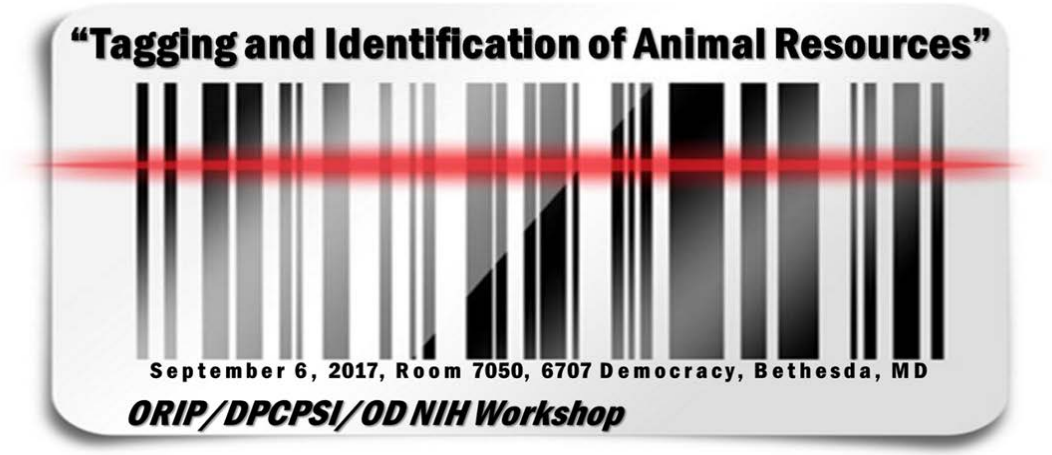

Tagging and Identification of Animal Resources Workshop

Division of Comparative Medicine
Office of Research Infrastructure Programs
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health

Wednesday, September 6, 2017
Bethesda, MD

Workshop Report

Table of Contents

A. Overview

On September 6, 2017, the Office of Research Infrastructure Programs (ORIP) at National Institutes of Health (NIH) convened a workshop on the "Tagging and Identification of Animal Resources." The goals of the workshop were to provide participants with (1) a thorough review of the current challenges and possible solutions for the reporting and unique identification of research resources, including model organisms used in biomedical research; (2) an overview of the Resource Identification Initiative (RII or Initiative) to create and maintain biological resource identifiers for NIH-supported animal repositories; and (3) a hands-on session for Resource Directors and NIH staff on acquisition and use of unique Research Resource Identifiers (RRIDs) to identify research resources and conduct citation analysis of their usage. Participants included representatives from academia, industry, and publishing.

Carrie Wolinetz, Ph.D., Acting Chief of Staff, Associate Director for Science Policy, NIH, gave a keynote presentation on NIH policies and concepts for rigor and reproducibility in biomedical research. The remainder of the 1-day workshop consisted of topical sessions related to the use of RRIDs. Each session included three to four individual presentations, followed by discussion that allowed for audience participation. Topics of the sessions were as follow:

1. Identification of Animal Resources in Scientific Literature and Grant Reports
2. Animal Repositories and Their Role in Supporting High-Standard Research
3. Common Publication Guidelines for Citing Animal Resources
4. Future Development of Animal Resource Identifiers
5. Hands-on Session with RRIDs.

B. Introduction and Welcome

Oleg Mirochnitchenko, Ph.D., Division of Comparative Medicine (DCM), Office of Research Infrastructure Programs (ORIP), welcomed participants and remarked on the importance of tagging and identifying animal resources, which aligns with new NIH initiatives and recommendations for biomedical research. He stated that the workshop will consist of presentations from stakeholders at all levels, including journals, staff from NIH Institutes and Centers (ICs), and NIH policy developers. He reflected on the 2013 DCM-sponsored symposium titled "Animals Models and Personalized Medicine," which yielded recommendations to develop the Initiative. Dr. Mirochnitchenko indicated that the NIH is looking forward to engaging discussions and shared experiences that will inform future improvements to the Initiative.

Stephanie Murphy, V.M.D., Ph.D., Director, DCM, ORIP, welcomed attendees to the meeting and expressed appreciation to the organizers for their efforts in planning the workshop. The DCM supports the development of new animal models, biological materials, informatics systems, and shared resources with the goal of distributing, characterizing, maintaining, and archiving these diverse resources for use by other biomedical researchers. One primary obstacle to using these resources in a more rigorous and reproducible manner is the absence of standards and sufficient unique identifiers that would allow easier identification in the current scientific literature, as well as authentication for using these resources for grant proposals. Guidelines on how to correctly report, accurately replicate, and extend findings from animal research resources for the development and maintenance of a

citation system are needed. This identification system will provide resource distributors and funding agencies a means of monitoring use of these critical model systems and biomaterials, while promoting rigorous and transparent research in all areas of science. In closing, Dr. Murphy conveyed to workshop participants the NIH's appreciation for their willingness to engage in these discussions on topics identified by the organizers.

C. Summary of Presentations and Discussions

Rigor and Reproducibility in Biomedical Research: Current NIH Guidelines

Dr. Wolinetz discussed current NIH guidelines for rigor and reproducibility in biomedical research. She stated that rigor and reproducibility remain the undercurrent of many of the policy and scientific initiatives at the NIH, including those related to clinical trials, data sharing, and stewardship. The NIH announced plans to enhance scientific reproducibility in biomedical research, which were detailed in the 2014 publication by NIH Director Dr. Francis S. Collins and Principal Deputy Director Dr. Lawrence A. Tabak. A multi-tiered approach focused on four areas was proposed: scientific premise, rigorous experimental design, consideration of sex as a biological variable, and authentication of key resources, which is the focus of today's workshop.

A major component to establishing the NIH policy is transparency in reporting methods, statistical approaches, and generalizability of research. Guiding principles for rigor and reproducibility include clarifying NIH's long-standing expectations, raising awareness and beginning culture shifts in the scientific community, prompting grant applicants to consider relevant issues, improving applicants' descriptions of research, demonstrating to our public stakeholders that the NIH is seriously considering their concerns, and ensuring that the NIH is investing in the best science and minimizing unnecessary burden. Dr. Wolinetz stated that the NIH Enhancing Reproducibility through Rigor and Transparency Policy timeline began in 2012 and has included pilot interventions for enhancement in 2013, a series of guide notices published in 2015, and updated instructions for grant applications, Research Performance Progress Reports (RPPRs), and the associated grant review language in 2016. The 21st Century Cures Act, which Congress passed in December 2016, also included language on rigor and reproducibility. In 2017, the NIH published a guide notice clarifying the authentication requirements based on input from the scientific community. Dr. Wolinetz remarked that the areas of concentration outlined in the 2014 Collins and Tabak publication have been incorporated as rigor elements into NIH research applications—authentication of key biological and/or chemical resources will be a new attachment separate from the research strategy.

Key biological or chemical resources are integral to any proposed research. Studies have reported on the misidentification of cancer cell lines, some of which had been used in NIH-funded research, and have served as the basis for implementing policy changes for authentication of key biological and/or chemical resources. From a policy perspective, the quality of key resources is critical to reproducibility, so they should be regularly authenticated; the resources may or may not be generated from NIH funds. Differences that could vary from laboratory to laboratory may occur over time and may possess qualities that could influence the research data. The NIH recognized that polices regarding authentication were not clear to the scientific community. For example, 44 percent of inquiries made to the Office of Extramural Research on rigor and reproducibility are related to authentication. Of that 44 percent, 31 percent indicate a lack of clarity in the policy, 30 percent ask about specific reagents, 24 percent are unsure about attachment instructions, and 17.5 percent request examples. To address this issue, the NIH clarified and provided examples of the authentication requirements for grant applications in the May 2017 notice, NOT-OD-17-068. The emphasis is on the authentication plan, which may include existing published consensus standards, not data reporting.

Dr. Wolinetz informed participants that the 21st Century Cures Act *Subtitle C, Section 2039*, "Enhancing the Rigor and Reproducibility of Scientific Research," requires the NIH Director to convene a working group under the Advisory Committee to the Director (ACD) to develop and issue recommendations for a formal policy—this will build on prior NIH efforts. The working group met on May 25, 2017, and provided a preliminary report to the ACD in June 2017 that included recommendations on the NIH grant application to develop checklists for applicants and reviewers and resources to maintain or support a database of validated and vetted materials. A full report is expected to be completed in early 2018, and NIH actions in response to the recommendations are likely to occur before June 2018; a report to Congress will be delivered in December 2018. Dr. Wolinetz emphasized that enhancing reproducibility is not just an initiative within the NIH, but is a team effort that requires multi-stakeholder engagement and ongoing input from the scientific community to develop genre-specific standards. In parallel, journal publishers have united to enhance publication practices for reproducibility. The overarching objective of the NIH is to be a leader in ensuring rigor and reproducibility in biomedical research that strengthens the scientific community without imposing undue burden.

Discussion

Terry Magnuson, Ph.D., The University of North Carolina at Chapel Hill, sought clarity on NIH's definition of reproducibility in the context of genetic drift of biological resources. Dr. Wolinetz clarified that the expectation is to develop a rigorous plan that would automatically drive the reproducibility. Transparency is a key factor to understanding discrepancies in data.

Natalie de Souza, Ph.D., *Nature Methods*, asked how the NIH would assess the impact of guidelines on the review process. Dr. Wolinetz noted NIH's ongoing efforts and commitments to assessing rigor and reproducibility, in which the Center for Scientific Review plays a key role.

J. R. Haywood, Ph.D., Federation of American Societies for Experimental Biology (FASEB), commented that addressing rigor and reproducibility should extend beyond NIH requirements to become standard practice in the scientific community. Dr. Wolinetz noted the challenge of interjecting change to a dynamic and large scientific community. Policies and guidelines are important tools to promote a cultural change, but training also will be critical to this deliberative process. Dr. Mirochnitchenko added that the NIH will rely on experts in the scientific community to assist with implementation of rigor and reproducibility tools. Successful outcomes will be reflected in the practices being used.

Session 1: Identification of Animal Resources in Scientific Literature and Grant Reports

Anita Bandrowski, Ph.D., SciCrunch Inc., discussed RRIDs, which are unique tags for key biological resources that are involved in all phases of the experimental life cycle. The RRID string is cited in the methods section of a publication and consists of a search-friendly identifier, a resource repository or authority identifier, and a local identifier (e.g., name and stock/catalog number). The SciCrunch RRID portal—which includes 2.5 million antibodies, 500,000 organisms, 80,000 cell lines, and 14,000 software projects—supports the NIH RII. She detailed the author workflow process: the journal directs the author to the RRID portal, the author searches for a resource, cites the found resource in the manuscript, and the paper is published. Data on a broad range of model organisms are aggregated within the RRID portal and are represented as subcenters, including *Caenorhabditis elegans* (*C. elegans*), *Drosophila*, and zebrafish. A master stock list of current and previous stocks from 25 participating stock centers is being maintained. RRIDs for model organisms are based on stock numbers that are unique to the stock center, providing a globally unique identifier; therefore, stock numbers must not be reused.

Dr. Bandrowski described SciBot, a curation tool used to find RRIDs in articles, look them up in the SciCrunch resolver, create Hypothesis.is annotations that anchor to the RRIDs, and display results.

Hypothesis.is has the capacity to capture user comments in the publication PDF files that will populate journal publisher websites and PubMed Central in annotations that comply with World Wide Web Consortium standards. Publishers are beginning to use robot (bot) software applications that run automated tasks, because they can significantly reduce the base curation time and allow human curators to focus on other tasks. Bots are less error prone, provide syntax that is always correct, and reduce the lag time between release of the HTML version of the paper and the base RRID.

Dr. Bandrowski summarized that RRIDs are uniform across publishers and resolvable (i.e., persistent unique identifiers), allowing the scientific community to know which resources are used in the publication and to identify other publications that also use the identical resource. For the researcher, RRIDs can improve tracking of the impact of resource; provide credit to tool makers, in addition to authors; and track problems with key biological resources.

Discussion

Ian Korf, Ph.D., University of California, Davis, pointed out that commercial software updates maintain the base identifier number with minor changes to accommodate the newer version and asked whether SciCrunch could do similarly for RRIDs. Dr. Bandrowski explained that the digital object identifier system differs from the RRID, which allows the user to track across the publication landscape. The syntax behind the RRID contains information similar to a software version update

Valentina Di Francesco, Ph.D., National Human Genome Research Institute (NHGRI), provided an update of the model organism databases (MODs) and the NIH Data Commons Pilot Phase. The Alliance of Genome Resources (AGR), a consortium of MODs, was established in 2016 to provide an integrated resource that would facilitate data access and use. Founding members include the NHGRI-funded FlyBase, Gene Ontology Consortium, Mouse Genome Database, Saccharomyces Genome Database, WormBase, and Zebrafish Information Network (ZFIN), as well as the National Heart, Lung, and Blood Institute–funded Rat Genome Database. Discussions are in progress to expand to other databases, such as the *Eunice Kennedy Shriver* National Institute for Childhood Health and Human Development-funded XenBase and Echinobase. The AGR aims to federate participating resources through a unified Web portal; standardize the acquisition, curation, and display of shared data types; design a unified data model and a modern scalable information architecture; and support the scientific community to fully leverage AGR resources by establishing a common Scientific Advisory Board (SAB) and a central office for communications, user support, training, and outreach. Dr. Di Francesco pointed out the implementation challenges that would need to be addressed and noted the two priority areas for the AGR: (1) data types, tools, interfaces, and outreach; and (2) infrastructure. Accomplishments to date include establishing a governance and communication infrastructure; an operational framework; an SAB; and 10 working groups to focus on priorities. Soft launches with limited front-end functionalities were released in March and June 2017, and product development is in progress with an expected official release in October 2017. Also in March 2017, the AGR convened a meeting with the SAB and representatives of the model organisms research community.

Dr. Di Francesco described the NIH Data Commons Pilot Phase. In the NIH Data Commons, such products of research as data, methods, or tools are treated as digital objects and shared in a virtual space. The digital objects must comply with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Initial data sets for the Data Commons Pilot Phase will include Trans-omics for Precision Medicine (TOPMed), Genotype-Tissue Expression (GTEx), and MODs/AGR. Applications were solicited through the Other Transaction Research Opportunity Announcement, have been reviewed, and are in the negotiation phase with potential awardees. A Data Commons Pilot Phase Consortium (DCPPC) kick-off meeting is scheduled to occur within 1 to 2 months following award selections to develop plans for activities for stage 1 of the project (i.e., 180 days); stage 2 will involve renegotiation of the terms of

the awards and an implementation phase. The governance structure of the Data Commons will consist of an operational team, a Commons working group along with NIH leadership, and an External Advisory Board. As a member of the DCPPC, the AGR is afforded the opportunity to participate in development of use cases and assist in adoption of NIH-wide FAIR solutions. Dr. Di Francesco noted that the AGR and the NIH Data Commons are complex projects that are in early stages of development and the potential exists for priorities to conflict.

Kristin Abraham, Ph.D., National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) presented on the NIDDK Information Network (dkNET) and its role in expanding use of RRIDs across the NIDDK. The dkNET connects researchers to research data, tools, materials, and services and spans across multiple projects, databases, and NIH ICs using the SciCrunch infrastructure. Aligning with NIH policy requirements for authentication, dkNET provides an opportunity to introduce FAIR principles and RRID concepts to NIDDK investigators, assist in implementing NIH rigor and reproducibility initiatives, and provide RRID resource services. Resource authentication modules have been developed within dkNET that can assist researchers in preparing new attachments of authentication of key biological or chemical sources by providing manual and automatic reports for authenticating a resource. Authentication alerts can be enabled to inform users of updates to resources they currently are using. Current efforts include expanding RRIDs to include NIDDK infrastructures (e.g., Centers or Cores) and NIDDK-unique resources not captured in standard sources.

Discussion

Kent Lloyd, D.V.M., Ph.D., University of California, Davis, wondered whether efforts to expand RRIDs to include NIDDK infrastructure would be applicable to other ICs. Dr. Abraham explained that infrastructure RRIDs are being developed initially for small-scaled centers and then will be shared with associated investigators, who will provide feedback on the uptake and impact.

Laurel Haak, Ph.D., Open Researcher and Contributor ID, (ORCID), Inc., called attention to efforts in the U.S. Department of Energy to develop RRIDs for centers.

Session 2: Animal Repositories and Their Role in Supporting High-Standard Research

Monte Westerfield, Ph.D., University of Oregon, described two zebrafish research resources, ZFIN and the Zebrafish International Resource Center (ZIRC), and unique zebrafish identifiers. The ZIRC provides resources, distributes information, and participates in research to improve zebrafish health. In 2016, 2,493 cell lines were imported into ZIRC; 103,720 animals were shipped to 515 laboratories; 188 antibodies were shipped to 167 laboratories; and 147 health diagnostic cases were provided for 101 laboratories. ZFIN contains information on 36,394 zebrafish genes, including more than 100,000 expression and phenotype annotations. Also represented are 20,842 zebrafish genotypes and 21,695 curated publications. ZFIN maintains the reference genome sequence, which was provided by the Wellcome Trust Sanger Institute. Every item provided by ZIRC has a unique and persistent ZFIN identifier. To date, ZFIN has approximately 5 million unique identifiers that are permanent and never deleted, even when the object becomes obsolete. The challenge lies in getting authors to use these identifiers in publications. The RRID attaches a prefix to the ZFIN identifier and can be generated automatically. For example, ZFIN:ZDB-GENO-070524-7 becomes RRID:ZFIN:ZDB-GENO-070524-7.

Dr. Westerfield discussed some of the challenges to using RRIDs. The RRID prefix redirects to SciCrunch, rather than to the primary data source, causing a synchronization problem with ZIRC and ZFIN curation. This redirect also could lead to incomplete or misleading information. He summarized that the research communities with MODs strongly support identifiers, but other communities and data types

that lack MOD resources need identifiers like RRIDs. The use of RRIDs with MOD data has several issues that need to be addressed.

Ian Korf, Ph.D., University of California, Davis, reported on the Mutant Mouse Resource and Research Center (MMRRC) repository, whose mission is to distribute and cryopreserve scientifically valuable, genetically engineered mouse strains and mouse embryonic stem cell lines with potential value to the genetics and biomedical research communities. He emphasized that understanding, documenting, and accurately reporting the genetic backgrounds of mouse models used in research are essential to achieving reproducible results. The MMRRC was established in 1998 in response to recommendations from an NIH-sponsored meeting on priority setting for mouse genomics and genetic resources and was initiated in 1999 by the NIH. In 2001, the website launched and strain acquisitions began, which was followed by strain distributions in 2002, and the addition of embryonic stem (ES) cell lines in 2003. Currently, the repository houses more than 30,000 mouse strains and cell lines. The benefits to using the MMRRC are twofold. First, the requesting investigator is provided access to unique mouse models that are not commercially available elsewhere, with assurances of specific pathogen-free health status mice and genetic quality control. Second, the donor investigator fulfills the NIH obligation to share resources, reduces the use of animal housing resources at their institution, generates a cryopreserved archive of the line/strain, and eliminates their direct shipment of mice to multiple requestors. Resources are distributed across the United States from four regional sites: The Jackson Laboratory, The University of North Carolina at Chapel Hill, University of Missouri, and University of California, Davis. An Informatics, Coordination, and Service Center coordinates activities for the MMRRC. Unique identifiers or RRIDs are included on the MMRC strain detail sheet (SDS).

Dr. Korf pointed out some challenges of using genetically engineered mouse strains regarding rigor and reproducibility, and he highlighted new strategies being used to address these challenges. Several genetic backgrounds are used in developing strains, which could affect the phenotype. Strains retrieved from the archive may experience genetic drift from the original background. The SDS collects information on the mutation, not the genetic background. In addition, the microbiome also could affect the phenotype and is difficult to track. The MMRRC is now using mouse universal genotyping arrays (MUGA) to describe the genetic backgrounds of strains to populate the SDS. Dr. Korf noted the efforts to track usage of MMRRC resources through journal article searches and contacting prior users.

Madeline A. Crosby, Ph.D., Harvard University, described FlyBase, a database of *Drosophila* genes and genomes and discussed issues related to literature curation, interactions with reagent providers, and a proposal to develop a standard reagent table. Bibliographies, data extracted from research articles (i.e., papers), and large-scale data sets are data sources for FlyBase. Stock and reagent sources are provided by Indiana University's Bloomington *Drosophila* Stock Center (BDSC) and *Drosophila* Genomics Resource Center (DGRC). Identification of genes described in a paper is a primary task of a genetic database, and the majority of papers use FlyBase symbols. General statements are reported less frequently, suggesting that the reporting of stock identifiers has improved. Stock center identifiers are being included more often, but there is inconsistency regarding location or format. Publication citation information usually is sufficient for stocks obtained from sources other than the public stock centers. Yet in some cases, source information is ambiguous, a description is repeated from a prior paper rather than citing that paper, and errors are common. Rather than guessing, FlyBase uses an unspecified or unknown record category to curate these types of data. Dr. Crosby emphasized that FlyBase associates data with the relevant genetic component, not the stock or genotype. Extensive links to stocks are provided in reports, tables and hit lists, and resource pages. Stock lists contributed by owners and compiled by BDSC are the source of all information in FlyBase, and the DGRC is the primary public resource for cell lines and molecular reagents data. Efforts are being coordinated to incorporate cell line RRIDs into FlyBase.

Dr. Crosby detailed a proposal for a standardized reagent table. The goals are to encourage use of reagent source and identifier information, facilitate information handling, and increase transparency and

reproducibility, as well as accuracy and efficiency. A standardized reagent table is expected to benefit researchers, journals, and post-publication users of the information. An updatable spreadsheet that is active during the course of a project to record reagents as they are used is the recommended standardized format. The original table was developed in consultation with other MODs and was updated to a compact version based on feedback from journals. Several journals have expressed interest in a standardized table.

Discussion

Dr. Lloyd noted the genetics expertise that may be required to understand the information in the standardized table and asked how that would affect the use. Dr. Crosby responded that FlyBase users are experienced with using identifiers, but recognized the need to develop guidelines to educate the broader research community. Dr. Lloyd commented that the standardized format would address descriptions of the data but the veracity and validity of the data remain a concern.

Dr. Bandrowski explained that SciCrunch has used bots to perform data verifications.

Session 3: Common Publications Guidelines for Citing Animal Resources

Emma Ganley, Ph.D., *PLOS Biology*, discussed the journal's experiences with mandates and policy implementation. In 2012, *PLOS ONE* launched the Reproducibility Initiative, a collaboration with Science Exchange and figshare to encourage authors to validate their work by facilitating collaboration with an unbiased expert, and began offering a Certificate of Reproducibility upon completion. The information necessary for validation is required, but the best method to obtain this information remains to be identified. The March 2014 update to the PLOS Data Policy required authors to submit a Data Availability Statement in addition to the standard manuscript data requirements. Since March 2014, more than 81,000 papers have been published with a data statement and less than 0.1 percent of submissions were rejected due to authors' unwillingness or inability to share data. Dr. Ganley stated that the *PLOS* Data Policy allows for minimum checks for compliance across publications, but not to the level of detail needed. Recognizing that mandates do not necessarily guarantee 100 percent compliance, considerations should be given to encourage data validations early on through incentives.

In January 2015, *PLOS* introduced the Research Resource Identification Initiative to two of its journals, *PLOS Biology* and *PLOS Genetics*, as a strong encouragement for authors to link to RRIDs for all relevant resources. *PLOS* authors who had used RRIDs indicated in response to a February 2016 survey that the most common reason to include identifiers was to ensure reproducibility. Others commented that RRIDs were easy to include and should be required, but they were concerned that a mandate would increase the cost of publications. Journals can provide tools, encourage, and suggest best practices to the scientific community, rather than mandating these types of changes. For example, *PLOS* partnered with Protocols.io in April 2017 to offer authors tools for sharing methodological details about their research.

Ann Goldstein, Ph.D., *Neuron***,** presented on the tagging and identification of animal resources at Cell Press. One major element of Structured, Transparent, Accessible Reporting **(**STAR) Methods being used at Cell Press is the Key Resources Table. The organisms used, source, and unique identifiers (e.g., RRIDs) are included in the table. The Experimental Model and Subject Details section of the publication provides information on animal husbandry and housing conditions, age and sex of the model organism used, and information on regulatory standards. Dr. Goldstein noted areas for improvement in tagging resources at Cell Press. The use of unique identifiers is encouraged, but has not been universally adopted. For example, 42 percent of papers published in *Neuron* and 8 percent of papers published in *Cell* from January to June 2017 that used animal models provided RRIDs. Also, the adherence to providing full genotype of the organism used varies. Cell Press welcomes suggestions of ways to improve reporting and tagging of model organisms.

Natalie De Souza, Ph.D., reported on reproducibility initiatives at the *Nature* journals. Four areas that journals and publishers can focus on to improve reproducibility in publications are education, infrastructure, policy, and ways to shift the incentives. *Nature*'s reproducibility efforts began in May 2013, and have focused primarily on raising reporting standards for methodology, data, materials, and code components of *Nature* papers. Central to the methodology and experimental design changes is the requirement to complete an 18-point reproducibility checklist for life science papers. Reporting unique materials and their availability and providing descriptions of animals used in the studies also are required. To achieve greater accessibility of reported information, *Nature* began publishing the reproducibility checklist in 2017, requested greater transparency in data representation, and began developing discipline-specific standards that will be published with the manuscript. In 2016, *Nature* began requiring data availability statements and data citations.

Dr. De Souza pointed out that materials reporting in *Nature* papers requires authors to make unique materials available to the public, report on the authentication of cell lines used, deposit resources (e.g., mutant strains and cell lines) in established public repositories, and provide accession numbers for resources. She noted that *Nature* journals currently do not publish a materials availability statement and do not encourage or require use of RRIDs. The main elements of *Nature Methods* code guidelines are the requirements to send software before the review process and a review of software by one or more referees.

General Discussion with Journal Editors

Dr. Mirochnitchenko commented on the overarching ideas regarding resource identification in papers and reproducibility, noting that funding agencies and principal investigators have a vested interest. He observed the undue hardship being placed on journals to perform resource identification checks for papers and suggested that the scientific community invest additional efforts to ensure that the RRIDs being provided to the journals are accurate.

Jonathan Pollock, Ph.D., National Institute on Drug Abuse, NIH, asked how journals reconciled costs for checking publication citations and wondered whether setting a default to reject submissions without the necessary identifiers was feasible. Journal editors explained that the standard checks will not be able to identify or rectify an incorrect RRID. Generating an all-inclusive table to capture and check all elements would be challenging.

Mr. Paul Donohoe, Somark Innovations, commented that the validity of RRIDs should be checked well before submission of manuscripts or grant proposals, and Dr. Haak suggested recording RRIDs in electronic laboratory notebooks.

Maryann Martone, Ph.D., University of California, San Diego, pointed out that reviewers are focused on the science, not on resource identifiers or other specific details. She suggested decoupling the paper workflow model to address the different level of review.

Dr. Haywood wondered about applying the identifier concept to address issues related to animal-based research. Studies to evaluate the animal husbandry effects on reproducibility are limited. Journal editors noted that information on husbandry is being captured, but it may not be reviewed in great detail.

Dr. Martone commented that research libraries could provide the level of expertise needed to perform resource validity checks prior to grant submissions. A participant added that research libraries currently are performing these types of checks and would be a resource to leverage.

Session 4: Future Development of Animal Resource Identifiers

Maryann Martone, Ph.D., discussed the past, present, and future of RRIDs. She noted the two NIH-funded projects that made the RRIDs possible: the Neuroscience Information Framework (NIF) and dkNET, which led to the development of a generic and customizable version of the NIF platform that later resulted in SciCrunch. Investigating the use of text mining to identify antibodies used in research papers and addressing the need to identify and track resources provided the rationale for RRIDs. She detailed the history of the RRIDs. Milestones included publishing a white paper in 2011 after the 2010 ORIP-sponsored meeting of the Linking Animal Models to Human Disease Initiative, where identifiers for antibodies and organisms were proposed. At the 2013 Society for Neuroscience publisher meeting, the launch date for a pilot project was established. In 2014, dkNET launched the Resource Identification Portal, and the RRID project started. The commercial version of the technology used in dkNET, SciCrunch.com, was founded in 2016, and efforts expanded beyond neuroscience to other disciplines in 2016–2017. Data show that in 2014, 115 papers from 25 different journals used RRIDs; in 2017, 4,211 papers published in 314 different journals used RRIDs, suggesting that the concept is being broadly adopted in the scientific community.

Dr. Martone remarked on the future direction for RRIDs and called attention to two new efforts: resource watch and data citation. Resource watch describes a process in which the dkNet and RRID ecosystem will allow, for the first time, a means to disseminate information on problematic research resources before, during, and after the resources enter the biomedical literature. The RII and FORCE11 (The Future of Research Communications and e-Scholarship [2011]) are actively involved in efforts to develop and broadly adopt formal systems for data and software citation. A publisher's Data Citation Roadmap linking RRIDs into the publishing workflow process has been introduced. The next steps for RRIDs and the RII will be to work with organism databases to resolve the remaining issues and integration of RRIDs with other identifiers.

Laurel Haak, Ph.D., presented on the role of persistent identifiers in research reproducibility. Persistent identifiers embedded in research workflows are enabling authoritative connections between researchers, organizations, and academic contributions. ORCID provides a persistent digital identifier that allows the researcher to connect with publishers to assert authorship, funders to assert an award, and employers to assert affiliation. Permission to collect or exchange data within the integration must be granted. Research workflows include sharing research via articles, data sets, and activities; presenting at a conference; and receiving an award. The researcher has the responsibility to register for an ORCID identifier and to use the identifier when interacting with research systems. Use of established integrations extends beyond papers. Researchers use special shared scientific facilities, equipment, and collections to develop data sets. These resources are not exposed in research outputs. Connecting with these resources has relevance for transparency, reproducibility, and ongoing funding for research projects. Efforts in the scientific community should focus on making resources citable, making it easier to cite funding, removing barriers and challenges to using RRIDs, and giving credit by including persistent identifiers in published papers and indexing them in abstracting systems.

J. R. Haywood, Ph.D., described FASEB's support of scientific standards through engagement of the scientific community. FASEB is comprised of 31 scientific societies representing more than 125,000 scientists. Efforts to support animal welfare include publishing a Statement of Principles that provided the guiding principle for animal research, working to reduce regulatory burden, and advocating for animal research as well as biomedical and biological scientists. Dr. Haywood remarked that FASEB has worked extensively to address regulatory burden, which involved conducting surveys in collaboration with the National Science Board to solicit input from scientists on reducing burden and partnering with the American Association of Medical Colleges and the Council on Governmental Relations to identify areas in the regulations and guidelines that represent burdens to research. FASEB advocates for research by writing letters of support for science, factsheets for public education, and technical reports that address the current state of biomedical research. He called attention to FASEB's Database of U. S. Providers of Research Organisms, which could be a resource for the RII. In 2016, FASEB issued recommendations on

reproducibility and responded to the ORIP request for information on environmental factors in animal facilities. Dr. Haywood stated that the goal is to bring the scientific community together so that the best information can be gathered and synthesized into rational policy in support of the individual investigator.

Panel Discussion: Animal Resources and the Challenges of Unique Identifiers

Marine Biological Laboratory (MBL)—Marcin Wlizla, Ph.D., reported that the National *Xenopus* Resource (NXR) repository at MBL received its first animals in 2012. Currently, the repository houses 8,000 adult animals and 120 strains, including wild-type, inbred, and transgenics. The repository established clear guidelines for naming stock resources in Xenbase. In 2016, NXR began using RRIDs for older stocks. Dr. Wlizla noted the lack of awareness of the RII and RRIDs in the scientific community. The NXR shares stock resources with the European *Xenopus* Resource Center. Efforts to incorporate the practice of using RRIDs in the European stock center are ongoing.

Indiana University—Annette Parks, Ph.D., described the animal resources at the BDSC, which was established in 1986. As the primary *Drosophila* supplier, the BDSC maintains 120,000 living stocks and distributes 215,000 stocks annually to laboratories worldwide. The BDSC has worked to provide a website and support system to aid researchers in locating references on the history and genetic components of resources. Donor information is provided on the website and the BDSC database is aligned with FlyBase. RRIDs are provided on stock reports and user information sheets. Dr. Parks noted two challenges: lack of an authority to financially support genotyping and providing identifiers for stocks that are not included in a public collection; and the *Drosophila* is component-based, not organism-based. The BDSC supports use of a standardized reagent table.

University of Kentucky—Stephen Voss, Ph.D., reported on Sal-Site and the *Ambystoma* Genetic Stock Center. The axolotl *(Ambystoma mexicanum)* is unique and has the oldest laboratory pedigree of any laboratory animal, dating back to 1863. Satellite resources available in the United States and Europe date back to the Sal-Site axolotl. The axolotl research community is relatively small and supports the RII and RRIDs.

University of Minnesota—Ann Rougvie, Ph.D., described the animal resources of the *Caenorhabditis* Genetics Center (CGC). The CGC maintains 19,969 strains; approximately 30,000 strains are shipped annually, and shipments were sent to 1,481 laboratories in the last year. There are 1,279 laboratories with registered strains. Registered strains can be identified by a unique strain number. RRIDs are used, and resource information is deposited in WormBase. Dr. Rougvie noted compliance issues with RRIDs that she attributes to the lack of awareness of the RII. She also mentioned that journals are not encouraging investigators to use RRIDS.

University of Missouri—Elizabeth Bryda, Ph.D., presented on the resources at the Rat Resource and Research Center (RRRC). The RRRC archives and distributes rat models and embryonic stem cells to researchers and provides repository services to the biomedical community. Dr. Bryda pointed out that the RRRC RRIDs currently use RGD in the identifier string. There are plans to replace the RGD designation, which will require coordination with SciCrunch.

The Jackson Laboratory—Laura Reinholdt, Ph.D., reported on unique identifiers in the MMRRC at The Jackson Laboratory. The MMRRC displays RRIDs on the biorepository website. Further implementation of RRIDs in the mouse model will depend on support from the International Mouse Strain Resource.

University of Missouri—Craig Franklin, D.V.M, Ph.D., described the process to assign RRIDs in the MU-MMRRC. He reiterated the challenges to using genetically engineered mice in the context of rigor and reproducibility that were noted earlier by Dr. Korf.

University of California, Davis (UCD)—Dr. Lloyd highlighted the unique features of the UCD animal resources. The Knock-out Mouse Phenotyping Program (KOMP2), a source that produces models, is RRID compliant. The Mouse Metabolic Phenotyping Center, which provides phenotyping services, also is RRID compliant. The KOMP Repository, a distributor, is not RRID compliant. The UCD-MMRRC, one of four archive and distribution repositories in the MMRRC, is RRID compliant. Dr. Lloyd provided an example of how the RRID could enable seamless integration between producers, phenotypers, repositories, and researchers. He noted two guiding principles for resources and unique identifiers: (1) each stock must have nomenclature authority in a MOD; and (2) the identifier should persist and resolve in perpetuity, even if stock is no longer available.

The University of North Carolina (UNC) at Chapel Hill—Dr. Magnuson described the UNC-MMRRC's optimized MUGA method, the common platform for genetic quality control of mouse stocks and ES cell lines in the MMRRC. The method discriminates between hundreds of mouse inbred strains, is low cost and easy to interpret, and is available to investigators upon request.

Discussion

Neil M. Thakur, Ph.D., Office of the Director, suggested including RRIDs in the RPPRs.

Dr. Lloyd asked how the questions or concerns raised in today's workshop would be prioritized and addressed. Dr. Mirochnitchenko explained that all possibilities will be addressed. Participants are welcome to send any additional comments to DCM following the meeting. Monthly meetings with the RII team will be considered for the immediate future, and discussions with NIH Office of Science Policy (OSP) will continue.

Dr. Lloyd suggested establishing an Advisory Board consisting, in part, of MOD coordinators/organizers.

In response to a query by Dr. Westerfield on establishing NIH policies on the use of RRIDs, Dr. Mirochnitchenko replied that the goal is to work with OSP to consider those options.

Dr. Magnuson wondered about resolutions to RRIDs of mice obtained from the MMRRC who later show signs of genetic drift. Dr. Martone explained that the objective of the RRID is to document the source of the stock for authentication purposes. Any problems encountered should be captured in the MODs so that it will be transparent to other investigators.

Session 5: Hands-on Session with RRIDs

Dr. Bandrowski led participants through a hands-on exercise use of the SciCrunch RRID system to create an account, become owners of a resource, and link to ORCID.

Discussion

Dr. Magnuson voiced concern on receiving credit in ORCID for being a distributor, not a developer, of a stock resource, which is cited in a journal publication. Dr. Martone pointed out that ORCID allows users to fully account for all their work, not just publications. Resources can be credited to an individual and linked using the category for "other". Not all resources are associated with publications that are cited.

Dr. Bandrowsi added that the project with ORCID is in the early stages of development. Input on implementation strategies are encouraged.

D. Closing Remarks

Dr. Mirochnitchenko remarked on the progress of the RII, which has become an established program within the NIH. Further improvements or enhancements to the RII will rely heavily on contributions from workshop participants and the broader scientific community. He highlighted nine key points relative to tagging and identification of animal resources:

1. Incorporating use of the RRIDs in all phases of the research workflow from education to design of a research project to grant writing to publication is a best approach, but means that efforts to address authentication as well as rigor and reproducibility would need to begin earlier.

2. As digital identifiers become more embedded into the day-to-day life of the biomedical community, engaging with experts to integrate connections and interactions with other systems, academic or otherwise, will be critical.

3. RRIDs require improvements, especially in communication with resources/primary sources of information.

4. RRID system should interact with research resources, including model organisms used in biomedical research.

5. The RRIDs are not replacing the requirements for qualifying the primary source; the major load of animal/biomaterial information verification should lie on the resources.

6. Developing and supporting projects with other NIH ICs will help to further shape the RII and be of significant benefit to them, and will allow individual investigators supported by specific ICs to document and acknowledge the contribution, conduct data validation and identify critical elements for rigor and reproducibility of research.

7. The RRID reference system should contain additional data including environmental and condition-of-use information.

8. Increasing awareness of and educating the scientific community about the RII are needed, including international community and biomedical research societies, which should help with implementation and policy recommendations.

9. Additional ways to encourage investigators to use RRIDs need to be explored and applied by all stakeholders.

Appendix A: Workshop Agenda

8:00 – 8:30 **Registration**

8:30 – 8:45 **Introduction and Welcome**

Dr. Oleg Mirochnitchenko, ORIP, NIH
Dr. Stephanie Murphy, Division Director, ORIP, NIH

Objectives: (1) Introduce the goals of the meeting; (2) Describe the mission of the DCM/ORIP and the needs of the biomedical research community and NIH program on standards and sufficient unique identifying information for research resources; (3) Introduce the meeting participants and agenda.

8:45 – 9:15 **Rigor and Reproducibility in Biomedical Research – Current NIH Guidelines**

Dr. Carrie D. Wolinetz, OD, NIH

Objectives: (1) Introduce the concepts of rigor and reproducibility in biomedical research; (2) Overview of current NIH guidelines for authentication of key biological resources; (3) Development of new research standards and available resources.

Session 1: Identification of Animal Resources in Scientific Literature and Grant Reports

9:15 – 9:35 **Dr. Anita Bandrowski, SciCrunch Inc, San Diego, CA**

Objectives: (1) Introduce the existing problems with unique identification of resources; (2) SciCrunch and development of the Research Resource Identification System (RRID); (3) Curation of NIH-supported animal resources and dissemination of a unique identification system.

9:35 – 9:55 **Dr. Valentina Di Francesco, NHGRI, NIH**

Objectives: Current status of Data Commons Pilot, Model Organism Databases.

9:55 – 10:15 **Dr. Kristin Abraham, NIDDK, NIH**

Objective: To introduce the NIDDK Information Network (dkNET) www.dknet.org, and its role in supporting NIDDK's efforts to expand the use of RRIDs across the NIDDK research community.

10:15 – 10:25 **Break**

Session 2: Animal Repositories and Their Role in Supporting High Standard Research

10:25 – 10:45 **Dr. Monte Westerfield, University of Oregon, OR**

Objectives: (1) Describe the Zebrafish International Resource Center (ZIRC) and its distribution activities; (2) Discuss the challenges of cataloging and monitoring the use of zebrafish mutant strains.

10:45 – 11:05 **Dr. Ian Korf, University of California, Davis, CA**

Objectives: (1) Describe the Mutant Mouse Resource and Research Center consortium (MMRRC) and its mission; (2) Discuss the need for mouse strain universal tagging and efforts to monitor the use of distributed animal resources.

11:05 – 11:25 **Dr. Madeline Crosby/Norbert Perrimon, Harvard University, MA**

Objectives: (1) Describe FlyBase, an online bioinformatics database and the primary repository of genetic and molecular data for the insect family Drosophilidae, and the Bloomington Drosophila Stock Center; (2) Illustrate the challenges and provide the potential measures to improve the ability of this repository as well as researchers to unambiguously identify the fly model and cite it in literature; (3) Introduce the standardized author reagent table that FlyBase (in consultation with other MODs) has developed and the response they have gotten from journals.

11:25 – 12:25 **LUNCH –** Meals and light refreshments are at the expense of attendees. (Attendees will be responsible for meals/light refreshments on their own, at their own cost. The government and/or government contractors are not involved in facilitating the provision of food and/or light refreshments.)

Session 3: Common Publication Guidelines for Citing Animal Resources

12:25 – 12:35 **Dr. Emma Ganley, Chief Editor, *PLOS*, CA**

12:35 – 12:45 **Dr. Ann Goldstein, *Neuron*, MA**

12:45 – 12:55 **Dr. Natalie De Souza, *Nature Methods*, NY**

Objectives: (1) Describe the need for improvement of reporting of animal resources in biomedical publications; (2) Introduce new instructions for authors and editorial oversight; (3) Comment on the need of the cultural shift and development of text mining tools.

13:05 – 13:25 **General Discussion With Journal Editors**

Session 4: Future Development of Animal Resource Identifiers

13:25 – 13:45 **Dr. Maryann Martone, University of California, San Diego, CA**

Objectives: (1) Describe SciCrunch achievements and future developments of the RRID initiative; (2) Provide current statistics on RRID usage by the biomedical research community and publishers; (3) Provide plans on coordinating RRID activities with other efforts on scholarly data citation.

13:45 – 14:05 **Dr. Laurel L. Haak, Executive Director, ORCID, MD**

Objectives: (1) Describe ORCID, Cross Ref, and the use of Digital Object Identifiers; (2) Discuss the requirements of a unique identification system; (3) Comment on ways to perform information validation.

14:05 – 14:25 **Dr. J. R. Haywood, FASEB, MI**

Objectives: (1) Describe the FASEB efforts in supporting animal welfare standards and practices; (2) Highlight the critical role of animal use standards in supporting rigor and reproducibility of animal research; (3) Discuss the need of collaboration among different stakeholders: researchers, government agencies,

publishers, informaticians, curators, and animal resource directors.

14:25 – 14:35 **Break**

14:35 – 15:25 **Animal Resources and the Challenges of Unique Identifiers**

Dr. Marcin Wlizla/Marko Horb (Marine Biological Laboratory, MA)
Dr. Annette Parks/Kevin Cook (Indiana University, IN)
Dr. Stephen Voss (University of Kentucky, KY)
Dr. Ann Rougvie (University of Minnesota, MN)
Dr. Elizabeth Bryda (University of Missouri, MO)
Dr. Ron Walter (Texas State University, TX)
Dr. Laura Reinholdt (The Jackson Laboratory, ME)
Dr. Craig Franklin (University of Missouri, MO)
Dr. Kent Lloyd (University of California, Davis, CA)
Dr. Terry Magnuson (University of North Carolina, NC)

15:25 – 16:00 **Discussion**

Session 5: Hands on Session with RRIDs

16:00 – 17:00 **Dr. Anita Bandrowski, SciCrunch Inc., San Diego, CA**

Objectives: (1) Introduce the RRID website and describe the steps needed for obtaining a unique identifier for an animal resource; (2) Describe and demonstrate how to use and validate identifiers; (3) Show the demo version of the dashboard for tracking identifier citations and explain how to use it; (4) Train workshop participants to use the tools.

17:00 – 17:20 **Closing Remarks/Recommendations**

Adjournment

Appendix B: Workshop Participants

Kristin Abraham, Ph.D.
National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH
abrahamk@mail.nih.gov

Anita Bandrowski, Ph.D.
SciCrunch Inc, San Diego, CA
abandrowski@ncmir.ucsd.edu

Patricia Brown, V.M.D.
Office of Laboratory Animal Welfare, Office of the Director (OD), NIH
brownp@mail.nih.gov

Elizabeth Bryda, Ph.D.
University of Missouri, MO
brydae@missouri.edu

Ben Busby, Ph.D.
National Center for Biotechnology Information, National Library of Medicine (NLM), NIH
busbybr@ncbi.nlm.nih.gov

Devon Crawford, Ph.D.
National Institute of Neurological Disorders and Stroke, NIH
devon.crawford@nih.gov

Madeline Crosby, Ph.D.
Harvard University, Boston, MA
crosby@morgan.harvard.edu

Natalie De Souza, Ph.D.
Nature Methods
n.desouza@us.nature.com

Valentina Di Francesco
National Human Genome Research Institute, NIH
vdifrancesco@mail.nih.gov

Paul Donohoe
Somark Innovations, UK
pdonohoe@somarkinnovations.com

Valerie Florance, Ph.D.
NLM, NIH
florancev@mail.nlm.nih.gov

Craig Franklin, D.V.M., Ph.D.
University of Missouri, MO
franklinc@missouri.edu

Emma Ganley, Ph.D.
Chief Editor, *PLoS*, CA
eganley@plos.org

Gallya Gannot, D.M.D., Ph.D.
National Institute of Dental and Craniofacial Research, NIH
gallya.gannot@nih,gov

Ann Goldstein, Ph.D.
Neuron, MA
agoldstein@cell.com

Rebecca Goodwin, J.D.
NLM, NIH
rebecca.goodwin@nih.gov

Jeffrey Grethe, Ph.D.
University of California, San Diego, CA
jgrethe@ucsd.edu

Franziska B. Grieder, D.V.M., Ph.D.
Office of Research Infrastructure Programs (ORIP), OD, NIH
griederf@mail.nih.gov

Laurel L. Haak, Ph.D.
Executive Director, ORCID, MD
l.haak@orcid.org

Brian Haugen, Ph.D., (Reporter)
Office of Extramural Research (OER), OD, NIH
haugenb@mail.nih.gov

J. R. Haywood, Ph.D.
Federation of American Societies for Experimental Biology, MD
haywoo12@msu.edu

Alena Horska, Ph.D.,
ORIP, OD, NIH
alena.horska@nih.gov

Pritty Joshi, Ph.D.
Division of Planning, Evaluation & Analysis, OD, NIH
pritty.joshi@nih.gov

Ian Korf, Ph.D.
University of California, Davis, CA
ifkorf@ucdavis.edu

Kent Lloyd, D.V. M., Ph.D.
University of California, Davis, CA
kclloyd@ucdavis.edu

Terry Magnuson, Ph.D.
The University of North Carolina at Chapel Hill, NC
tmagnuson@unc.edu

Maryann Martone, Ph.D.
University of California, San Diego, CA
mmartone@ucsd.edu

Willie McCullough, Ph.D.
ORIP, OD, NIH
mccullow@mail.nih.gov

Oleg Mirochnitchenko, Ph.D.
ORIP, OD, NIH
oleg.mirochnitchenko@nih.gov

Stephanie Murphy, VM.D., Ph.D., DACLAM
ORIP, OD, NIH
stephanie.murphy@nih.gov

Dina N. Paltoo, Ph.D.
Office of Science Policy, OD, NIH
dp249p@nih.gov

Annette Parks, Ph.D.
Indiana University, IN
anlparks@indiana.edu

Jonathan Pollock, Ph.D.
National Institute on Drug Abuse, NIH
jpollock@ngmsmtp.nida.nih.gov

Laura Reinholdt, Ph.D.
The Jackson Laboratory, ME
laura.reinholdt@jax.org

Molly Romick
Charles River Laboratories, MA
molly.romick@crl.com

Ann E. Rougvie, Ph.D.
University of Minnesota, Twin Cities, MN
rougv001@umn.edu

Cynthia L. Smith, Ph.D.
Mouse Genome Informatics, The Jackson Laboratory, ME
cynthia.smith@jax.org

Neil M. Thakur, Ph.D.
OER, OD, NIH
thakurn@mail.nih.gov

Patricia A. Valdez, Ph.D.
OER/Office of Extramural Programs, OD, NIH
patricia.valdez@nih.gov

Desiree Von Kollmar, M.S.
ORIP, OD, NIH
drat@mail.nih.gov

Stephen Voss, Ph.D.
University of Kentucky, Lexington, KY
srvoss@uky.edu

Ron Walter, Ph.D.
Texas State University, TX
rwalter@txstate.edu

Harold L. Watson, Ph.D.
ORIP, OD, NIH
watsonh@mail.nih.gov

Monte Westerfield, Ph.D.
University of Oregon, Eugene, OR
monte@uoregon.edu

Marcin Wlizla, Ph.D.
Marine Biological Laboratory, MA
mwlizla@mbl.edu

Carrie D. Wolinetz, Ph.D.
Office of Science Policy, OD, NIH
carrie.wolinetz@nih.gov

Wang Xujing Ph.D.
NIDDK, NIH
wangx21@mail.nih.gov

Max Zaretskiy
Highrise Consulting, Inc.
max@highriseconsulting.com

Sige Zou, Ph.D.
ORIP, OD, NIH
zous@mail.nih.gov