Data Staging Platform
High-Level Architecture

Organizational data sources
- MSM schema-based, container model
- Distributed systems architecture
- Cloud | On-premises

Data Curation and Machine Learning

Phase 2: Automated data classification and governance
Data Staging Platform Dataflow

1. Utilizing ADF’s cloud, on premise capability of data ingestion.
2. Custom pipelines to orchestrate the data pipeline ingestion.
3a. Delta lake containing bronze, silver and gold copies of data.
3b. *Future state: Automated data classification and governance using Microsoft Purview classification service and data policies.*
4. Ingested emissions data, curated and updated for MSM ingestion and/or ML model training.
5. Azure Data Factory data orchestration for ML model training; creating delta tables which would be queried using synapse serverless pool.
6. Custom Power BI reports using alternative, industry report provider formats (i.e. GRI, SASB).
7. Customized MSM User Experience (UX), dynamic PowerApps UI extensibility for customer specific brand/campaigns.
Maureen Data Systems
Microsoft Sustainability Manager Data Staging Platform