Cristyn Filla

Data Curation I

Repository Profile

February 10, 2023

My chosen dataset for this quarters final project contains data on the rock-climbing routes found in Yosemite National Park. When considering what repository I would select to store and disseminate the curated data that I produce from this dataset, the goal user for this data was at the forefront of my mind. I had a couple of options for who I could identify the goal user as, including those interested in land use data such as the National Park Service (NPS). However, since I was drawn towards this dataset as a result of my own interest in the data as a climber, I chose to prioritize the recreational climbing community as my ideal user population. (Though it should be noted that a repository which stores data for climbers on climbing routes will likely still be of use to NPS and others interested in this data for non-recreational purposes).

After having no success finding the recreational data repositories, I believed would be most likely to interface with the goal users on the re3data database I began to search for repositories the user community may be using already by using key search terms in Google and Reddit. This search turned up a limited selection of options, however, there was one repository that caught my eye due to its community-centric mission and messaging. OpenBeta is the repository I chose to house my climbing dataset as it has an open-data policy and is actively being used by climbers to research and locate climbing routes across the United States. By choosing this repository my dataset will maintain a high level of visibility and accessibility by the climbing community. The passion and expertise of the volunteer team running OpenBeta helped to support this decision. Given the often informal, “clique”-like nature of the climbing community, finding a formalized system for information storage and dissemination is challenging, however, OpenBeta appears to have sparked a community movement led by volunteers who are also professionals in data and information fields. This lends integrity and thoroughness to the repository maintenance.

The OpenBeta repository is a crowdsourced repository of climbing route data from across the U.S. This means that anyone can submit data to the repository however, the data that is submitted must be relevant to the collection scope -- climbing. In order to contribute to the repository, however, the contributor does need to go through the minimum threshold of downloading and accessing the platform OpenTacos which is used for data submission. While account creation, login, or ownership is not required to access the OpenBeta site or to access or download the data from the repository, account creation and login are required to access the submission platform. Creating a login for OpenTacos, however, is extremely straight forward and does not require any monetary investment.

Once logged into OpenTacos the submitter is able to submit JSON, CSV, or Phyton pickle files to the repository. (Conversely the data stored in the OpenBeta repository can be accessed in any of these file formats). Direction on how to convert your dataset into one of these three formats is provided by the repository. The official style guide for the data is still a work in progress, however, the data must be substantiated observational and/or experiential data within the domain of outdoor recreation/climbing. There is guidance on the data and metadata that should be submitted, however, since the repository is working to fill gaps in on the ground knowledge for climbers there is a “more is more” mind set around unstructured data pertaining to route location, description, and beta (i.e. guidance on how to successfully perform the climb). Metadata for grade (climb difficulty) and area/region (latitude and longitude) is standardized for searchability and useability.

While, as mentioned previously, the style guide for the data is still a work in progress, OpenBeta does provide what is essentially a Submission Information Package (SIP) to those looking to contribute to the repository. This includes information on climb grading evaluation to help determine the difficulty rating to be applied to climbs in the data set; a manual on how to use the data submission platform, OpenTacos; and expectations around data submission, including describing the nature of open-source data so that the data contributor has full transparency on how the data they add to the repository may be used/accessed by others.

There is a team of volunteers able to assist in the submission process. This team can be accessed through the support line email provided on the site. However, working with a volunteer for the repository in real-time may prove to be difficult given the ad hoc manner of the volunteer positions and repository maintenance. In order to supplement this human assistance, OpenBeta also maintains a GitHub page that includes instructional materials on submission processes and protocol as well as examples of how to apply the OpenBeta standards to datasets as needed.

There are several methods of access on OpenBeta. The one that likely experiences the highest level of user interface is the database query set up on the OpenBeta home page. This allows to search for climbs or areas across the repository and access data that contains those key words or tags. Alternatively, direct file download is available by navigating to OpenBeta’s GitHub page. While metadata is consistent across entries there is no clear metadata standard being followed by OpenBeta and no directions on metadata standards for submitters.

The OpenBeta repository is an open data repository and thus provides a minimal dissemination information package. This is due to the fact that all data placed in the repository is open for dissemination, sharing, and use by anyone who accesses the OpenBeta site. Information regarding the open access of all data stored on OpenBeta is provided to the submitter prior to submission so that they are aware of the dissemination implications of submitting data to OpenBeta.

Overall, while OpenBeta is in some ways a repository that is still in its infancy and working through developments, I believe that it remains an excellent repository choice for the Yosemite National Park climbing route data in question as it ensures clear and fair access of the data to those most likely to use the data. Increased submission of data to the repository will also likely help to enhance the features of OpenBeta since increasing traffic and data use from the site has the potential to increase demand for enhancement which can push for those improvements to occur faster.

References

“About: Openbeta Documentation.” *About | OpenBeta Documentation*, OpenBeta, https://docs.openbeta.io/.

Support@openbeta.io. “Why OpenBeta.” *Open Collaboration Climbing Platform*, OpenBeta, https://collective.openbeta.io/why.

OpenBeta. “Docs.openbeta.io/Docs/How-to-Contribute at Develop · OpenBeta/Docs.openbeta.io.” *GitHub*, OpenBeta, https://github.com/OpenBeta/docs.openbeta.io/tree/develop/docs/how-to-contribute.