**Using Zebrafish as a Model System for Studying Motor Axon Guidance & Motoneuron Disease: A Teaching Case for Data Management**

By Julie Goldman

**Summary of Teaching Points**

**Module 2: Types, Formats & Stages of Data**

* The challenges in conducting a multiyear research project with living specimens
* Instrument data that needs to be exported to a common or open format for analysis, storage, etc.
* Data in digital and paper formats
* Paper lab notebook inconsistencies and lack of standardization

**Module 3: Contextual Details**

* No use of a data dictionary
* No file naming conventions
* Lack of synchronization between data sources
* No standards for data documentation

**Module 4: Data Storage, Backup and Security**

* Use of personal computers
* No plan for storage of data files
* No security and backup plan for digital and hardcopy data (lab notebooks)

**Module 5: Legal and Ethical Issues**

* IACUC-related documentation and compliance
* Need to clarify funding purposes (NIH vs. private)

**Module 6: Data Sharing and Re-Use**

* Research team uses web-based (cloud) applications to share images and data

**Module 7: Plan for Archiving and Preservation of Data**

* Need for preservation-friendly image and document file formats and media
* Use of model organism repository

**Case Narrative**

A graduate student is working with zebrafish as a model to research and develop drugs for motoneuron diseases. Motoneuron diseases are progressive neurological disorders that destroy motor neurons, the cells that control essential voluntary muscle activity such as speaking, walking, breathing, and swallowing. She is a part of a lab team that focuses on two diseases, Spinal Muscular Atrophy (SMA) and Amyotrophic Lateral Sclerosis (ALS), and specifically on the mechanisms of motor axon growth. The primary investigator (PI), a post-doctoral researcher (postdoc), a lab technician, and she, a graduate student, comprise the lab team. The lab has been working on this research since 2005 and it has received multiple NIH grants. The lab also receives research support funding from private companies, organizations, and families affected by these diseases.

There is research supporting that low levels of Survival of Motor Neuron (SMN) protein cause muscle atrophy and weakness. The graduate student is investigating how low levels of SMN protein compromise motor neurons. By using a genetic model of SMA in zebrafish, the graduate student can collect fluorescent images of axon outgrowth on a microscope and use a scoring system (severe, moderate, mild, or no defect) to count visible axons, quantifying how SMN levels affect axon development.

Degeneration of motor neurons is also the cause of muscle weakness and atrophy in ALS disease. By using a genetic model of ALS in zebrafish, the student can study *superoxide dismutase 1, soluble* (SOD1), which is a gene responsible for an enzyme that protects the body from free radicals. If free radicals accumulate in the body, they can damage DNA and the proteins produced within cells. The resulting axonopathy, the disruption of normal function in axons, causes defects in the neuromuscular system. The student observes the effects of the mutant SOD1 gene in zebrafish using fluorescence microscopy. The resulting images of juvenile zebrafish with the mutation reveal curved spinal cords, which inhibit the fishes’ motility.

The majority of the lab team’s research involves “general lab work”: DNA manipulation, polymerase chain reaction (PCR), running agarose gel electrophoresis, western blots, and microscope imaging. Since the lab is so small, it only produces around 1-2 files each per day or per experiment. Their Thermo Scientific™ NanoDrop™ spectrophotometer, which measures light transmittance or reflection intensity as a function of the light source wavelength, and their Bio-Rad (RT)-qPCR machine, which amplifies a single or a few copies of a piece of DNA across several orders of magnitude, produce Microsoft™ Excel™ files. The western blot images are scanned and any agarose gels are read on a gel box and then images are scanned and printed for densitometry quantification, the measure of light absorption through the medium. The labs’ microscopes produce Tagged Image File Format (TIFF) files and JPEG (.jpeg) files. The team uses Excel™ and SPSS™ for data analysis. The team edits image files in ImageJ™ (a Java™-based image processing program developed by the NIH) or Adobe™ Photoshop™. The team uses Portable Document Format (PDF) (.pdf) and PowerPoint™ (.pptx) files for figures and publications.

Typical data flow involves analysis of direct microscope images of manipulated fish samples, or gel images of DNA/protein analysis. These image files are produced on computers attached to the lab equipment, and then these files are analyzed on a computer and/or sent to the experimenter’s personal computer. The student always prints out a hard copy of any images and pastes them in her paper lab notebook.

The zebrafish are housed in a facility in the same lab building. There are 10-12 people working with the fish for different research projects. Everyone has their own fish, and each person is responsible for the upkeep of his or her own gene lines, as well as the cleanliness of their tanks. There is a common stock of fish for breeding and controls. The fish manager is in charge of breeding and general facility organization, and Institutional Animal Care and Use Committee (IACUC) compliance. Each person has his or her name, fish name, date of fish birth, and stock number written on his or her fish tanks. There is a database on a shared Google Drive™ that each member can sign into to make notes and update his or her fish information and individual experiments. This is a big change from the paper notebook log previously used.

The graduate student’s lab does not have any standardized way to document its data or any naming conventions for saving and locating documents and images. It also does not have a data dictionary. Files usually just involve the person’s name and a title or description meaningful to him or her. Paper lab notebooks also have a personal feel to them. The graduate student feels her notebook is comprehensible and easy to follow. The team has always used paper notebooks; she has not heard any news of switching to an electronic version. The team also uses their personal computers for lab work; she feels each person is responsible for his or her own data management. She backs up the lab files on her computer to an external hard drive. When the graduate student began working in the lab, she was given CDs with data from previous graduate students. However, now the lab team shares its data with each other using Google Drive™ and Dropbox™.

Generally, among her university peers and colleagues, the graduate student feels she can get any data she needs pre-publication; however, publications are considered the primary “electronic” form of data conservation in her lab. The graduate student feels that once the research is published, then anyone who wants it has access to the relevant data in the article. Her lab would share reagents, enzymes, and animals that were used in any of its published studies. They also share and use data via repositories that house data on genetically modified zebrafish lines, such as ZFIN, the NIH-funded zebrafish model organism database.

**Discussion Questions**

1. Types of data

a. What types of data are being collected for this study?

b. What would be needed in a data management plan to describe use of live specimens?

c. What analytical methods and mechanisms will be applied to the researcher’s data?

d. What type of outcome data will be generated?

2. Contextual details

a. What impact would the lack of naming conventions have on later data access?

b. What other contextual details would the researcher specifically need to document to make her data meaningful to others?

c. What kind of metadata would she need to upload her data to the ZFIN repository?

3. Data Storage, Backup, Security

a. Where and on what media will the data from each data source be stored?

b. How, how often and where will the data from each source be backed up?

c. How long following the completion of her study will she need to store the data?

d. What impact would the proprietary software have on storage and access?

4. Data protection/privacy

a. Who will own any copyright or intellectual property rights to the data?

5. Policies for reuse of data

a. Will the data be restricted to be re-used only for certain purposes or by specific researchers?

b. Are there any reasons not to share or re-use data?

c. How will having public and private funding affect data sharing and reuse?

6. Policies for access and sharing

a. Could there be a use for the graduate student’s data not in the published article? Explain your answer.

b. What process should be followed to gain future access to the researcher’s study data?

c. How could the team’s use of lab notebooks be improved?

d. Describe the issues related to using cloud or web-based data sharing services.

e. How does the graduate student plan to link her datasets to her published article?

7. Archiving and preservation

a. What is the long-term strategy for maintaining, curating and archiving the data?

b. What data will be included in an archive?

c. Where and how will it be archived?

d. What other contextual data or other related data will be included in the archive?

e. How long will the lab need to keep data beyond the life of the project?

Creative Commons LicenseThis work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States License](http://creativecommons.org/licenses/by-nc-sa/3.0/us/). You are free to re-use part or all of this work elsewhere, with or without modification. In order to comply with the attribution requirements of the Creative Commons license (CC-BY), we request that you cite:

* **Editor**: Lamar Soutter Library, University of Massachusetts Medical School
* **Title**: New England Collaborative Data Management Curriculum
* **URL**: <http://library.umassmed.edu/necdmc>

Revised June 12, 2015