

Questionnaire:

*Section 1: Consent*

Thank you for your interest in participating in the research project “Peer Review of Research Data Submissions to ScholarsArchive@OSU”. The goal of this research project is to evaluate the reusability of the research datasets hosted in Oregon State University’s institutional repository, ScholarsArchive@OSU, using domain level expert peer-review. Currently, it is not clear how useable datasets in ScholarsArchive@OSU are to researchers in the field for which the data was originally published. We also want to learn more about what makes a dataset review process useful, so that the efforts of the researchers and the repository reviewers are focused in aspects that enhance the reusability of the datasets. Your participation in this study is voluntary, but all required questions must be answered in order for your responses to be included in the study. The only piece of personal information that will be collected in the questionnaire is your name and e-mail, so that we can manage the answers that we receive. Once our dataset is complete, and before we analyze the dataset, we will remove the names of the reviewers, so your answers will not be linked to your name. Information about the dataset you are reviewing will be in the survey. We will use all the reviews to research the reusability of the datasets in our institutional repository, will publish our findings in one (or more) research articles in scientific journals, and may present our results at professional conferences. Anonymized aggregated data will be shared in a data repository, ScholarsArchive@OSU, for potential future reanalysis. We may use quotes from your review in our publications. Again, this information will never be linked to your name or your personal information in any way. The risks for you associated to your participation in this study are small, because we are not collecting any personal information about you in our survey, only your opinion about the dataset you will be reviewing. There is a

small possibility that you could be identified by the information you give in your responses associated to the dataset your are reviewing. The communications that we maintain by e-mail could be intercepted, arrive late or incomplete, or contain viruses. We will make every effort to minimize these risks. There may be risks related to the study that are not yet known to the researchers. This study is conducted by AUTHOR and AUTHOR at the INSTITUTION Library. If you have any questions feel free to contact AUTHOR (Principal Investigator; AUTHOR'S E-MAIL).

Do you consent to participate in this study?

I consent to participate in this study (1)

I do not consent to participate in this study (2)

Rule: If consent is given, proceed to the rest of the questionnaire.

### *Section 2: Information about the reviewer*

Please provide information about the reviewer and the dataset that will be reviewed.

Q1 What is your Name? [text box]

Q2 What is your e-mail address? [text box]

Q3 What is the title of the dataset you are reviewing? [text box]

Q4 Provide the link to the dataset you are reviewing. [text box]

### *Section 3: Repository Record*

Please evaluate the quality of the repository record. This is the page of ScholarsArchive@OSU that has information about the dataset and gives you access to the files.

Q5 Is the record sufficiently descriptive to help you determine whether the dataset is of interest?

Do you find the abstract and/or keywords helpful in evaluating the dataset? [text box]

Q6 Are there other elements that could be added to the record that would help a user evaluate the relevance of the data? [text box]

Q7 What is your overall evaluation of the quality of the Repository Record? Give a value between 1 (very bad) and 10 (excellent). [scale 1-10]

#### *Section 4: Documentation*

Please evaluate the quality of the documentation accompanying the dataset. This would be one of the files that you will download. It is usually (but not always) named readme.txt .

Q8 Does the dataset documentation contain contact information for a responsible party? Does the contact information appear to be robust to changes in personnel and organizational affiliations of researchers? [text box]

Q9 Does the documentation provide sufficient contextual information to help you understand why the data was created and what it was originally used for? [text box]

Q10 Does the document provide a comprehensive description of all the data that is there? Is there sufficient quality information describing the format and content of the data, including column header descriptions, units, etc? [text box]

Q11 Are the methods and any data-processing steps described in sufficient detail to allow others to reproduce these steps? Is the processing chain visible and well documented? [text box]

Q12 Can all internal references ( e.g. URL/DOI) be resolved to real entities? Are the external electronic references stored in a trusted repository? [text box]

Q13 Does this documentation, in combination with the repository record metadata, give you enough information to determine your rights to reuse the dataset? [text box]

Q14 What is your overall evaluation of the quality of this dataset's documentation? Give a value between 1 (very bad) and 10 (excellent). [scale 1-10]

*Section 5: The data itself*

Please evaluate the quality of the dataset itself. These would be the downloadable files from the record, excluding the documentation files.

Q15 Are the data easily readable, e.g. do they use standard or community formats? [text box]

Q16 Are the data of high quality e.g. are error limits and quality statements adequate to assess fitness for purpose, is spatial or temporal coverage good enough to make the data useable? [text box]

Q17 Are the data values physically possible and plausible? [text box]

Q18 Are there missing data that might compromise its usefulness? [text box]

Q19 What is your overall evaluation of the quality of this dataset? Give a value between 1 (very bad) and 10 (excellent). [scale 1-10]

*Section 6: Overall quality*

Evaluate the overall quality of the dataset

Q20 Overall, is enough information provided (in metadata also) to enable the data to be re-used? Give a value between 1 (dataset not reusable at all) and 10 (dataset very easily reusable). [scale 1-10]

Q21 What did you learn about your own data sharing practices from conducting this peer review? [text box]

Q22 Is there anything else you would like to share with us? [text box]