**User Stories**

Users seeking data:

|  |  |
| --- | --- |
| Goal | User Story |
| Find data about the organizations’ activities | I am a government official for the City of Oakland. I want to locate data about how many animals are taken in and adopted out or otherwise leave the city’s shelters to determine the staffing need for each shelter. |
| Find data about revenue | I am a philanthropist looking to donate money to local shelters. I want to view data about shelter revenue and costs to determine which shelters experience the highest funding needs. |
| Use data for education and activism | I am an animal rights activist looking to compile informational resources to educate the community on the need to adopt animals instead of purchase from breeders. |

Users contributing data:

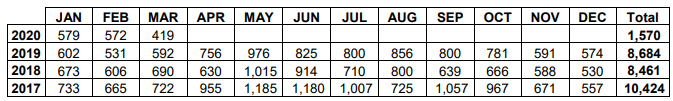
|  |  |
| --- | --- |
| Goal | User Story |
| Understand ideal data submission frequency | I am an animal shelter employee looking to find out how frequently I should submit new data to the repository. |
| Understand terminology the repository will use | I am an animal shelter director looking to find out how shelters will be classified (e.g. “animal pound,” “kill shelter,” “no-kill shelter,” “adoption facility”) to determine the impact the labeling will have on the facility’s reputation and public image. |
| Understand metadata requirements | I am in charge of recordkeeping at an animal shelter looking to determine the metadata I should be including for the data I submit with my data packages. |

**Use Cases**

* You should begin with a brief 3-5 sentence overview of the data, publisher, and the relevance to your quarter project.

This dataset I am using for this use case is published by the Contra Costa County Animal Services agency. The data is the March 2020 update of the total animal intakes, intake types, intake breeds, intake ages, and similar data for all “outcomes,” or animals leaving the facilities. This is relevant to my quarter project because it is data about a Bay Area animal shelter, which is the data for which I am creating a protocol.

* Elements that characterize the data available (or at least published in summary form) on the web:
  + Domains served by the data: Animals, Animal Welfare, Pets, Government Agencies, Non-Profit Organizations
  + Obligation / Motivation for publishing the data: The Contra Costa Animal Shelter Performance Reports landing page states: “The monthly reports compare operational performance for Animal Services in various areas against our performance from the prior year.” (<https://www.contracosta.ca.gov/6820/Performance-Reports>)
  + Potential ways that the domains will use the data: Data will be used by citizens, animal welfare organizations, government agencies, and philanthropists.
  + Quality: Because this is a PDF containing mostly graphs and charts with a few tables, the data is mostly unstructured. The tables showing monthly figures are structured but have no context. For example, the following table shows figures for each month from January 2016 to Match 2020, but it does not give any indication of what the data is for. These numbers could show the number of animals taken in, number of animals euthanized, number of bags of dog food purchased, number of times a dog was taken for a walk, etc. Without context, the data really has no structure.



* + Size: It is difficult to determine the size of the dataset because of its presentation as a report with charts and tables.
  + Type / Format: PDF with charts and tables.
  + Rate of Change: A new report is published monthly, adding on to the report from the previous month. At the end of the year, a year-end report is compiled, and individual monthly reports are no longer available. This report is for March 2020. January 2020 and February 2020 are also available. In theory, April through December 2020 will be available, and then a 2020 annual report will be published, and the individual monthly reports will be removed from the website.
  + Lifespan: The lifespan of this data should be indefinite because it is historical data that will not change.
  + Potential audiences: Citizens, animal welfare organizations, government agencies, and philanthropists.
* Positive aspects of the published data.

This particular report includes historical data and the new data simply builds on it. Rather than submitting only numbers for March 2020, the report includes March 2020 as well as all data going back to January 2016. It would be easy to integrate all of this data at once rather than having separate packages for each month or year.

This data also adheres to standards published by the Asilomar Accords (<https://www.shelteranimalscount.org/docs/default-source/DataResources/2004aaccords5.pdf?sfvrsn=31c1ff76_0>). These standards govern data collection “in order to facilitate the data collection process and assure consistent reporting across agencies.”

* Negative aspects of the published data.

This data is published in PDF format, which is not easy for intake into a repository that integrates data from many different datasets. This data would need to be submitted in a standardized data package that is more easily integrated with other data, such as a .csv file.

* Challenges that will face data publishers, users, or meeting best practices over time. See: <https://www.w3.org/TR/dwbp-ucr/#general-challenges>

Publishing this data in PDF format with charts and tables instead of a structured table presents a major challenge to data curators who want to integrate this data with other datasets.

* Requirements for DWBP: Select two requirements for your data source, and justify why this is a requirement (~3 sentences) (See: <https://www.w3.org/TR/dwbp-ucr/#requirements-1>

1. **R-FormatMachineRead:** *Data should be available in a machine-readable format that is adequate for its intended or potential use*

I selected this requirement because it is crucial for the submitted data to be machine-readable so it can be integrated with other data. Although a PDF is technically readable, most of the data in a PDF is noise and is not relevant to the repository, at least insofar as integrating it with other data. This repository essentially will have integers, dates, and short text descriptions (e.g. “dog”), whereas the data in a PDF has a lot of unnecessary text that would make it difficult to quickly add to the repository.

1. **R-FormatStandardized:** *Data should be available in a standardized format. Through standardization, interoperability is also expected.*

This requirement is related to the machine-readability format. The data should be standardized so that the datasets from various animal shelter agencies that will be submitting data packages that contain the same type of information. The information included in the Contra Costa Animal Services dataset is adequate, but not all datasets include all of this information. Adhering to the Asilomar Accords standards is a good start to ensure the data is standardized upon receipt by the repository.

1. **R-AccessBulk:** *Data should be available for bulk download*

This requirement relates to the other two requirements. The data should be able to be downloaded in bulk to facilitate expeditious submission to the repository.