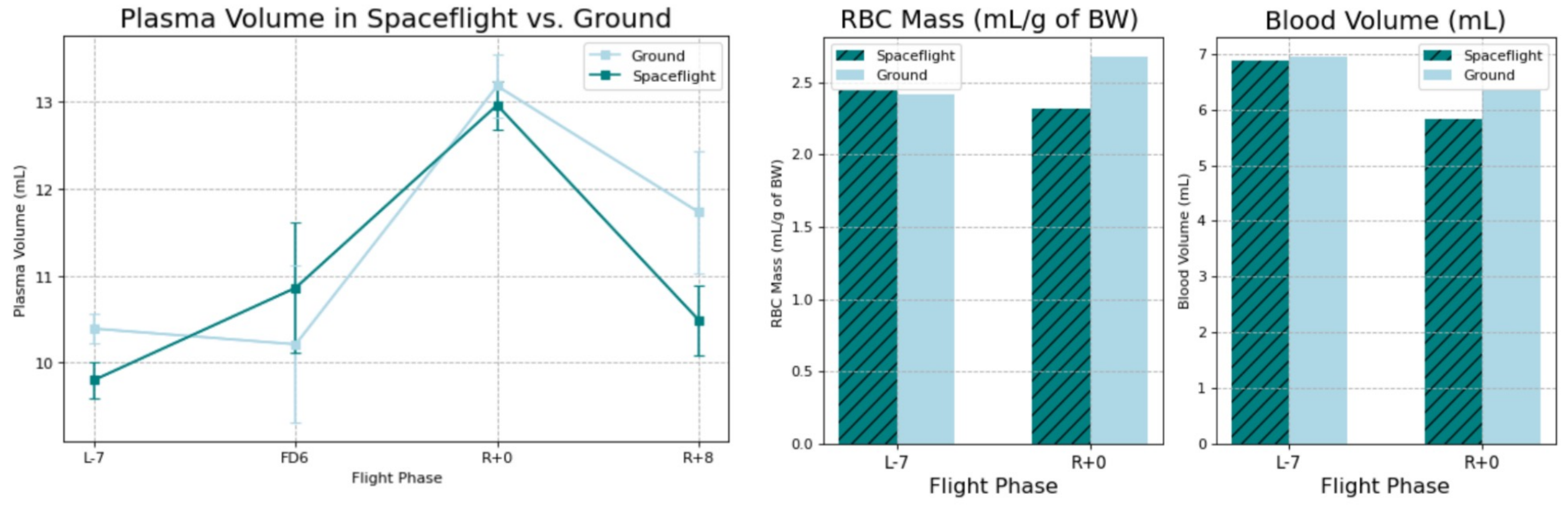


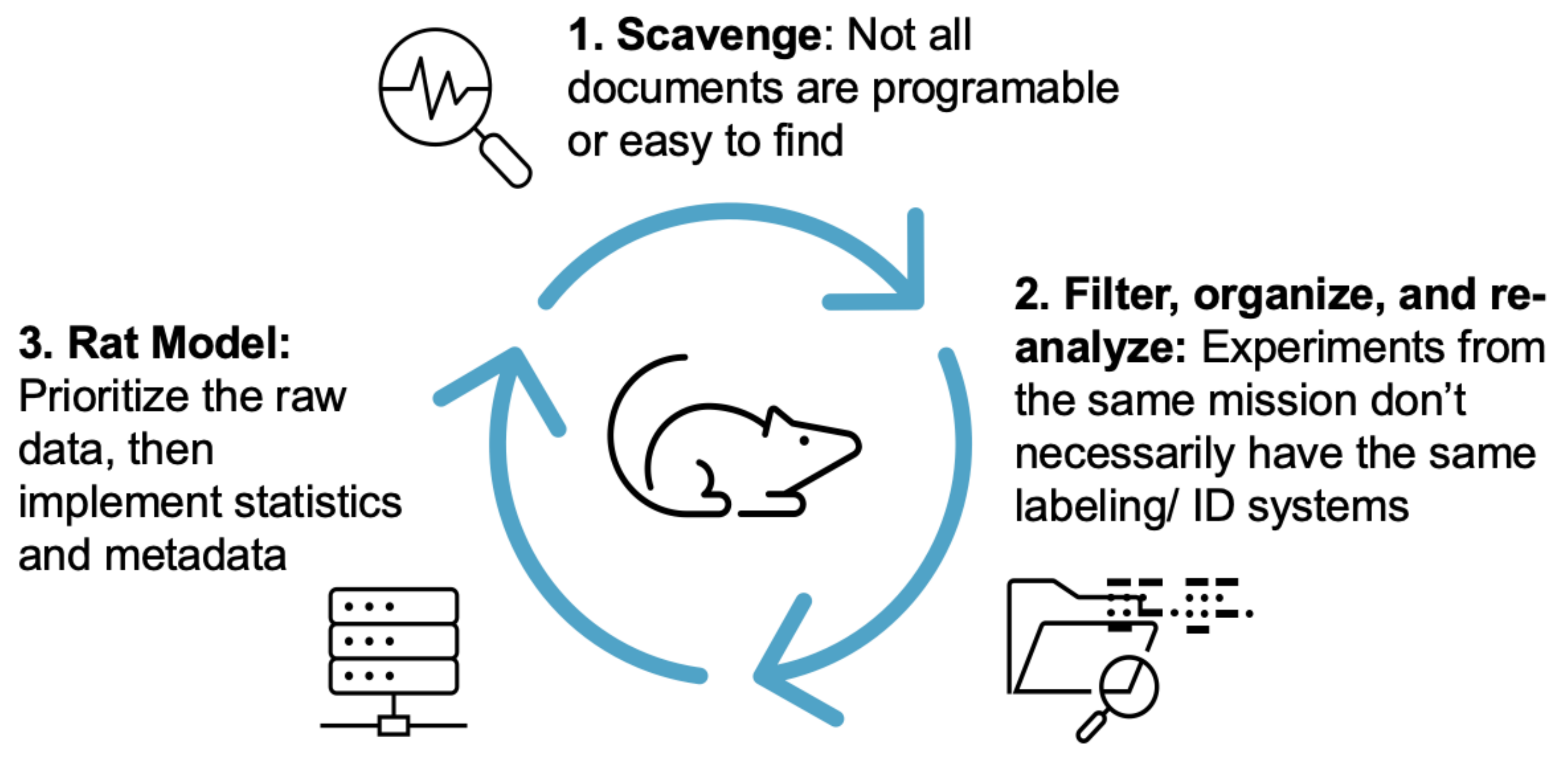
Figure 5: Plasma and blood volume in relation to RBC mass was recorded during pre-launch (L-), flight (FD), and recovery (R+). There is an increase in plasma volume in response to spaceflight. There is a slight overall decrease post-flight when compared to the ground controls.

Data Frame

	ID	measurement	unit	error	factor_value	house	date	experiment	PI	tissue	assay	file_name
0	98	0.6405	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	Thymus	Weight	OrgnWt_R.csv
1	98	0.02	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	R.Adrenal	Weight	OrgnWt_R.csv
2	98	0.0213	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	L.Adrenal	Weight	OrgnWt_R.csv
3	98	12.4863	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	Liver	Weight	OrgnWt_R.csv
4	98	1.4402	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	R.Kidney	Weight	OrgnWt_R.csv
5	98	1.4149	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	L.Kidney	Weight	OrgnWt_R.csv
6	98	1.3446	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	Heart	Weight	OrgnWt_R.csv
7	98	1.0265	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	Spleen	Weight	OrgnWt_R.csv
8	98	373.35	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	n/a	Body Weight	OrgnWt_R.csv
9	102	0.6565	g	n/a	Control	DFPT	R+9	ARC Project Data	ARC	Thymus	Weight	OrgnWt_R.csv

Table 1: Sample Data Frame organizing data and meta data in a 13-column longform data frame. The stacking nature of the data frame allows users to specify what group of data they would like to reanalyze. Specifications include "Factor Value", "Assay" , and more.

Lessons Learned



1. Scavenge: Not all documents are programable or easy to find

2. Filter, organize, and re-analyze: Experiments from the same mission don't necessarily have the same labeling/ ID systems

3. Rat Model: Prioritize the raw data, then implement statistics and metadata

Figure 6: Figure delineating the lessons learned throughout each step of the pipeline reffered to in Figure 2.

Conclusion and Deliverables

The decision to reanalyze research from SLS-2 was made on broader methodological questions on data engineering principles and research gaps about biology in spaceflight. Our methods are intended to facilitate the data organization of older spaceflight studies to perform further studies without the impediment of spaceflight experimentation costs. We developed a pipeline that could be applied to other studies of similar nature to use its results to bridge differences between space life science experiments. It's a form of systematically updating previous reviews and giving more power and purpose to older legacy data. After testing our methods with datasets from the SLS Missions, we successfully wrote code that produces a programmable data frame that creates accurate and logical plots that may be used for spaceflight data reanalysis.

Future Direction

With 12 datasets (produced about 3,000 data points) processed to the SLS-2 data frame, there are still over 200 more datasets to add to the data frame. It takes ~7 minutes to add to a dataset to the data frame because of some of the user-inputted aspects of the code. It will take about 46 hours and 20 minutes to compute all of the data sets from SLS-2. The project is an integrated, modeling and database system that will support space biomedical research and operations in a variety of fundamental ways (Hart, 2016). Such data reanalysis support system will enable the construction, validation and utilization of important responses of the whole human body to the types of stresses experienced during space flight and low-gravity environments. The discoveries produced with the reanalysis of these space life data sets will be essential for future research that addresses gaps on astronaut health and capabilities for its application on future missions for long duration spaceflight. With our data pipeline and database, we collected and integrated past and current rat data across many physiological disciplines into an operationally useful form that will not only summarize knowledge in a convenient and novel way but also reveal gaps that must be filled via new research to effectively ameliorate biomedical risks. In conclusion, we created a new data source and data pipeline that other scientists could use to perform reanalysis on datasets of similar nature to identify unseen nuances and patterns to make new discoveries. Since we want to make this project an accessible source and start an open data repository, our program and code was uploaded to Mendeley so this source can serve as an additional tool for other researchers in the field.

Acknowledgements

The development of this project would not be possible without the help of Crystal Rain Fowler,. Joshua Alwood, and the NASA GeneLab and ALSDA teams. We also include out thanks to the SLSTP staff and management for their support and guidance.

POC: (Eva Soler, undergraduate, ens2136@columbia.edu)

NASA Space Biosciences SLSTP: https://www.nasa.gov/ames/research/space-life-sciences-training-program

NASA Space Biology Program: https://science.nasa.gov/biological-physical/programs/space-biology