**Data & Metadata Profile**

McFadden-Hiller, J. E., Beyer, D. E. Jr., & Belant, J. L. (2016) Spatial Distribution of Black Bear Incident Reports in Michigan [data set]. Available as Supporting Information for: *PLoS ONE 11*(4): e0154474. <https://doi.org/10.1371/journal.pone.0154474>

Accessed from: <https://figshare.com/articles/dataset/Spatial_Distribution_of_Black_Bear_Incident_Reports_in_Michigan/3208564?file=5037811>

The data selected for this profile are public reports of black bear incidents in Michigan between 2003 and 2015. The data also include the region and location of the encounters, as well as the land cover of the area. These data come from reports received by the Michigan Department of Natural Resources (MiDNR) of bear related incidents. Dean E. Beyer compiled these data from MiDNR. Also included are data from eMODIS Normalized Difference Vegetation Index (NDVI) and National Land Cover Database (NLCD) based on the geographic regions of study in regard to vegetative food sources and land cover respectively.

Key stakeholders of the data are the three creators credited in the repository record and their supporting institutions. Two of the creators are researchers with the Carnivore Ecology Laboratory of the Forest and Wildlife Research Center at Mississippi State University. The other creator is associated with the Wildlife Division of MiDNR. In addition, wildlife managers and property owners looking to better understand circumstances and reduce black bear incidents, researchers investigating ecological impact on human-animal interactions have potential interest in the data.

There are two .xslx data files included in the dataset. Each file contains a metadata tab, and a data tab. In the data tab are entries for black bear encounters, randomized units from each geographic region used for comparison and modeling (McFadden-Hiller et al., 2016), and corresponding ecological data for each entry. The entries in “S1 Dataset” are encounters from 2003-2011 and were used in analysis; the entries in “S2 Dataset” are from 2012-2015 and were used in the evaluations of models (McFadden-Hiller et al., 2016). The data are under a Creative Commons Attribution 4.0 International (CC by 4.0) license, meaning that the data are able to be shared and adapted, but must attribute credit to the creators and indicate if any changes were made. The data are Excel Workbooks (.xslx files) and while a proprietary format, they can be accessed using a variety of widely available software including Excel, Numbers, Google Docs, or open software like OpenOffice.

The metadata tabs in each of the data files contains descriptions of the contents for the column headings and do not include any other information. The files themselves are not named in a way that makes them easily identifiable, and there is no title or creator information included in the metadata. There is no inclusion of the methods used in gathering the data or how the models were computed. Metadata citations can be exported from the repository as XML files according to DataCite and Dublin Core schemas among others. These metadata do include title, creator, identifier, relation, description, date and subject elements with appropriate values. They are however, separate from the dataset download and require intentional searching to locate and export when accessing the repository. Additionally, the subjects listed make the exported metadata seem auto generated by the repository, not intentionally described by the creator, as they do not fully represent the content of the data. The repository website misrepresents the data by including “Hematology” and “Cancer” as related categories to the dataset which are not relevant to the research at all.

In order to improve the discoverability of the data set in the repository the keyword and category information would need to be modified to better reflect the content of the data. Metadata including license information, geographic or temporal coverage, and accurate subject keywords would also help with discoverability in the repository (Wu et al., 2019). Some of these metadata may have been provided by the creator when submitting the data set to the repository, but are not included with the data download and would be useful to have if the data were to be separated from the record or placed within another repository. DataOne also recommends including instructions on how to cite the dataset in the metadata which could also aide in discovery (Strasser et al., n.d.).

To make the data more useable by new researchers, more information about how the data were collected and generated would be needed in the metadata. As described in the DataOne Best Practices Primer, the metadata should include by whom, why, and how the data were collected (Strasser et al., n.d.). Descriptions of the column headers are helpful in interpreting the data, but the parameters for calculations are not included, which makes reproducing and verifying the calculations impossible, and makes using the data more difficult. Including these fields in the metadata would assist users in understanding and using this data in new research.

The data are linked to an accompanying article that was published in *PLOS One* called Spatial Distribution of Black Bear Incident Reports in Michigan (McFadden-Hiller et al., 2016). In Google Scholar six publications cite the accompanying article when searching its DOI, but do not cite the data directly. Searching the DOI of the data specifically only results in the discovery of its own analysis in the original publication. According to the Figshare repository the data have been downloaded 38 times, so there is a possibility that there have been or will be citations of the data in the future. Without any title, creator, or relation metadata included in the dataset it is feasible that it may have been referenced without the DOI or other proper citation.

When these data are directly linked to their accompanying article, the metadata included in the dataset make it easier to interpret the figures. Without, there is less clarity and detail in the metadata which could impede identification or reuse of the data. Exported XML files do assist in remedying this problem, but they do not completely represent the data and are not a part of the dataset provided. Additional metadata would make these data more discoverable and would help to ensure their longevity and usability.

**Repository Profile**

Repository Chosen: Forest Service Research Data Archive, <https://www.fs.usda.gov/rds/archive/>

The repository chosen for the data on Spatial Distribution of Black Bear Incident Reports in Michigan is the Forest Service Research Data Archive (FSRDA). The repository is described by the Registry of Research Data Repositories as a “repository for the long-term preservation and distribution of citable research data sets that are broadly relevant to forest and grassland ecology, and the economic and social interactions of humans with these ecosystems” *(2019, General Description*). While the FSRDA is geared towards employees of the United States Forest Service (USFS), submissions are open to non-employees as well. The FSRDA is a small repository with a focused collection policy. Putting the data in a larger repository with a general focus may have increased the volume of people who could potentially see the data; submitting the data to a more focused repository will allow it to be more easily discovered by those interested in similar areas of study and will hopefully add to the possibility of its reuse. More general repositories, like Dryad, require a CC0 (No Rights Reserved) license for any submitted data. The data for black bear incidents are under a CC BY 4.0 (Attribution 4.0 International) license which is incompatible with that requirement. The FSRDA has a requirement for Data Citation which fits with the CC BY 4.0 license.

The FSRDA has a collection policy that limits submissions to those from Forest Service researchers, Joint Fire Science Program-funded researchers, or data that researchers have collected with Forest Service funds or on Forest Service lands. Those who are not affiliated with the USFS must contact the archive team in order to submit data. The data in the dataset come from reports from the Michigan Department of Natural Resources (MiDNR) and include data collected on USFS land so they would fit under this collection policy. MiDNR operates in conjunction with the USFS under “good neighbor authority” (Michigan Department of Natural Resources, 2016; US Forest Service, n.d.), meaning that they have formal arrangements to work together to support organizational needs. Although the data comes from MiDNR reports, this formal partnership with the USFS enforces the idea that this data would be beneficial to have in a FSRDA. Users of and other contributors to the FSRDA would have potential interest in this data either when looking at regional data for Michigan or in studying ecological or anthropogenic effects on bear incidents across regions.

The FSRDA supplies potential submitters with guidelines for what should be included in the Data Submission Package on their website. Data Submission Packages should include data sets, metadata, any additional files that should be archived with the data like maps or research notes, and a file index. The Data Submission Package is then submitted to the archival team along with any associated publications or links to online publications and a completed and signed submission form. The repository requires consulting with the archival team before any data submission occurs to determine how data should be sent. The archival team is also available to contact for any questions in preparing data, and they provide multiple points of education about and resources for assistance in writing and formatting metadata. Included in the provided support is a “How To” document (Forest Service Research Data Archive, n.d.-b) that walks researchers through the process of preparing files and developing metadata as well as compiling and submitting the data package. Appendices in this document provide an example File Index and descriptions of important metadata fields. The FSRDA uses version 2 of the Federal Geographic Data Committee’s (FGDC) Biological Data Profile of the Content Standard for Digital Geospatial Metadata (BDP). Other standards are allowed, however, if the data necessitates. Metadata are preferred in HTML or XML format but can be converted by the archival team if that is not possible. The requirements for metadata standards may change as the FGDC endorsed ISO 19115: Geographic information – Metadata in 2010 (Federal Geographic Data Committee, n.d.). The USFS mentions that when a compliant North American Profile of ISO 19115 is available it is expected to be made a federal standard and conversion tools for the metadata will be made available (Forest Service Research Data Archive, n.d.-a).

In order to access data within the repository, there is no need to create a registered account or login. Datasets are found through searching the data catalog with a keyword, phrase, title, author, or DOI search. Datasets can then be sorted by things like funding agency and year of publication. The only publicly available download option for datasets is through direct link, assumably because it is a smaller database. More options for data access may become available as the database gains more datasets. With the archive team available to consult on data submissions and to help convert metadata formats, it is possible that they would be able to make arrangements for other access mechanisms like an automated script. The file directory and metadata can be viewed separately before downloading the data publication package and are also included in the package along with the dataset and any supplemental files provided by the author.

While the FSRDA is a focused repository with a smaller number of datasets, it is a government affiliated archive that provides an archival team to support all researchers submitting data. The repository ensures that data and metadata archived are checked for accuracy and completeness. The long-term preservation of these data in a focused repository will help to increase use of the data within related fields.

**Recommended Citation**

APA: McFadden-Hiller, J. E., Beyer, D. E. Jr., & Belant, J. L. (2016) Spatial Distribution of Black Bear Incident Reports in Michigan [data set]. Available as Supporting Information for: *PLoS ONE 11*(4): e0154474. https://doi.org/10.1371/journal.pone.0154474

**Considerations for Long-term Preservation**

In considering the long-term preservation of this data, it should be noted that the files are in a proprietary format, as .xslx files. They are accessible through open-source software.

**Copyright License**

The data are covered under a Creative Commons Attribution 4.0 license (CC BY 4.0). This means that the data may be shared and adapted but must attribute credit, link to the license, and indicate any changes. Full license information can be found at <https://creativecommons.org/licenses/by/4.0/legalcode> and should be reviewed before use.

**Human Subjects Consideration**

There are no human subject considerations for this dataset.

**References**

Federal Geographic Data Committee. (n.d.). *ISO Geospatial Metadata Standards*. Retrieved February 11, 2021, from https://www.fgdc.gov/metadata/iso-standards

Forest Service Research Data Archive. (n.d.-a). *Metadata Standards*. Retrieved February 11, 2021, from https://www.fs.usda.gov/rds/archive/metadata/standards

Forest Service Research Data Archive. (n.d.-b). *Submitting Data*. Retrieved February 10, 2021, from https://www.fs.usda.gov/rds/archive/submittingData

McFadden-Hiller, J. E., Beyer, D. E. Jr., & Belant, J. L. (2016). Spatial Distribution of Black Bear Incident Reports in Michigan. *PLOS ONE*, *11*(4), e0154474. https://doi.org/10.1371/journal.pone.0154474

Michigan Department of Natural Resources. (2016, July 19). *DNR and U.S. Forest Service cooperate on “Good Neighbor Authority.”* https://www.michigan.gov/dnr/0,4570,7-350-79137\_79770\_79873\_80003-388998--,00.html

Registry of Research Data Repositories. (2019, August 7). *Forest Service Research Data Archive*. http://doi.org/10.17616/R3404J

Strasser, C., Cook, R., Michener, W., & Budden, A. (n.d.). *DataOne Primer on Data Management: What you always wanted to know*. 11.

US Forest Service. (n.d.). *Shared Stewardship*. Retrieved February 10, 2021, from https://www.fs.usda.gov/managing-land/shared-stewardship

Wu, M., Psomopoulos, F., Khalsa, S. J., & de Waard, A. (2019). Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. *Data Science Journal*, *18*, 3. https://doi.org/10.5334/dsj-2019-003