# DSP Concepts

## Contextualizing Sound Data Management

*"Great machine learning is not about algorithms, but how organized you are.* People are still learning how to build ML teams that are scalable. If you're putting 80% of your time into data, you better be retaining what you did. Otherwise you're burning money. You might as well just throw dollars into a trash can and light it on fire."

—**Josh Morris**, Machine Learning Engineering Manager at DSP Concepts

**DSP Concepts** is the global leader in embedded audio technology and creator of *Audio Weaver*, the audio development platform that makes audio innovation easy. DSP Concepts equips engineers with real-time workflows to quickly stand up prototypes, collaborate and modify designs across teams, and deploy to the most popular chipsets.

Quilt on AWS

DSP Concepts uses Quilt to organize their repository of machine learning datasets and models in Amazon S3. Quilt integrates data files, metadata, models, and documentation into versioned data packages in Amazon S3. Powered by Amazon OpenSearch Service, Quilt makes it easy for the DSP Concepts team to quickly find datasets based on metadata or content queries.

| Challenge | How Quilt Helps |
|---|---|
| **Poor data management** means new engineers take weeks or months to learn where everything is and how it works | Onboarding team members of DSPC to Quilt has been fast. It's really proven to us the value of having everything stored there, all the shared documentation, and having a nice API to retrieve stuff. With the guys on my team, we did probably an hour of training. I just sat them down, explained how we were organized, showed them how Quilt works, and three days later one had his first package in Quilt." |
| **Numerous, divergent copies of files and uncertain data lineage** make it impossible to trust and safely reuse models. | "We know we can use the data if there's no issues. Because not only do we know where the data is, we know the history associated with it. We have documentation. We know how it's changed over time. We can even jump back to a previous version if someone did something jacked up to the data." |
| **Location-based file management** does not guarantee the integrity or consistency of files | "I can go literally look at the commit messages, and then tie a [Quilt] hash to a model that's been trained. So I know at what point in history this model was trained and I know where the data set was at that point. Contrast that with a team where we just had an S3 prefix we all maintained, and it was a community effort. But you don't |

| | |
|---|---|
| | know if your coworker  decided that this data set was suboptimal and changed it for some reason.” |
| **No way to link git repositories** to the models and datasets associated with that source code | “We always maintain a parallel git repo that's dedicated to each package. So if we need to do any pre-processing or cleaning, we've got the code to do that. Or if we need to transform it into another package from some source package, we've got a repo associated with how that's created too. So not only are we getting data, but we're using hashes to track how our models change over time.” |

Quilt brings seamless collaboration to S3 by connecting people, pipelines, and machines using visual, verifiable, versioned data packages. Quilt Data is an Amazon Advanced Technology Partner. Amazon Web Services provides secure, cost-effective, and scalable big data services that can help you build a Data Lake to  collect, store, and analyze massive volumes of heterogeneous data.

**Visit [quiltdata.com](quiltdata.com) to learn more about seamless collaboration in Amazon S3.**

Quilt