

Climate Data Lake Project

Overview and Goals

The Massachusetts Data Office (MDO), a division of EOTSS, is launching a new project to create a **Commonwealth Climate Data Lake** in partnership with multiple agencies across the state.

The Climate Data Lake seeks to:

- Enable secure, centralized access to climate-relevant datasets from across the Commonwealth.
- Support data exploration, transformation, integration, and visualization by staff, researchers, and the public.
- Ensure appropriate public access through a mass.gov front-end while protecting privacy through data risk reviews.

This project will entail designing and building the infrastructure to host climate data, establishing classification and metadata standards, enabling governance workflows, and launching a public-facing site.

Outcomes

The project is complete when the following deliverables have been accepted by EOTSS and participating agencies:

- The centralized Climate Data Lake is operational in the Commonwealth's cloud environment.
 - Datasets from at least three agencies have been ingested, classified, and documented.
 - A public-facing mass.gov data portal is live, providing access to approved climate data.
 - Privacy risk assessments are conducted for all public-facing datasets.
 - Governance and access workflows are in place for internal staff and external researchers.
-

Approach

The MDO Project Manager will provide weekly status updates to key stakeholders via email.

Work will be organized into **biweekly sprints**, following Agile delivery practices.

Deliverables will be demoed in recurring check-ins. Feedback will be incorporated within the same sprint or used to inform the next sprint.

Close collaboration will be maintained with data owners at participating agencies to ensure proper classification, quality review, and onboarding of datasets.

Project Milestones and Timeline

Milestone	Description	Proposed Date
Milestone 1. Project approval	MDO receives go-ahead from EOTSS leadership and initial participating agencies	Week of 6/17
Milestone 2. Project kickoff	Kickoff meeting with stakeholders to align goals, approach, and team roles	Week of 6/24
Milestone 3. Initial agency datasets identified	Agencies deliver metadata and sample files for prioritized climate datasets	Week of 7/1
Milestone 4. Data classification and standards finalized	MDO and agencies align on data domains, tags, privacy tiers, and publishing guidelines	Week of 7/8
Milestone 5. First dataset ingested into Climate Data Lake	End-to-end ingestion of one sample dataset using agreed architecture	Week of 7/15
Milestone 6. Internal researcher access established	Internal permissions workflow built and tested for secure analyst access	Week of 7/22
Milestone 7. Public data review and risk mitigation	Privacy reviews completed on proposed public datasets; mitigation plans documented	Week of 7/29
Milestone 8. Mass.gov front end prototype complete	Mass.gov interface built to expose approved climate data to public users	Week of 8/5
Milestone 9. Stakeholder UAT	Participating agency staff and external researchers conduct user testing	Week of 8/12
Milestone 10. Launch & handoff	Launch of public portal, documentation delivery, and transition to operations	Week of 8/19