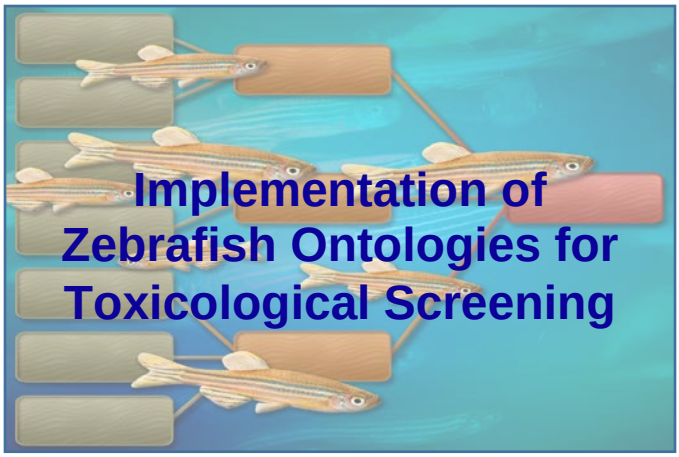

Zebrafish researchers and ontologists met to discuss the current state of ontology usage in zebrafish toxicological screening, barriers to large-scale use of ontology, and ways to encourage greater researcher buy-in.

The “Implementation of Zebrafish Ontologies for Toxicological Screening” Information Session was held on April 4 and 5, 2017, in Research Triangle Park, North Carolina. The workshop brought together approximately 30 zebrafish researchers and data scientists representing primarily government and academia. Participants discussed the current state of ontology usage in zebrafish toxicological screening, barriers to large-scale use of ontology, and ways to encourage greater researcher buy-in of ontology use. An ontology is a systematic means of naming in which terms are defined, relationships between them are established, and terms are organized in a hierarchy. Given the variability in data output from zebrafish screening assays, greater untilization of ontologies and data structure should facilitate analysis of data collected from multiple laboratories. The information session began with invited presentations by zebrafish researchers, who each described their test method protocols, measured endpoints , and data analyses. These presentations highlighted variability in protocols among researchers and the lack of consistent data annotation. Data scientists presented use cases to illustrate how ontologies can improve comparison of data between differing protocols and create connections to additional data. Participants agreed that greater collaboration between zebrafish researchers and ontologists should be encouraged and made several recommendations that would facilitate improvements in zebrafish screening data analysis.

Session 1: Overview of Zebrafish Screening Data and Practices

David Reif, Ph.D., North Carolina State University, began the session with a presentation of his recent work comparing zebrafish toxicology datasets generated in three different laboratories. Four major elements affect data analyses and harmonization across the data sets. The first element is the exposure protocol in which the experiment was conducted. For example, protocols used exhibit variations in chemical exposure regimen (i.e., static vs. static-renewal), treatment concentration, and spacing of concentration. The second element is phenoptypic: different laboratories collect different phenotypic endpoints to address different scientific questions. The third element is resolution in the data, which is directly influenced by study design. Responses observed in embryos may be pooled within each well, or they may be presented as individual data. Differences in time points evaluated and the spacing between time points also impact the level of resolution that can be achieved. The last element, reproducibility, is affected by a number of factors, including chemical delivery, automation, throughput, and historical data.

Given the near-infinite chemical space and assay design permutations required to address specific research questions, ontologies provide an opportunity to increase data utility by helping to define relationships within the data. This includes defining analogous data endpoints that might be referred to with varying terminology and helping to define any hierarchy that exists within the data. Bayesian methods can statistically optimize the weighting of relevant endpoints and allow for the comparison of pooled versus individual zebrafish data. Using such approaches, the four key elements affecting harmonization can be addressed through a combination of informatics and data sharing.

Melissa Haendel, Ph.D., Oregon Health and Science University, provided an overview of ontologies and how they can be employed to help structure data. Existing biomedical ontologies lack uniformity, making it difficult to perform analyses across species and endpoints without the development and use of additional "bridging" ontologies.

Other participating investigators shared the zebrafish protocols employed in their laboratories:

- Arantza Muriana, Biobide – BBD Biophenix S.L.
- Stephane Padilla, Ph.D., U.S. Environmental Protection Agency
- Stefan Scholz, Ph.D., Helmholtz Centre for Environmental Research
- Don Stedman, Ph.D., Pfizer, Inc
- Robert Tanguay, Ph.D., Oregon State University
- Mitch Willbanks, U.S. Army Corps of Engineers

These presentations exemplified the challenges described by Reif including use of different environments, phenotypes, resolution, resulting in decreased reproducibility. Attempting to mandate rigid, standarized protocols across laboratories would be impractical, but the development of toxicological and experimental ontologies for zebrafish would aid in comparison of datasets.

Workshop participants agreed in principle that a minimum set of endpoints should be defined that would be collected and reported in all future zebrafish studies.

Session 2: State of the Science of Behavioral Testing in the Zebrafish Larvae

Randall Peterson, Ph.D., Health University of Utah, summarized a system developed in his laboratory to describe toxicant-induced changes in zebrafish behavior. Data collected are reduced to a string of numeric features that can be used to rapidly compare data across and within large datasets. This approach has been used to identify small molecules with antipsychotic-like behavioral profiles[1]. Peterson emphasized that the availability of publicly available ontologies would facilitate ongoing research.

[1] Bruni et al. 2016. Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. Nat. Chem. Biol. 12:559–566; doi:10.1038/nchembio.2097.

Session 3: Small Group Use Case Analyses and Brainstorming Exercise

This discussion session was led by Lyle Burgoon, Ph.D., U.S. Army Engineer Research and Development Center, and Yvonne Bradford, Ph.D., The Zebrafish Information Network. Participants were charged with identifying what experimental elements were critical for queries they might wish to run on a data set. Elements and issues discussed in the session include:

- The identification and definition of phenotypes, factors affecting the observation and severity of phenotypes, and whether the definition and severity of a reported phenotype is comparable across laboratories
- Data analysis within and across laboratories, including the utility of binary and single-point data, requirements for concentration-response data, and the value of point-of-departure estimates or other toxicity estimates
- Factors affecting data quality, such as the requirement for positive, negative, and solvent controls and inclusion of historical control data

After identifying some of the elements that affect the interpretation of data queries, researchers were asked to identify the information from zebrafish data sets that they would potentially find most useful. Common responses included phenotypes measure and altered, chemical coverage, laboratory- and species-specific effects, extrapolation of zebrafish embryo data to larvae and adult fish, extrapolation of zebrafish embryo data to mammals, cross-laboratory or cross-platform comparison of model data and data sets, and the use of zebrafish embryo data to prioritize chemicals for further toxicological evaluation.

Data Requirements

While it would be ideal for all data generated during the course of a study to be reported, providing this level of experimental metadata requires significant time and effort for data entry. Therefore, a minimum set of metadata was recommended that included descriptions of adult fish husbandry, chemical handling and chain of custody, exposure conditions for embryos and larvae, and data collection and analysis methods. The group believed that a publicly accessable internet “webform” for data entry could be developed to facilitate this process. Such a form should support both easy hand-entry of data and uploading of datasets in text file format. It was noted that the webform should be tested using existing data sets prior to deployment.

Data Usage

Metadata provides information about other data, and in the case of zebrafish assays often provides key information about an experiment was conducted. A balance between the time and cost to enter metadata and the added value of that metadata to future analyses should be considered. The group agreed that the availability of the metadata was more important to them than the specific data reported.

While there have been attempts at mapping experimental data to ontologies, they have yet to be widely implemented. Each ontology has its own vocabulary, and for this to be effectively utilized, documents and data must be uploaded to a database utilizing the ontology structure. Likewise, use of ambiguous abbreviations needs to be avoided . Such issues could potentially be resolved by the ontology creators implementing a standardized format that includes a "data dictionary" and providing specific instructions for data cleanup.

Summary data found in the peer-reviewed literature often fails to provide sufficient information on how summary scores were derived and whether endpoints were weighted in the analysis. Participants briefly discussed whether the development of an open-source, annotated zebrafish data repository could be developed and used to store in depth, annotated data for further analysis.

Informing Biological Issues using Zebrafish Data, and Community Development and Planning

There is a widely recognized lack of available annotated data for metanalyses. There also are potential difficulties involved in getting researchers to share data, finding appropriate ways to store data, and how best to encourage collaboration between the zebrafish community, ontologists, and the research community at large.

Future community development activities were identified that would encourage the reporting of protocols and endpoints, as well as the associated metadata, to allow for more efficient use of data. The development of a web form and data storage to capture this information was envisioned as a collaborative effort between the National Toxicology Program and the broader zebrafish community. The database should be created by organizations who have some degree of control over the database structure, and should contain an analytical package that could be used behind a researcher's firewall. Potential incentives for researchers to share their data would include the free availability of data analysis tools, provisional sharing of data, and owner control of data access. The database should establish standards for data sharing and encourage data contributions by the larger scientific community.

Major Recommendations

There is a need for:

- A comprehensive group analysis of a compiled data set should be conducted to demonstrate the advantages of utilizing ontologies to harmonize zebrafish data so that they can be easily compared.
- An online database for collection of zebrafish data. Features that such a database should incorporate include a web form for easy submission of data and output of revised data strutures to permit interlaboratory and ontology mapping.
- Incentives to contribute to the database and use collaborative workspaces should be identified and communicated to researchers using zebrafish as a model species.