# Development of Spatial Data for California Wastewater Systems

ADVISORY GROUP SUMMARY – JANUARY 2026 MEETING

# SPATIAL DATA SUMMARY

This document outlines the activities necessary to identify the spatial location and boundaries of wastewater systems and facilities on the Wastewater Needs Assessment (WWNA) project's final Facilities List. Knowing the spatial location and boundaries of facilities and systems is essential to the Risk and Solutions analysis in the WWNA, but is also useful for broader purposes.

Given the lack of existing, full data repositories,[1] whether public or private, there is a need for both the collection and production of spatial datasets of the location of various wastewater facilities and system types analyzed in the Wastewater Needs Assessment (WWNA).

The key deliverables expected from the task are:

1. The UMASS-led modeling of unsewered areas and associated production of maps
2. The UCLA and OWP-led collection, coordination, and organization of sanitary sewer shapefiles (SSGO systems) and associated production of maps
3. The UCLA-led collection, coordination, and organization of wastewater facility (NPDES and WDR- permitted facilities) shapefiles and associated production of maps

The University of Massachusetts Amherst (UMASS) unsewered model[2] is described in greater detail in the Phase 1 report[3] and other Advisory Group meetings, and is not detailed here.[4] There are, however, three other types of facilities and systems for which we need to collect or generate spatial data: 1) National Pollutant Discharge Elimination System (NPDES) facilities, 2) Waste Discharge Requirements (WDR) facilities, and 3) Sanitary Sewer System General Order (SSSGO) systems.

This summary will focus on Tasks 2 and 3 listed above. The purpose of these tasks is both to be used in potential subsequent Water Board internal-use and public mapping, as well as to be used in subsequent tasks of the analysis in Phase 2. As such, these

---

[1] Tools such as the new USEPA sewershed projections (https://www.epa.gov/cwns/sewersheds) may be useful in the future, but at the moment they do not appear preferable in terms of accuracy and granularity to the data sources and methods we propose to user here

[2] See: https://innovation.luskin.ucla.edu/wp-content/uploads/2025/07/UMass-Unsewered-Model-Executive-Summary-ENGLISH.pdf

[3] See: https://www.waterboards.ca.gov/water_issues/programs/waste_discharge_requirements/docs/wastewater-needs-assessment-phase-1-report.pdf

[4] The UMASS model uses geospatial data and machine learning to predict, at the parcel level, whether properties require wastewater infrastructure and if they are unsewered or sewered. The UMASS model has been described in detail at previous Advisory Group meetings. See: https://youtu.be/iqQu_49HRho

tasks are largely a means to the end of other tasks in the WWNA, particularly the Risk and Solutions assessment tasks of Phase 2, and more broadly beyond the WWNA.

**Spatial Analysis Development**

Because of the lack of spatial data in this space, we must rely on a mix of spatial identifiers for systems which range in accuracy and granularity. As such, the collection, validation, and synchronization of spatial data for all systems analyzed remains an area for major improvement in future WWNA iterations, which we will detail further in the Final Report "Roadmap." We note that while California needs considerable additional effort beyond the WWNA to characterize its wastewater systems' spatial profile, this is true for all states we analyzed, except for Massachusetts.[5]

The ideal method to utilize the spatial data is to obtain the actual boundary service area of each system in the form of a "shapefile" polygon.[6] SSSGO systems are likely the only facility system type of the three which is likely to have a considerable proportion of shapefiles. The Water Board is requiring such systems to provide a Sanitary Sewer System Service Area Boundary Map by no later than December 31, 2025 (Order 2022-0103-DWQ). However, many representatives of the systems have voluntarily submitted a shapefile before the required deadline. With shapefiles provided, we can also spatially attribute boundaries to some NPDES and WDR facilities which work in tandem with SSSGO systems.

However, even with these shapefiles there are numerous questions and apparent irregularities with some of the boundaries, which we are analyzing further. In the absence of shapefiles, which will be the case even for many SSSGO systems, we will be relying on point data for each facility or system in the form of a valid address or latitude/longitude points to characterize the approximate location of the system. We will aim to characterize the centroid of the system wherever possible, but in our early analysis within the California Integrated Water Quality System (CIWQS) to geolocate these facilities, even obtaining a single, legitimate address or latitude/longitude point for some systems (especially WDR systems) may prove difficult.

Again, within the WWNA, spatial data is being collected for instrumental use in the Risk and Solutions assessments. As further discussed in companion Advisory Group Summaries for the January 2026 Advisory Group meeting, to apply risk variables such as socioeconomic or climate change vulnerability to facilities and systems we will use GIS analysis to construct system-level profiles of the population served by that facility or system. For example, we will use demographic and income data from the U.S. Census Bureau to estimate each system's served population by racial/ethnic group or identify those with high proportions of historically marginalized populations. These characteristics can be assigned at the system level or, with further analysis, distributed

[5] For instance, see https://www.mass.gov/info-details/massgis-data-massdep-estimated-sewer-system-service-area-boundaries

[6] See: https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm

to individual wastewater facilities or their defined service area boundaries to better reflect local demographics.

For solutions analysis, spatial data is also important in evaluating potential physical integration or connection and collaboration opportunities between inadequate facilities or systems which might benefit from connection to another well-functioning facility or system. This analysis will rely on measures of simple spatial proximity between systems in general, but also between key facilities or infrastructure components of interest, where possible.