

 On July 9, 2024, EPIC's Technology Program hosted a webinar with partners from across the Environmental Protection Agency (EPA) on the agency's new **Community Water System Service Area Boundary Dataset** (and accompanying **drinking water non-compliance metric** in EJScreen).

 This first-of-its-kind, comprehensive map of drinking water service area boundaries is the culmination of years of effort and sustained collaboration with EPA and others. [Our recent webinar](#) gave more than 300 advocates, state and federal agency staff, water data experts, and others the chance to celebrate and explore what this milestone means for drinking water justice efforts moving forward.

 Below, we've compiled instructive questions (*scroll down!*) that came up during the webinar, and answered them. We've also collated resources for easy reference:

Quick links + resources on this work:

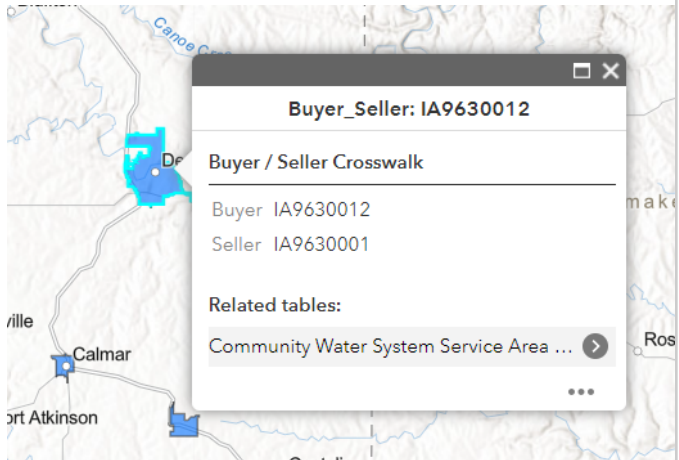
1. **Webinar Recording:** <https://www.youtube.com/watch?v=DcDUM1IpuJo&t=1467s>
2. **Webinar Slides:**  2024 07 EPA and EPIC Service Area Boundary Webinar Slides
3. **Overview of the data:**
<https://www.epa.gov/ground-water-and-drinking-water/community-water-system-service-area-boundaries>
4. **Use the data viewer** (the "i" button in the bottom bar) **to download the data:**
<https://epa.maps.arcgis.com/apps/webappviewer/index.html?id=5a9963ca7a594fbd9b0df40050ed704e>
5. **EPA's Methods + Data download:** https://github.com/USEPA/ORD_SAB_Model
6. **EJScreen Trainings + Office Hours:**
 - a. Training Date and Time: July 10th at 12pm EST. Link:
<https://usepa.zoomgov.com/j/1610600521>
 - b. Training Date and Time: July 24th at 12pm EST. Link:
<https://usepa.zoomgov.com/j/1610600521>
 - c. Office Hours Date and Time: August 21 at 12pm EST. Link:
<https://usepa.zoomgov.com/j/1610600521>



Q&A: About EPA's SABs Dataset, Modelling, etc.

Curious if the tool identifies people (or buildings like schools or other) on no "system"? i.e. private wells, or without even basic clean water?	I think the answer to that is a bit nuanced. Community water systems are defined as serving 15 connections or 25+ people year-round. If an area is outside of a boundary, it could still have water. For the most, residential areas outside of service boundaries are on private water. Other places like schools or gas stations or campgrounds may be served by transient or non-transient non-community systems, which are still regulated but do not have all of the testing requirements that community water systems have. We are planning to map these systems in the future.
Is this scoring different than the ETT scoring? Or is it the same score that informs the serious violator list?	It's a modified version of the ETT score. It counts the ETT violation points for unresolved violations, without the discount for formal enforcement, and without the extra points for the oldest unaddressed violation.
Does the dataset link service areas to water system ID?	Yes, the PWSID is in the data.
Do the models include privately owned water utilities?	Yes, when we say 'public' we mean that they are classified as community water systems and are subject to the Safe Drinking Water Act. They may be publicly or privately owned.
How do you estimate population (is it based on Census Blocks)? Are you able to distinguish between permanent residential population and seasonal population?	The water system provides the 'population served'. Regarding the second half, it's up to the utility/state on how they count the population served.
For the states with a large amount of public data, what did the small amount of modeling work consist of? Were there identified gaps or deficiencies in the available data?	Some states publish water districts in addition to service areas which we try to avoid. Some state boundaries were adjusted to remove overlapping areas or to correct 'nested' systems.
Does the dataset report if a given location is getting state supplied or modeled data?	Yep! You are able to see that in the metadata.
One of the challenges that we (SimpleLab) encountered in the modeled boundaries was not just estimating the size of the boundary, but figuring out the location. The lat/lons in SDWIS are often not accurate; often reflecting an administrative building rather than the actual service location. Did your model have a way of improving on this?	There are several other variables in the model that help to account for SDWIS locations that may be far away. We use fuzzy text matching across several geographies and system information, building density, population density and parcel data to name some of them. We also perform some transformations such as finding the mean center for systems with multiple intakes / treatment plants
It seems like the model is largely informed by data from coastal states. How much does model accuracy vary by geography?	There is some variation, but we are confident it performs well across geographies. Although only 3 states, those 3 represent about 10% of blocks in the country and have good samples of both rural and urban areas.



Did you find that the state metadata consistently provided small system coverage as comprehensively as larger system coverage? If not, how did you account for the difference?	Yes, state sourced data was more comprehensive with regards to larger systems than smaller systems. Mobile home parks or small home owners associations, for example, typically don't have geospatial boundaries delineated. If the state didn't have these boundaries delineated we modeled them. Within the service area boundary attribute table you can see what the source of the data is and associated metadata like date created.
How are service areas verified? What is the percentage of verified status vs unverified status service areas in the dataset? Thank you!	Approximately 40% of the data were sourced from states, and the remaining 60% with no state-supplied boundaries were modeled using machine learning.
Does the model distinguish between wholesale and retail systems where boundaries might overlap?	If a system is just a wholesaler, it is not mapped. Within the popup window (pictured below) you can see the buyer/seller relationship by selecting the "buy/seller crosswalk" relate table: 
Does detection of "non-compliance " cover water safety at people's taps, rather than at the water supply sources? --Just thinking about lead or other contaminants/toxics picked up by water when it sits in and or is locally distributed through lead piping or other combinations of piping/welds that can react with water at point of delivery? And how to deal with this?	Most rules under the Safe Drinking Water Act enforce compliance at the water system, not in people's homes. An exception is the Lead and Copper Rule , which requires utilities to sample for lead and copper at taps in homes and workplaces.
I wonder if there is any such thing as a large "private" water system, like an "investor owned water utility" -similar to what happened in the development of IOUs in the energy systems sector? Are there small "private " water systems that are excluded from this study?	Any drinking water system that has at least 15 service connections or serves at least 25 people at least 60 days per year, is regulated as a public water system (PWS) under the Safe Drinking Water Act. PWS may be publicly or privately owned.



	Systems that don't meet the definition of a PWS, such as residential private wells, aren't regulated under SDWA and aren't part of EPA's service area boundary data set. The service area boundary data set is so far only for community water systems, which are PWS that serve residential communities.
--	---

Q&A: Analyses on the data

Have you all conducted any preliminary demographic analysis on which communities ARE and ARE NOT served by public water systems?	We have done some very preliminary analyses but nothing to report with confidence. It's something we are definitely looking to dive into soon.
In aggregating these data and modeling these boundaries, did you find that there was a lot of temporal variation in system service area boundaries? or did these tend to be generally pretty static?	Anecdotally, we have heard the boundaries are relatively static and change minimally, if at all, on an annual basis. As new information becomes available about the water system boundaries, it will be updated in the national dataset.
Are there plans to perform more rigorous validation and analysis of the errors the models have? Any plans for quantitative measurement of the confidence by region or water system type?	Our technical guide provides some more quantitative analysis/validation: https://epa.maps.arcgis.com/home/item.html?id=609326a2d1024f2da42ac53f68f40db6
This is great! For the construction of the non-compliance index, do you have documentation/the code that was used to go from SDWIS data to its computation?	You can find more about the non-compliance index here: https://www.epa.gov/ejscreen/ejscreen-indicators-overview-drinking-water-non-compliance
As previously stated, there are several data sets/maps. What are your thoughts/suggestions for selecting between them? For example, CA has a comprehensive statewide, accurate map and a SAFER program that rates and tracks water systems based on various metrics of risk. What should practitioners think about when deciding to use CA dataset vs EPA datasets?	There are two questions in here: 1. The service area boundaries dataset: the national dataset uses state-data when available. As such, the underlying data on who serves water to who should be identical and you can use the CA data or the EPA data, whatever is easiest for you. 2. The drinking water justice metric: The data on drinking water non-compliance metric from EPA does not include the range of data available at a state-level. If you are working in a state like CA with robust metrics - and a strong history of tracking and reporting on this data - defer to the state information.
Are there plans to retrospectively estimate who was drinking what water 10 years ago? 20 years ago?	Not at present from EPA, but this is a great project for an organization to take on.



Have you all conducted any preliminary demographic analysis on which communities ARE and ARE NOT served by public water systems?	We have not, but some colleagues in the Office of Policy are writing a paper on this topic currently. They are using our boundaries to do this demo analysis. It'll probably be a few months until it's published. Wes Austin is leading this effort—austin.wes@epa.gov Some other papers that discuss this issue: McDonald, Yolanda J, and Nicole E Jones. 2018. "Drinking Water Violations and Environmental Justice in the United States, 2011–2015." <i>American Journal of Public Health</i> 108 (10): 1401–7. Mueller, J Tom, and Stephen Gasteyer. 2021. "The Widespread and Unjust Drinking Water and Clean Water Crisis in the United States." <i>Nature Communications</i> 12 (1): 3544. Bae, Junghwan, and Michael J Lynch. 2023. "Ethnicity, Poverty, Race, and the Unequal Distribution of US Safe Drinking Water Act Violations, 2016-2018." <i>The Sociological Quarterly</i> 64 (2): 274–95.
Based on the violation trends it seems the larger the water system, the safer/better. I have heard of some utilities being absorbed into others. Is there any anticipated policy or advocacy in that realm - absorbing smaller utilities to better serve users and ultimately help data purposes such as this as well?	Let me point you to this: The Water System Restructuring Assessment Rule US EPA There are a few public webinars coming up in July (July 17 and July 24). Public comment period is open through July 29, 2024.
Is there some way to identify percentage of lead/copper lines by service territory?	This is out of scope for the service area boundaries data project, but many folks are working on lead/copper identification and there are resources for how to do so here.
Does this dataset contain information related to PFAS pollution?	Not yet. EPA has issued a final rule to regulate PFAS in drinking water. Maximum contaminant limits will take effect in 2029, and violations will then be counted in the drinking water non-compliance indicator.



Is the Drinking Water Non-Compliance dataset itself also available for download, or even more practical, available as a hosted layer on e.g. ESRI's living atlas? Thank you for this webinar!	Yes! You can download with the "I" button in the bottom bar you can download the data: https://epa.maps.arcgis.com/apps/webappviewer/index.html?id=5a9963ca7a594fbd9b0df40050ed704e Or from GitHub: https://github.com/USEPA/ORD_SAB_Model/tree/main/Output_Data
---	---

Q&A: Expansions from here

Is this data now overlayed in the CJEST tool?	Not yet! Just EJScreen for now but definitely worth exploring how to get within CEJST as well.
Are there any plans to incorporate any non-community public water systems in this dataset?	Yes definitely, especially if it's clear that is useful to folks!
Love to see the shoutout to Mobile Home Park CWS, as they come up over and over again in struggling small systems conversations. Is there any potential coordination with HUD (beyond CBDG funding) on mapping/solutions for these systems, since HUD/state housing agencies have broader oversight on MHP living conditions?	We think there is definitely potential here. Over half of community water systems serve < 500 people and these systems lack the treatment and testing capabilities that larger cities like Chicago or New York have. As we continue to develop this work into the future we hope to get even closer to 100% of all systems
If a state or water utility representative would look at their service area boundary on this dataset and would want to improve its accuracy, what could they do?	Please reach out to OW OGWDW DWIDD DWDA@epa.gov with the updated data and it will be fixed in the dataset.
How likely are we to see a similar wastewater map in the near future?	Stay tuned! If this is something you are keen to work on, please reach out to Jessie; jessie@policyinnovation.org
Congrats EPA team, this is really great. What's next for this team in terms of project development / project maintenance?	As new or improved service area boundary data are provided to EPA, this service area boundary dataset will be updated. EPA is currently exploring options to map non-community water systems as well.