**Intro**

LakeFS is an open-source data versioning tool designed to manage data lakes, allowing version control and management of large datasets. While LakeFS offers several benefits, such as simplified data versioning, there are several reasons why we have decided not to use it for the FAIR MAST project. This document outlines the primary reasons for our decision.

**1. Access Complexity**

One of the main drawbacks of LakeFS is that users must access data through LakeFS itself rather than directly from S3. This adds an extra layer of complexity for our users to access the data, whereas we would prefer to use direct links to S3 to retrieve the data.

**2. Vendor Lock-in**

By using LakeFS as our versioning control and management system, we put ourselves at risk of vendor lock-in. If LakeFS were to go out of business or discontinue support, the FAIR MAST project would put itself into a difficult position with our data tied to a system that is no longer maintained. Adding LakeFS also introduces another dependency into our data infrastructure. Each additional tool increases the risk of potential failures and adds complexity to the system.

**3. Additional Installation and Maintenance**

With LakeFS acting as our access layer, the users would be required to install and maintain their own instance of LakeFS which can take up time and introduce complexity. If we are trying to keep data access as simple as possible, adding this extra installation (and maintenance that comes with it) reduces that simplicity.

**6. Learning Curve**

Adopting LakeFS means that our users needs to learn how to use a new tool, which can be time-consuming.

While LakeFS offers valuable features for data versioning and management, the drawbacks as stated above, outweigh the benefits for our specific use case. Git Repo containing some test notebooks on using LakeFS found here: https://github.com/jameshod5/lakefs-trial