

Apéndice F del Plan de Monitoreo del Aire de la Comunidad

Iniciativa de Monitoreo Móvil Estatal de California (SMMI)
Aclima's Plan de gestión de datos (v3.0)



16 de mayo de 2025



La Iniciativa de Monitoreo Móvil Estatal es parte de Inversiones Climáticas de California, una iniciativa estatal que utiliza miles de millones de dólares de Tope y Comercio para reducir las emisiones de gases de efecto invernadero, fortalecer la economía y mejorar la salud pública y el medio ambiente, particularmente en comunidades desfavorecidas.

Contenido

Contenido	2
Glosario de abreviaturas	3
Acerca de Aclima	5
Acerca de la Iniciativa Estatal de Monitoreo Móvil	5
1. Requisitos y entregables	6
1.1. Resultados del plan de gestión de datos.....	6
1.2. Elementos relevantes del SOW.....	6
2. Conceptos clave	7
2.1. Niveles de datos de Aclima.....	7
2.2. Canal de gestión de datos.....	9
3. Ingesta de datos	10
3.1. Definición y aplicación del Nivel de Datos 0.....	11
3.2. Flujos de trabajo de ingesta de datos.....	11
3.2.1. Publicar.....	11
3.2.2. Ingesta (datos del nodo móvil de Aclima).....	11
3.2.3. Ingesta (datos PML).....	12
4. Transformación de datos	13
4.1. Definición y aplicación de los niveles de datos 1, 2a y 2b.....	13
Nivel de datos 1.....	13
Nivel de datos 2a.....	13
Nivel de datos 2b.....	14
4.2. Flujos de trabajo de transformación de datos.....	14
4.2.1. Procesamiento de datos de nivel 1.....	14
4.2.2. Procesamiento de datos de nivel 2a.....	14
4.2.3. Cambio de hora.....	14
4.2.4. Marcado de datos.....	15
5. Modelado de datos	17
5.1. Definición y aplicación de los niveles de datos 3-4.....	17
Nivel de datos 3.....	17
Nivel de datos 4.....	17
5.2. Flujos de trabajo de modelado de datos.....	18
6. Almacenamiento de datos	20
7. Revisión de datos y garantía de calidad	21
7.1. Descripción general.....	21
7.2. Revisión de la implementación activa.....	22
7.2.1. Flujo de trabajo de revisión de la implementación activa.....	22
7.2.2. Alertas personalizadas.....	24
7.3. Revisión posterior a la implementación.....	24

7.3.1. Flujo de trabajo de revisión posterior a la implementación.....	25
7.4. Revisión posterior a todas las implementaciones.....	27
7.4.1. Flujo de trabajo de revisión posterior a todas las implementaciones.....	28
7.5. Revisiones de datos.....	29
8. Transferencia de datos.....	30
8.1. Esquema de datos de nivel 2a.....	30
8.1.1. Descripción general del esquema.....	30
8.1.2. categorías de esquema y utilidad.....	32
8.1.3. Tabla de búsqueda de códigos AQS.....	32
8.1.4. Rellenar la categoría de método.....	33
8.1.5. Nota general sobre los códigos AQS.....	33
8.2. Transferencia de datos de nivel 2a.....	33
8.2.1. Alcance y objetivos de la transferencia de datos finalizada.....	33
8.2.2. Disponibilidad de los datos finalizados a través de Google Cloud Storage.....	33
8.2.3. Formato de archivos de datos.....	34
8.2.4. Frecuencia de entrega y tamaño estimado de archivo.....	35
8.2.5. Obligación de alimentos.....	36
8.3. Vías de transferencia propuestas.....	36
8.3.1. Ruta preferida: Integración de Snowflake.....	37
8.3.2. Ruta alternativa: Interfaz de usuario y funciones de las herramientas de la consola de Google Cloud.....	38
8.3.3. Seguridad de la vía de transferencia.....	39
9. Conjuntos de datos y visualizaciones de StoryMap.....	40
9.1. Conjuntos de datos para ayudar a identificar fuentes.....	40
9.1.1. Conjuntos de datos basados en mejoras.....	40
9.1.2. Conjuntos de datos basados en la concentración ambiental.....	41
9.2. Conjuntos de datos para ayudar a identificar ubicaciones con un impacto desproporcionado.....	41
9.3. Enfoque de visualización de conjuntos de datos.....	45
9.3.1. Plantilla común de StoryMap.....	46
9.3.2. Capa de datos basada en mejoras, derivada de los sensores Aclima.....	47
9.3.3. Capa de datos basada en mejoras, derivada de instrumentos PML.....	47
9.4. Suministro de datos y transferencia de propiedad de StoryMaps.....	49

Glosario de abreviaturas

ADR- Ticket de revisión de datos de AMN

AMN- Nodo móvil de Aclima

AMPERIO- Plataforma móvil Aclima

API- Interfaz de programación de aplicaciones

ACAMPAR- Plan de Monitoreo del Aire de la Comunidad

CERP- Plan de Reducción de Emisiones Comunitarias

CNC- Comunidad nominada constantemente

demócrata- Modelo digital de elevación

DEP- Ticket de implementación

DQO- Operaciones de calidad de datos

Póngase en fila- Óxido de etileno

GCP- Plataforma de Google Cloud

GCS- Almacenamiento en la nube de Google

SOY- Gestión de identidad y acceso

IoT- Internet de las cosas

L1- Datos de monitoreo de la calidad del aire móvil de nivel 1

L2- Datos de monitoreo de la calidad del aire móvil de nivel 2

L3- Datos de monitoreo de la calidad del aire móvil de nivel 3

L4- Datos de monitoreo de la calidad del aire móvil de nivel 4

CLAVIJA- Grupo de expertos del proyecto

PML- Laboratorio Móvil Socio

SAPP- Plan de Muestreo, Análisis y Presentación

SMMI- Iniciativa de Monitoreo Móvil Estatal

SEMBRAR- Alcance del trabajo

TVOC- Compuestos orgánicos volátiles totales

VER- Ticket de verificación de implementación

Acerca de Aclima

Aclima es una Corporación de Beneficio Público y una empresa de tecnología climática que impulsa acciones audaces para reducir las emisiones, proteger la salud pública y garantizar un aire limpio para todos. Gracias a su red de sensores móviles, Aclima mide la contaminación atmosférica y los gases de efecto invernadero en ciudades, regiones y estados con una resolución a nivel de manzana. Las herramientas en línea de la empresa traducen miles de millones de mediciones científicas en mapas interactivos y análisis diseñados para fundamentar y potenciar la toma de decisiones sobre aire limpio, con un enfoque en herramientas intuitivas que apoyan a las comunidades, especialmente a las más afectadas por la contaminación atmosférica. Entre los clientes de Aclima se incluyen organismos reguladores, empresas de servicios públicos, empresas y comunidades que trabajan para reducir las emisiones y proteger la salud pública y el medio ambiente.

Acerca de la Iniciativa Estatal de Monitoreo Móvil

La Iniciativa Estatal de Monitoreo Móvil (SMMI) está diseñada para obtener un conjunto completo de datos sobre contaminantes criterio, contaminantes atmosféricos tóxicos y gases de efecto invernadero, crear un portal de datos para que el público acceda y visualice los datos de la SMMI, y promover la participación comunitaria inclusiva para comprender mejor y abordar las preocupaciones de la comunidad. Este proyecto brinda la oportunidad de complementar las actividades estatales de monitoreo del aire de la AB 617, involucrando a comunidades más allá de las actualmente seleccionadas bajo el Programa de Protección del Aire Comunitario, proporcionando datos para subsanar las deficiencias en el monitoreo del aire y respaldar medidas adicionales para reducir las emisiones y la exposición.

1. Requisitos y entregables

1.1. Resultados del plan de gestión de datos

El propósito de esta documentación es proporcionar una descripción general de alto nivel del sistema de gestión de datos patentado de Aclima y sus capacidades, y explicar cómo se aplica esto para proporcionar datos brutos finalizados de alta calidad y productos de datos derivados.

El sistema de gestión de datos de Aclima recopilará y procesará datos sin procesar a través de múltiples niveles de procesamiento, almacenamiento y etiquetado durante la campaña de recolección de SMMI, incluyendo datos de los Laboratorios Móviles Asociados (PML). Posteriormente, el sistema transferirá los datos sin procesar finalizados a CARB mediante un contenedor de datos en la nube y los proporcionará directamente para facilitar las visualizaciones y los informes mediante StoryMaps.

1.2. Elementos relevantes del SOW

Esta documentación cumple directa o parcialmente con los siguientes elementos del Alcance de Trabajo (SOW) de SMMI y las subtareas de Aclima:

- Elemento 6.1 del SOW de CARB: Establecer un procedimiento de gestión de datos
 - Subtarea 6.1.1 de Aclima: Borrador del plan de gestión de datos desarrollado y documentado
 - Subtarea 6.1.2 de Aclima: Iteración del Plan de Gestión de Datos a través del proceso CAMP, con foco en el análisis de datos y la comunicación de resultados para apoyar la acción.
 - Subtarea 6.1.3 de Aclima: Aprobación final del plan de gestión de datos
- Elemento 6.2 del CARB SOW: Establecer mecanismos de transferencia de datos
 - Subtarea 6.2.1 de Aclima: Transferencia de datos a CARB: recopilación de requisitos de interfaz
 - Subtarea 6.2.2 de Aclima: Transferencia de datos a CARB: documento de requisitos de interfaz finalizado
 - Subtarea 6.2.3 de Aclima: Transferencia de datos a CARB: recopilación de requisitos del esquema de datos

2. Conceptos clave

2.1. Niveles de datos de Aclima

Las categorías de datos que Aclima pondrá a disposición de CARB y las comunidades abarcarán mediciones de 1 Hz para análisis y desarrollo, agregaciones de datos tomados durante el período de observación y alertas de baja latencia para la detección de señales de alta concentración.

Aclima organiza estos datos en niveles que reflejan el grado de procesamiento aplicado, desde el nivel más bajo (Nivel 0 o L0) en la lectura del sensor hasta los análisis modelados de alto nivel (Nivel 4 o L4), que sintetizan los puntos de datos individuales en información práctica y resúmenes de datos para su difusión mediante visualización e informes. Ver [Tabla 2.1A](#) continuación se describen todos los niveles de datos. Esta clasificación se basó en [investigación publicada](#), y se ha adaptado para describir los niveles de procesamiento de datos móviles de Aclima. Los niveles de datos L0-L2 también son datos de series temporales continuas. L3 y L4 se diferencian de los niveles de datos L0-L2 en que son conjuntos de datos más agregados (ya sea espacial o temporalmente) o se extraen de los datos de series temporales originales, por ejemplo, un clúster geoespacial de detecciones de mejora. El propósito de definir estos niveles de datos es comunicar los pasos generales de procesamiento de datos asociados con los datos y productos de datos entregados a CARB y otros usuarios finales de datos y visualizaciones SMMI, y ayudar a informar a CARB y a los usuarios finales sobre el tipo y la naturaleza de los datos proporcionados.

Tabla 2.1 Niveles de procesamiento de datos de Aclima. Los asteriscos (*) indican los niveles de datos proporcionados a CARB o en apoyo de la comunicación no científica y la visualización comunitaria.

Nivel de datos	Nombre	Definición	Ejemplo
0	Señal sin procesar	Señal original producida por el sensor.	Voltaje, número digital, datos espectrales de masa sin procesar.
1*	Magnitudes geofísicas intermedias	Derivado de datos de nivel 0 utilizando principios físicos básicos o ecuaciones de calibración.	Concentración en ppb o $\mu\text{g}/\text{m}^3$.
2a*	Magnitudes geofísicas estándar	Estimación utilizando el sensor más las	NO_2 derivado de O_3 y O_x ($\text{O}_3 + \text{NO}_2$).

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

		<p>mediciones físicas a bordo asociadas.</p> <p>También puede incluir pasos simples de procesamiento de señales, como suavizado o filtrado de frecuencia para corregir los artefactos del sensor.</p>	<p>Correcciones de temperatura y humedad a las estimaciones de sensores utilizando relaciones derivadas empíricamente en lugar de principios físicos básicos.</p> <p>Mediciones corregidas para la deriva del sensor utilizando un método de procesamiento de señal de filtro de paso de banda alta.</p>
2b	Magnitudes geofísicas estándar, ampliadas	<p>Similar al nivel 2a pero utilizando fuentes de datos externas para la corrección de artefactos. No se prevé su uso para SMMI.</p>	<p>Cantidades normalizadas utilizando datos meteorológicos o de contaminantes cubricados.</p>
3*	Cantidades geoespaciales agregadas	<p>Productos de datos geoespaciales derivados típicamente de datos L1, L2a o L2b de mediciones de un solo contaminante o una combinación de múltiples contaminantes y utilizando enfoques de agregación estadística estándar de mediciones (es decir, media, mediana, máximo, etc.)</p>	<p>Mapa de ubicaciones de eventos de mejora individuales de TVOC, tóxicos especiados u otros contaminantes.</p> <p>Grupos de mejoras de metano, en combinación con etano, que indican ubicaciones de fugas de gas natural.</p> <p>Mapas continuos de concentraciones ambientales que utilizan únicamente agregación estadística básica de mediciones, sin aplicar métodos de modelado avanzados.</p>
4*	Cantidades geoespaciales agregadas combinadas con	<p>Productos de datos geoespaciales agregados derivados típicamente de</p>	<p>Mapa de promedios de contaminantes derivados de datos de medición en</p>

	fenomenología espacio-temporal modelada	datos L1, L2a o L2b utilizando métodos avanzados de modelado estadístico en el espacio y el tiempo. Se podrán incorporar fuentes de datos externas o información contextual como mediciones meteorológicas, datos topográficos, tipo de carretera, etc.	combinación con un método de reconstrucción estadística o un método de regresión del uso de la tierra Mapa continuo de estimaciones probabilísticas de la probabilidad de una fuga de metano en un área Mapa que muestra puntos críticos estadísticamente significativos de concentraciones promedio de contaminantes mediante el método Getis-Ord (Gi*)
--	---	--	--

2.2. Canal de gestión de datos

Los pasos clave en la gestión de datos de Aclima se resumen en [Tabla 2.2A](#) continuación, y representado visualmente en [Figura 2.1A](#) continuación, junto con los módulos técnicos que habilitan el pipeline. Estos se describen en las siguientes secciones.

Tabla 2.2: Una descripción general de los principales componentes de la gestión de datos desde el dispositivo hasta la transferencia de datos sin procesar finalizados a visualizaciones alimentadas por API.

Canalización de gestión de datos		
1	Publicar	Publique datos de sensores de 1 Hz junto con metadatos, incluidos números de serie de dispositivos y mediciones auxiliares como temperatura específica del sensor, humedad relativa, presión, latitud y longitud (GPS) y caudal desde dispositivos remotos a la nube.
2	Ingerir	Ingesta de datos en la nube. Los datos de nivel 0 del archivo en la nube nunca se eliminan ni se modifican. Código personalizado que se ejecuta en servicios de computación en la nube para extraer datos de terceros para capas de datos contextuales según sea necesario.
3	Transformar	Procesar datos en mediciones geolocalizadas con significado físico a lo largo de la ruta de L0→L1→L2.

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

		Marcar datos no válidos provenientes de fallas de instrumentación, períodos de calentamiento, lecturas fuera de rango, etc. a través de actividades de control de calidad (QA/QC).
4	Modelo	Transformaciones de datos para crear productos de datos de nivel superior a partir de datos L1/L2, incluidas agregaciones espaciales y temporales y análisis de datos de múltiples contaminantes.
5	Almacenar	<p>Etiquete y almacene todos los niveles de datos procesados en un lago de datos. Haga que los datos sean accesibles para su posterior procesamiento, análisis y entrega directa a los clientes.</p> <p>Conserve instantáneas de datos en etapas de procesamiento críticas, como por ejemplo al finalizar la calibración.</p>

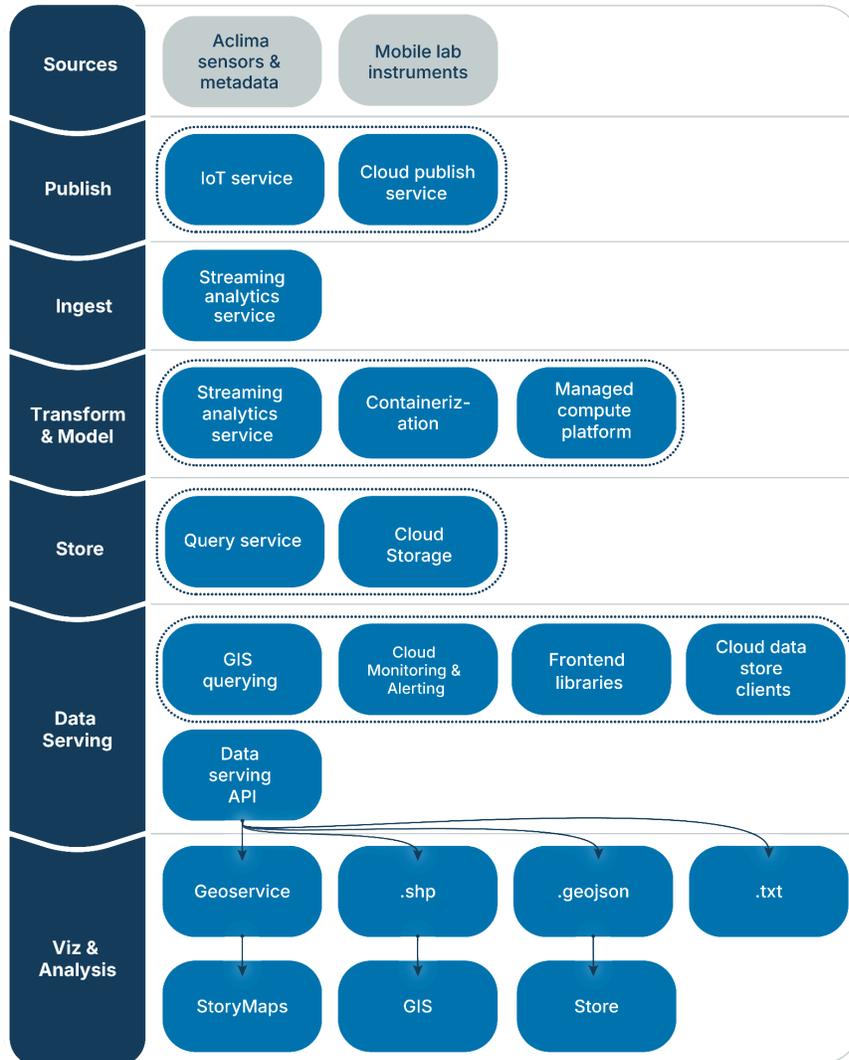


Figura 2.1: El flujo principal de información desde la publicación de datos de sensores e instrumentación de 1 Hz en el punto de medición, pasando por el procesamiento, el almacenamiento, las alertas y la distribución de datos y productos de análisis, hasta el resumen de los hallazgos para la comunidad.

3. Ingesta de datos

Ingestión de datos describe el proceso mediante el cual los datos inalterados de nivel más bajo de los sensores del nodo móvil Aclima (AMN) se envían al backend de almacenamiento y procesamiento de datos de Aclima.

3.1. Definición y aplicación del Nivel de Datos 0

Nivel de datos 0 (L0) describe la señal bruta inalterada producida por un sensor. Dependiendo del sensor, puede ser un voltaje, un número digital, datos espectrales de

masas sin procesar u otro formato. Un ejemplo de datos L0 son los conteos sin procesar de un sensor electroquímico. Los datos L0 generalmente no son útiles para el usuario final y no se incluirán en el conjunto de datos entregado para SMMI, sino que se conservarán indefinidamente en su estado original.

Tenga en cuenta que, en algunos casos, cuando el sensor cuenta con su propio procesamiento de datos interno, es posible que no se pueda acceder a los datos L0, y los datos de nivel más bajo disponibles para ciertos flujos de medición pueden ser L1 o L2. Por ejemplo, para un contador óptico de partículas, la señal bruta se considera en conteos por unidad de tiempo (p. ej., conteos/s), pero el procesamiento de datos interno de este sensor transforma los datos en conteos por unidad de volumen (p. ej., conteos/L), que se considera L1 en el marco utilizado aquí. En todos los casos, los datos de nivel más bajo disponibles son los que publica la AMN, se incorporan a la nube y se conservan indefinidamente en su estado original. A continuación, se describen estos pasos con más detalle.

3.2. Flujos de trabajo de ingesta de datos

3.2.1. Publicar

La AMN sondea los sensores físicos y los equipos de referencia de laboratorio para obtener datos cada segundo. Las mediciones sin procesar de los sensores móviles se registran junto con la fecha, hora y ubicación de cada medición, así como otros parámetros de diagnóstico críticos, como la temperatura, la humedad relativa, la presión, la latitud y longitud (GPS) y el caudal específicos del sensor. Todas las lecturas sin procesar de los sensores, la ubicación y los datos de diagnóstico se empaquetan en mensajes estructurados con una marca de tiempo de red asociada y se transmiten al backend en la nube de Aclima a través de un módulo LTE interno, utilizando un protocolo de transporte ligero del Internet de las Cosas (IoT) ampliamente adoptado. Si no hay conexión de red disponible en el momento de la recopilación de datos, estos se almacenan localmente y se publican posteriormente una vez que se dispone de conexión. Hay suficiente almacenamiento local disponible para hasta 60 días de recopilación de datos.

3.2.2. Ingesta (datos del nodo móvil de Aclima)

Los paquetes de datos de los sensores en el nodo móvil se envían a la nube a través de una puerta de enlace y luego se reenvían a un servicio de mensajería que separa la publicación de mensajes del consumo. Los mensajes se procesan mediante canales de datos que se ejecutan como trabajos en la nube. Los datos se almacenan en múltiples archivos en la nube para garantizar la conservación a largo plazo de los datos de nivel más bajo.

Se incorporan conjuntos de datos adicionales de terceros para respaldar el procesamiento de datos operativos y de productos, así como las operaciones de control de calidad (QA/QC). Estos pueden incluir datos meteorológicos modelados, modelos digitales de elevación (DEM) y datos estacionarios de monitoreo de la calidad del aire de diversas fuentes.

3.2.3. Ingesta (datos PML)

Los laboratorios móviles asociados (PML) gestionan la ingesta de datos por sí mismos, según sus procedimientos operativos estándar internos y dadas sus capacidades únicas de hardware y software.

Aclima no incorporará datos PML directamente en la canalización estándar. Los datos PML se recibirán después del procesamiento y se gestionarán mediante la línea de comandos o interfaces de arrastrar y soltar proporcionadas por el proveedor de nube de Aclima. Cada equipo PML contará con un contenedor de datos dedicado en la nube y las cuentas de lectura y escritura adecuadas para transferir datos a estos contenedores.

Los datos de PML se formatearán con el mismo esquema de posprocesamiento que los datos de AMN, y los indicadores de control de calidad y los códigos de error se rastrearán en el mismo formato. Aclima transmitirá estos datos a CARB sin modificaciones a través del mismo contenedor de datos en la nube que los datos de AMN (véase la [sección 8.2](#) la categoría "ID del vehículo" de los datos puede utilizarse para identificar los datos de PML, así como la categoría "Organización" incluido en los nombres de los archivos de datos. Los equipos de PML seguirán las mismas prácticas recomendadas que Aclima para conservar los datos L0 (o los datos de nivel más bajo disponibles) indefinidamente y sin modificaciones.

4. Transformación de datos

Transformación de datos describe cómo el nivel de datos 0, ahora ingerido y almacenado en el backend de Aclima, se transforma en los niveles de datos 1, 2a y 2b.

4.1. Definición y aplicación de los niveles de datos 1, 2a y 2b

Nivel de datos 1

Nivel de datos 1(L1) se deriva de los datos de Nivel 0 (señal bruta del sensor) utilizando principios físicos básicos o ecuaciones de calibración. Mientras que los datos L0 pueden ser datos de voltaje o datos espectrales de masas brutos, los datos L1 son magnitudes geofísicas intermedias, como una concentración en ppb o $\mu\text{g}/\text{m}^3$. Un ejemplo de datos L1 es la concentración en unidades de ppb que se ha derivado de la señal bruta como corriente (en amperios) de un sensor electroquímico y se ha convertido a concentración según una sensibilidad calibrada (en amperios por ppb) y una corriente de compensación (en amperios), pero antes de cualquier corrección empírica por humedad o temperatura.

Nivel de datos 2a

Nivel de datos 2a(L2a; también denominado "sin procesar" y "finalizado" en otra documentación del proyecto) es generalmente el formato final de los datos para la entrega de datos SMMI, aunque en algunos casos puede incluirse la entrega de datos L1 (véanse los ejemplos a continuación). Aclima define los datos L2a como señales de sensor transformadas en magnitudes geofísicas de medición, estimadas utilizando la señal del sensor más las mediciones físicas asociadas directamente relacionadas con el principio de medición, como las mediciones de temperatura y humedad relativa en la apertura del sensor para la corrección de interferencias, incluyendo los artefactos marcados y con los parámetros de calibración finales aplicados. También se puede incluir procesamiento adicional de señales en las transformaciones para producir datos L2a, como el suavizado o el filtrado de frecuencia.

Algunos ejemplos de datos L2a incluyen:

- $\text{P.M}_{2.5}$ derivado de recuentos de partículas resueltas por tamaño utilizando una densidad de partículas asumida
- NO_2 derivado como una diferencia entre O_3 y O_x ($\text{O}_3 + \text{NO}_2$) medido desde dos sensores en el mismo AMN.
- Correcciones de temperatura y humedad a las estimaciones de los sensores
- Suavizado de datos de carbono negro mediante reducción de ruido optimizada (ONA) u otro método

Dependiendo del flujo de medición, se podrían enviar datos L1 o L2a (o ambos) a CARB. Por ejemplo, el recuento de partículas por unidad de volumen, con resolución de tamaño, se considera L1 y se planea enviarlo además de $PM_{2.5}$ (categorizado como L2a) porque ambas mediciones pueden ser útiles para la categorización de las fuentes de emisión. Los datos de cierta instrumentación de investigación operada por los PML pueden considerarse L1 si utilizan únicamente los principios físicos básicos y un factor de calibración para traducir la cantidad medida (p. ej., la cantidad de luz absorbida mediante la Ley de Beer a una longitud de onda específica) a unidades físicas de concentración.

Nivel de datos 2b

El nivel de datos 2b (L2b) corresponde a datos L2a, pero utiliza fuentes de datos externas para la normalización o corrección de artefactos, como datos de monitoreo estacionario o datos meteorológicos. Este nivel de datos se incluye aquí para mayor exhaustividad, pero actualmente no se propone su uso en ningún resultado derivado de SMML.

4.2. Flujos de trabajo de transformación de datos

4.2.1. Procesamiento de datos de nivel 1

Las señales sin procesar se transforman en magnitudes físicas calibradas (datos L1) que representan las concentraciones de contaminantes, utilizando principios físicos básicos y parámetros de calibración específicos del sensor. Como parte de esta transformación, se suele consultar la base de datos de metadatos de Aclima para vincular los parámetros de calibración con el ID de referencia único del sensor que informa. El método de transformación de la señal sin procesar del sensor a magnitudes físicas calibradas depende del tipo de sensor y del principio de detección. Para ciertos flujos de medición, esta transformación se realiza internamente en el dispositivo mediante métodos especificados por el fabricante y no está controlada por el proceso de transformación de datos de Aclima.

4.2.2. Procesamiento de datos de nivel 2a

Los datos de Nivel 1 se transforman en datos de Nivel 2a mediante diversos métodos, según el método de detección y el flujo de medición. Estos métodos pueden ir desde simples diferencias entre dos mediciones simultáneas (p. ej., NO_2 de O_3 y O_x), combinación de múltiples contaminantes en una sola cantidad (p. ej. $PM_{2.5}$), correcciones de temperatura y humedad y/o métodos de procesamiento de señales.

4.2.3. Cambio de hora

Un paso adicional de transformación que se realiza en los datos L1 y L2a antes de su uso para generar productos de datos derivados de mayor nivel y antes de su entrega a CARB

es la aplicación de un desfase temporal para tener en cuenta el tiempo finito que transcurre entre el muestreo de aire en la entrada de las líneas de muestreo fuera del vehículo y su detección dentro del AMN. Aclima utiliza un desfase temporal fijo en toda la flota para cada sensor, basado en el caudal volumétrico y la longitud y el diámetro interior de las líneas de muestreo asociadas a dicho sensor. Los PML utilizan un enfoque similar, adaptado a cada instrumento y a la disposición de los laboratorios móviles. Esta transformación es importante para alinear los conjuntos de datos de múltiples contaminantes con la ubicación real del muestreo, en lugar de la ubicación de detección.

4.2.4 Marcado de datos

Como parte del proceso de transformación de datos, un sistema automatizado marca los puntos de datos individuales como no válidos según umbrales predeterminados de valores de contaminantes y mediciones de diagnóstico auxiliares (p. ej., temperatura, caudal, presión y humedad). Este marcado automatizado detecta algunos casos comunes de producción de datos no válidos, pero no todos los casos posibles. Por ello, existe un sistema manual para identificar y marcar cualquier posible punto de datos integrado en el proceso de transformación. Se incluye una descripción detallada de este proceso, conocido como revisión de datos, en [Sección 7](#). Tanto los indicadores automáticos como los manuales se aplican a los datos L1 y L2 y se publican junto con los archivos de datos entregados para SMMI. La integración de estos indicadores en los datos se describe en [Sección 8](#).

Tabla 2.3: Una visión general de cada modalidad y el nivel de datos L (provenientes de AMN y PML)

Parámetro	Método	Nivel de datos
pm_1.0	amn-pm	L2a
pm_2.5	amn-pm	L2a
pm_ch_1_count	amn-pm	L1
pm_ch_1_mass	amn-pm	L2a
pm_ch_2_count	amn-pm	L1
pm_ch_2_mass	amn-pm	L2a
pm_ch_3_count	amn-pm	L1
pm_ch_3_mass	amn-pm	L2a
pm_ch_4_count	amn-pm	L1
pm_ch_4_mass	amn-pm	L2a
pm_ch_5_count	amn-pm	L1
pm_ch_5_mass	amn-pm	L2a
pm_ch_6_count	amn-pm	L1
pm_ch_6_mass	amn-pm	L2a

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

c2h6	amn-c2h6	L2a
ch4	amn-ch4	L2a
no	amn-no	L2a
no2	amn-no2	L2a
o3	amn-o3	L2a
voc	amn-voc	L2a
blackcarbon	amn-blackcarbon	L2a
co2	amn-co2	L2a
co2	amn-co	L2a

5. Modelado de datos

Modelado de datos describe el procesamiento posterior de datos de nivel inferior mediante agregación estadística estándar o análisis avanzados para obtener magnitudes geofísicas distribuidas espacialmente. La principal distinción entre los datos de nivel inferior (L2 e inferiores) y estos niveles de datos superiores radica en la transformación de datos indexados como series temporales continuas en datos discontinuos (p. ej., eventos específicos en el tiempo) o agregados a lo largo de la dimensión temporal, representando así instantáneas geospaciales en el tiempo. El propósito de estos conjuntos de datos de nivel superior es transformar los datos de monitoreo móvil en visualizaciones y análisis fáciles de interpretar para obtener información que aborde de forma más directa los objetivos de monitoreo.

5.1. Definición y aplicación de los niveles de datos 3-4

Nivel de datos 3

Nivel de datos 3 (L3) Constituye productos de datos geospaciales agregados derivados de series temporales de nivel inferior mediante métodos de agregación estadística estándar. Estos productos de datos pueden provenir de mediciones de un solo contaminante o de una combinación de múltiples contaminantes para inferir los tipos de fuentes de contaminación.

Algunos ejemplos de datos L3 incluyen:

- Ubicación de eventos de mejora individuales de TVOC, tóxicos especiados u otros contaminantes
- Grupos agrupados espacialmente de eventos de mejora de metano y etano, que indican ubicaciones de fugas persistentes de gas natural
- Mapas espacialmente continuos de concentraciones ambientales que utilizan únicamente agregación estadística básica de mediciones, sin modelos probabilísticos avanzados aplicados

Nivel de datos 4

Nivel de datos 4 (L4) se define de forma similar a L3, pero utiliza modelos probabilísticos avanzados. Los datos L4 también pueden incorporar fuentes de datos externas, como (entre otras) datos meteorológicos, datos de sitios estacionarios, capas de datos sobre el tipo de uso del suelo o tipo de carretera, o datos topográficos.

Algunos ejemplos de datos L4 incluyen:

- Un mapa continuo de promedios de contaminantes con intervalos de confianza

- asociados derivados mediante un método de reconstrucción estadística
- Un mapa de estimaciones probabilísticas de la probabilidad de una fuga de metano en un área
- Un mapa que muestra puntos críticos estadísticamente significativos de concentraciones promedio de contaminantes mediante el método Getis-Ord (Gi*)

5.2. Flujos de trabajo de modelado de datos

Los flujos de trabajo de modelado funcionan para transformar datos L1/L2 en datos L3/L4 para crear análisis de datos centrados en la identificación de ubicaciones de fuentes de contaminantes y áreas sobrecargadas por contaminantes conocidos.

Tanto L3 como L4 incluyen dos categorías generales de análisis.

1. *Análisis basados en mejoras* identificar eventos en el tiempo y el espacio donde las concentraciones de un contaminante se pueden distinguir de forma medible de las concentraciones ambientales de fondo.
2. *Análisis basados en la concentración ambiental* identificar gradientes espaciales en las concentraciones ambientales promedio o típicas observadas para contaminantes atmosféricos específicos o combinaciones de contaminantes. Los datos específicos de L3 y L4 que se generarán para SMMI se detallarán en los CAMP tras comprender mejor las preocupaciones y prioridades de la comunidad. Las posibles opciones para estos análisis y los detalles adicionales sobre los tipos de análisis se discuten en [Sección 9](#).

Los datos fuente subyacentes para todos los conjuntos de datos L3 y L4 serán los datos finales (como L1 y L2a) incluidos en la entrega a CARB. Antes de incluirlos en el flujo de trabajo de modelado, se eliminan los datos marcados como no válidos.

Flujo de trabajo de modelado básico para análisis basados en mejoras

1. Identificación de eventos de mejora en la serie temporal de contaminantes mediante la "búsqueda de picos" (L1/L2a→L3)
2. Análisis de múltiples contaminantes para la categorización de eventos de mejora (L3→L3)
3. Agrupación espacial o clusterización de eventos de mejora y generación de estadísticas básicas por cluster (L3→L3)

Flujo de trabajo de modelado básico para análisis basados en la concentración ambiental

El flujo de trabajo básico de modelado para análisis basados en la concentración ambiental puede variar según el tipo de análisis y el nivel de datos a generar. A continuación, se presentan algunos ejemplos de flujos de trabajo:

Agregación estadística básica de concentraciones ambientales (L3)

1. Agregue datos de series temporales a unidades espaciales (es decir, segmentos de carretera, hexbins, etc.) para definir un único "pase" o "visita", calculando, por ejemplo, la media (L1/L2a→L3)
2. Agregue las medias de un solo paso durante todo el período de monitoreo utilizando, por ejemplo, la mediana (L3→L3)

Reconstrucción estadística de concentraciones ambientales¹ (L4)

1. Agrupar los datos de series temporales por unidad espacial de elección durante todo el período de monitoreo (segmentos de carretera, hexbins, etc.) (L1/L2a→L1/L2a)
2. Los datos de series temporales agrupados espacialmente se combinan con datos meteorológicos, topográficos e información sobre el tipo de carretera. El resultado es una estimación probabilística con intervalos de confianza en la unidad espacial definida en el paso 1 (L1/L2a→L4).

Si bien estos son flujos de trabajo potenciales y podrían no reflejar completamente lo que se incluye en los CAMP y las visualizaciones finales para SMMI, establecen un marco para comunicar los flujos de trabajo de modelado asociados con los diferentes productos de datos derivados. Este marco se utilizará en los CAMP y en el material descriptivo que acompaña a las visualizaciones públicas finales.

¹ Los detalles específicos de este enfoque están fuera del alcance de este documento, pero se comunicarán en los CAMP si se elige este método para el análisis.

6. Almacenamiento de datos

Un paso esencial en el proceso de gestión de datos es *almacenamiento de datos*, desde la ingesta de datos de Nivel 0 hasta las operaciones computacionales que respaldan el procesamiento de datos en los Niveles 3 y 4.

La estrategia de almacenamiento de datos de Aclima incluye los siguientes componentes:

1. Uso de almacenes de datos en la nube escalables y estándar de la industria para proporcionar almacenamiento persistente a largo plazo de múltiples niveles de datos.
2. Uso de herramientas de computación en la nube para respaldar procesos de streaming y por lotes para transformaciones eficientes de datos de gran volumen entre niveles, con capacidad de escribir en todos los almacenes de datos.
3. Una arquitectura de almacenamiento que admite operaciones de relleno para volver a transformar datos entre niveles, cuando los parámetros de calibración actualizados y el marcado de datos indican la necesidad.
4. Uso de una interfaz de consulta central lista para usar para datos en varios niveles.
5. Uso de una base de datos de metadatos en la nube para poder conectar los datos al hardware de los sensores físicos y a los vehículos donde se instalaron los sensores, y para rastrear los parámetros de calibración.
6. La capacidad de escribir archivos planos para la entrega final a los clientes y la disponibilidad de una interfaz de usuario basada en la nube, herramientas de línea de comandos y API para acceder a estos datos.

Para el proyecto SMMI, Aclima proporcionará a CARB acceso a los datos almacenados durante tres meses tras la finalización del contrato. Posteriormente, Aclima retirará el acceso a CARB y deshabilitará el almacenamiento de datos en la nube.

7. Revisión de datos y garantía de calidad

El sistema de gestión de datos incorpora soporte para la revisión de datos, definida como el marcado manual o automático de señales automatizadas de las series temporales de sensores. Esta sección destaca los componentes de ingeniería del proceso de revisión de datos y control de calidad. Los detalles científicos de la revisión de datos se pueden encontrar en los documentos de control de calidad (QA/QC) correspondientes (que se incluirán en los CAMP).

7.1. Descripción general

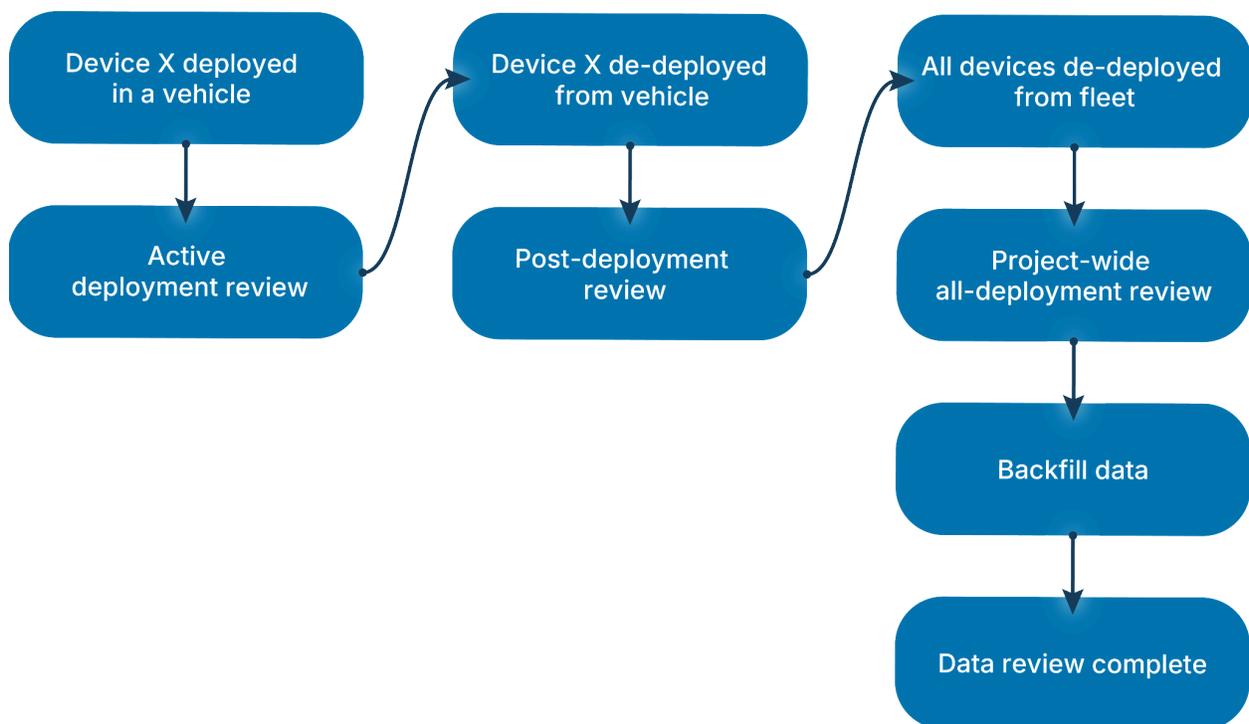


Figura 7.1 Descripción general del proceso de revisión de datos y control de calidad. Cada etapa de la revisión incluye actividades manuales y automatizadas.

7.2. Revisión de la implementación activa

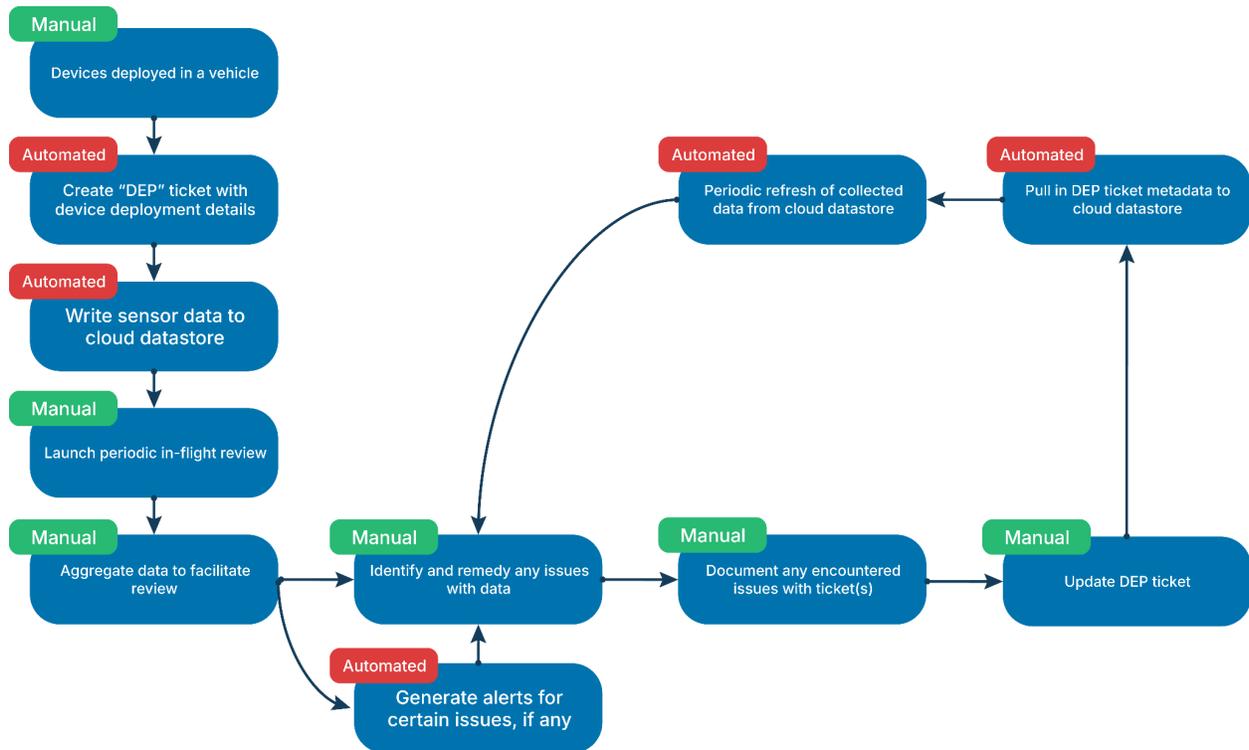


Figura 7.2:Proceso de revisión de datos de un dispositivo individual mientras su implementación está en curso o activa. Las marcas "Manual" indican el trabajo realizado por el equipo de Operaciones de Calidad de Datos (DQO).

Cada implementación de un dispositivo comienza cuando se instala o despliega en un vehículo. Desde su instalación hasta su desinstalación, la implementación se considera activa. Los datos de los sensores de los dispositivos se revisan periódicamente a medida que avanzan las implementaciones activas. Las principales etapas de este proceso se describen a continuación.

7.2.1. Flujo de trabajo de revisión de la implementación activa

Paso 1: Crear un ticket "DEP" con los detalles de implementación del dispositivo

La instalación de un dispositivo en un vehículo marca el inicio de su despliegue en dicho vehículo. Cada instalación activa la creación automática de un ticket de despliegue único (DEP) en el sistema de tickets estándar de Aclima. Cada ticket DEP se utiliza como repositorio central para que los revisores de datos registren información sobre todos los problemas de calidad de los datos y de hardware detectados, el mantenimiento realizado,

las comprobaciones estandarizadas del sistema realizadas (p. ej., comprobaciones de flujo, de conectividad, etc.) y cualquier otra información relevante durante un despliegue.

Paso 2: Escribe los datos del sensor en el almacén de datos en la nube

Una vez instalado y asociado un dispositivo a un vehículo, se puede encender para iniciar la recopilación de datos. Durante la recopilación, los datos del sensor se transmiten desde la AMN y se almacenan en la nube.

Paso 3: Iniciar una revisión periódica durante el vuelo

Una vez que los dispositivos recopilan datos, el equipo de Operaciones de Calidad de Datos (DQO) comienza a realizar verificaciones periódicas de los datos para detectar y solucionar cualquier problema de calidad de datos o sensores a medida que surja.

Paso 4: Agregue datos para facilitar la revisión

Para visualizar los datos recopilados, los datos de los sensores deben importarse a una base de datos y agregarse temporalmente (por ejemplo, a 1 minuto o 1 hora) para que la revisión de datos sea más eficiente. Posteriormente, están disponibles para su manipulación manual y se verifican mediante la señalización automatizada de problemas. Los datos subyacentes no agregados (a 1 Hz) siguen siendo accesibles para los revisores cuando es necesario centrarse en intervalos cortos (de minutos a horas) de interés.

Paso 5: Identificar y solucionar cualquier problema con los datos

El equipo de DQO visualiza manualmente los datos de recolección a medida que llegan para identificar, descartar y resolver rápidamente los problemas que surgen en la categoría. Los datos se actualizan automáticamente y con frecuencia en estas herramientas.

Paso 6: Generar alertas para ciertos problemas, si los hubiera

Ciertos problemas de calidad de datos o de sensores tienden a mostrar el mismo comportamiento cada vez que ocurren y pueden traducirse en una regla automatizada para alertar a los usuarios cuando se identifica este comportamiento.

Paso 7: Documente cualquier problema encontrado con los tickets

Cuando se detectan problemas mediante alertas automatizadas o personalizadas, o durante la revisión manual de datos, se documentan mediante un ticket de Revisión de Datos de AMN (ADR). Las alertas generadas manualmente, también conocidas como "omisiones", se adjuntan a los fragmentos de datos que el equipo de DQO desea omitir.

Las razones para marcar datos como no válidos (omisión) pueden incluir:

- Tiempo de calentamiento del sensor más largo de lo habitual: a veces, los sensores de un dispositivo tardan más de lo habitual en calentarse. Los datos medidos lo reflejan, mostrando un comportamiento de la serie temporal uniforme, con aumentos o descensos constantes, durante la primera hora del turno de un vehículo.
- Fuga: el tubo de muestra se desconecta del dispositivo y este comienza a tomar muestras desde el interior del vehículo, lo que genera mediciones poco realistas.
- Falla general del sensor: un sensor falla por completo mientras está implementado y comienza a informar datos no físicos.
- Bloqueo de flujo o flujo defectuoso: el camino del flujo se obstruye debido a un filtro obstruido o un objeto que ingresa al camino del flujo, o la bomba de flujo comienza a fallar, lo que genera tasas de flujo fuera de los límites aceptables.

Paso 8: Actualizar el ticket DEP

Los tickets de omisión generados durante la revisión activa de datos se vinculan a los tickets de problemas de ADR que los provocaron, para facilitar la trazabilidad en la base de datos de Aclima hasta el motivo de la omisión. Además, todos los tickets de problemas y omisiones se vinculan a su ticket DEP correspondiente.

Paso 9: Extrae metadatos de DEP al almacén de datos en la nube

Los tickets de omisión creados durante la revisión activa de datos se incorporan periódicamente al almacén de datos en la nube, donde se guardan sus detalles hasta que se incorporan para generar indicadores de validez y control de calidad para el producto de datos final durante el relleno posterior a la implementación.

Paso 10: Actualización periódica de los datos recopilados del almacén de datos en la nube

Las tablas de datos que almacenan datos de sensores y metadatos de tickets se actualizan periódicamente a medida que se recopilan nuevos datos, y el proceso de revisión de datos activos se repite para esta siguiente sección de datos.

7.2.2. Alertas personalizadas

Además de la revisión manual y las alertas automatizadas básicas, el equipo de revisión de datos puede definir alertas basadas en umbrales para alertar sobre condiciones específicas que puedan indicar un comportamiento atípico de un sensor determinado que deba resolverse. El equipo utiliza visualizaciones para obtener más información sobre el problema, generar un ticket de incidencia e iniciar el proceso de resolución.

7.3. Revisión posterior a la implementación

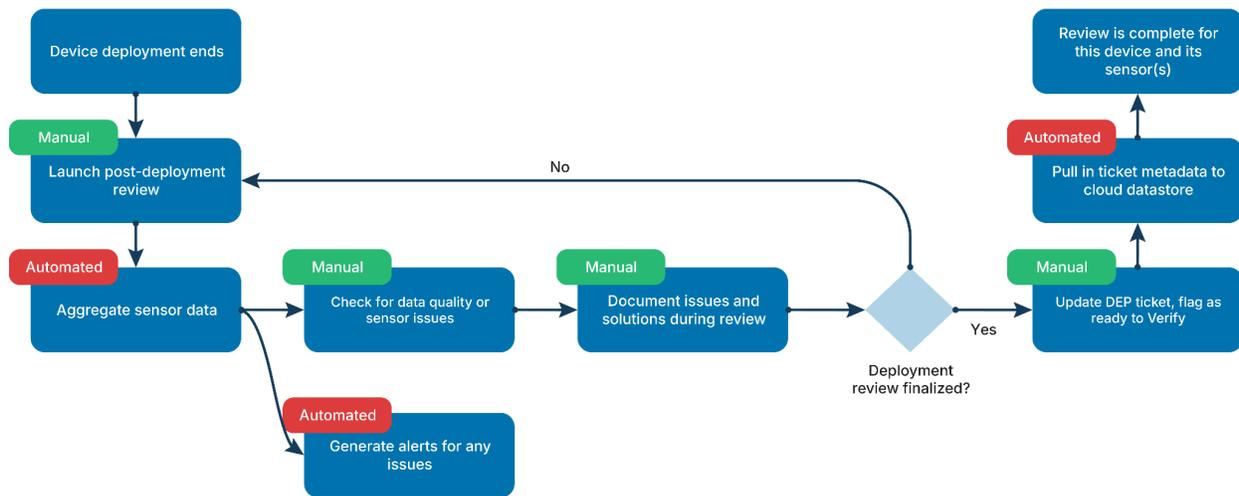


Figura 7.3 Proceso de revisión de datos para un solo dispositivo una vez finalizada la implementación. Consulte el texto a continuación para obtener información detallada sobre cada paso.

La fase de revisión posterior a la implementación es una revisión final de todos los datos de los sensores recopilados durante una implementación determinada, para garantizar que se detecten y aborden todos los problemas no identificados durante las revisiones de datos activas.

7.3.1. Flujo de trabajo de revisión posterior a la implementación

Paso 1: Iniciar la revisión posterior a la implementación

Una vez finalizada la implementación de un dispositivo, el equipo DQO inicia una revisión posterior a la implementación de todos los datos de los sensores recopilados durante la implementación de ese dispositivo, para garantizar que se detecten todos los problemas que puedan haber pasado desapercibidos durante el período de implementación antes de que se verifiquen finalmente los datos.

Paso 2: Agregar datos del sensor

Los metadatos del ticket que detallan la implementación del dispositivo se utilizan para consultar todos los datos de los sensores asociados a dicha implementación durante su duración. Estos datos se compilan en una base de datos.

Paso 3: Generar alertas para cualquier problema

Cualquier incidencia de problemas de calidad de datos bien caracterizados se marca mediante alertas en un panel accesible para el equipo de DQO. Estos problemas se cuantifican en una vista resumida para toda la implementación, lo que ayuda al equipo de DQO a abordar eficazmente su solución durante la revisión posterior a la implementación.

Paso 4: Verifique la calidad de los datos o problemas con el sensor

El equipo de revisión utiliza diversas herramientas de visualización para detectar problemas de calidad de datos o sensores. Cualquier problema detectado se registra mediante un ticket de Revisión de Datos de AMN (ADR). Este ticket de ADR está vinculado al ticket DEP de la implementación, de modo que todos los problemas de calidad de datos o hardware detectados por dicha implementación se pueden rastrear desde un lugar centralizado.

Paso 5: Documentar los problemas y las soluciones durante la revisión

De manera similar a la revisión de implementación activa, los problemas encontrados durante la revisión posterior a la implementación se documentan mediante tickets de problemas y/o omisiones de ADR.

Paso 6: Verificación finalizada de la revisión de la implementación

Una vez que un miembro del equipo DQO ha concluido una revisión posterior a la implementación, actualiza el ticket DEP de la implementación para reflejar que los sensores del dispositivo de implementación han recibido su revisión de datos inicial.

Paso 7: Actualizar el ticket DEP y marcarlo como listo para verificar

Si se finaliza la revisión, la implementación está lista para ser completada y el ticket de DEP recibe una marca de "Listo para verificar".

Paso 8: Extraiga los metadatos del ticket al almacén de datos en la nube

Los metadatos de los tickets ADR, DEP y VER de Aclima se extraen al almacén de datos en la nube. Esta información es referenciada por flujos de trabajo de reposición posteriores que reprocesan los datos de los sensores y los marcan como "Verificados" según los metadatos de los tickets adjuntos.

7.4. Revisión posterior a todas las implementaciones

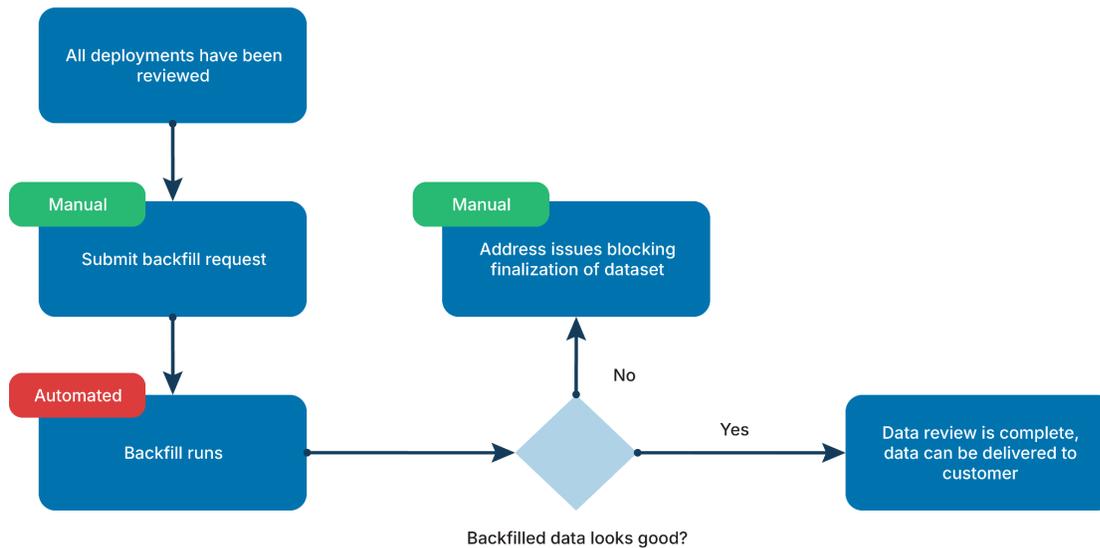


Figura 7.4:Proceso de revisión de datos una vez que se completan todas las implementaciones de AMN.

Una vez que se hayan revisado todas las implementaciones relevantes para el contrato SMMI durante el período de tiempo pertinente, se realizará un *relleno*. Se ejecuta el proceso de los datos recopilados. El proceso de reabastecimiento reprocesa todos los datos recopilados, junto con los metadatos relacionados con la implementación, para generar datos verificados listos para ser entregados al cliente.

Definición de relleno

Los procesos de revisión y calibración de datos durante y después de la implementación generan un conjunto de indicadores de datos determinados manualmente y parámetros de calibración actualizados que deben aplicarse. Para que estas correcciones se apliquen, *relleno* debe activarse para reprocesar las transformaciones de datos de Nivel 1 que hacen uso de parámetros de calibración del sensor y cualquier agregación posterior (es decir, flujos de trabajo de modelado para producir datos L3 o L4) para todos los datos recopilados en la categoría.

Un reabastecimiento es un proceso mediante el cual se reejecuta la totalidad (o una parte) del flujo de trabajo diario para incorporar los parámetros actualizados, y es fundamental para la plataforma de datos de Aclima. Las tablas del almacén de datos incluyen versiones de datos asociadas con el código utilizado para procesarlos y rastrear su linaje. Tenga en cuenta que un reabastecimiento nunca altera los valores almacenados inmediatamente después de la ingesta (normalmente datos L0 y, en algunos casos, datos L1).

El proceso de reposición genera además metadatos que categorizan las marcas de datos y detallan el período en que se generaron. También se incorpora el estado de verificación de los tickets de revisión de datos. Estas anotaciones estarán disponibles para CARB en los archivos de datos entregados.

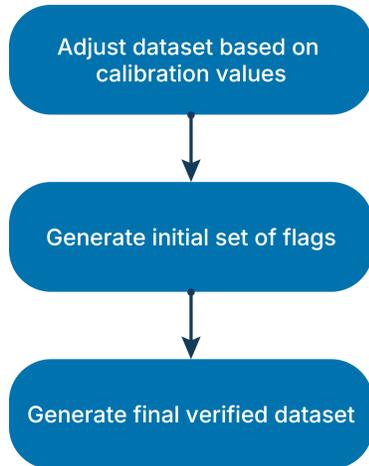


Figura 7.5 Resumen de las tres etapas principales del relleno. En el paso 1, se consideran los ajustes de los valores de calibración y se aplican si es necesario. En el paso 2, los metadatos generados previamente que indican datos potencialmente inválidos se convierten en indicadores formales. En el paso 3, una verificación final verifica todos los indicadores y prepara los datos finales de Nivel 2a para su entrega.

7.4.1. Flujo de trabajo de revisión posterior a todas las implementaciones

Paso 1: Enviar solicitud de reposición

El equipo de DQO envía una solicitud para completar todos los datos revisados, incluyendo los detalles utilizados para dar formato al flujo de trabajo de reposición. Una vez completados estos detalles, se programa e inicia la reposición.

Paso 2: Ejecuciones de relleno

Se ejecuta una secuencia de flujos de trabajo para reprocesar los datos del sensor junto con los metadatos de los tickets de revisión de datos generados durante los procesos de revisión activa y posterior a la implementación.

Durante el relleno, el flujo de trabajo de relleno hace referencia a los metadatos del ticket de omisión para marcar cualquier dato aplicable como no válido, a fin de garantizar que los datos omitidos se puedan excluir o filtrar fácilmente de los productos de datos finales si así se desea.

Además de los metadatos de los tickets de omisión, también se incorporan los metadatos de los tickets de metadatos de implementación (DEP) y metadatos de verificación (VER) que se marcaron como Listos para verificar, para marcar todos los puntos de datos aplicables como Verificados.

Paso 3: Verifique que los datos rellenos se vean bien

Tras finalizar el proceso de reposición, el equipo de revisión vuelve a consultar los datos reprocesados para actualizar sus herramientas de visualización de datos como última comprobación y garantizar que los datos verificados estén listos para su entrega a CARB. Si los datos reposición superan esta comprobación, el equipo de DQO puede concluir el proceso de revisión de la implementación para ese período contractual y sus implementaciones.

Paso 4: Abordar los problemas que bloquean la finalización del conjunto de datos

Si se identifican datos erróneos, el equipo DQO básicamente repite los pasos 1 a 3 para corregir el problema a través de otra ronda de revisión enfocada, solicita un seguimiento específico para resolver los problemas pendientes y luego verifica los datos completados una vez más.

7.5. Revisiones de datos

Los socios de Partner Mobile Lab (PML) entregarán datos finalizados con la misma cadencia y el mismo esquema al depósito de almacenamiento de datos en la nube accesible a CARB, como se describe en [Sección 8.2](#). Sin embargo, durante un período posterior a la finalización del periodo de recopilación de datos (hasta uno o dos años, o incluso más), podrían continuar las investigaciones. Esto podría dar lugar a revisiones necesarias, como las de los parámetros de calibración, que CARB podría desear difundir interna o externamente.

Antes de que finalice el contrato, los PML enviarán actualizaciones a Aclima. Aclima subirá el nuevo archivo al almacén de datos en la nube, incrementará el número de revisión en el nombre del archivo y actualizará el archivo README del almacén de datos (consulte [Sección 8.2.3](#)).

Una vez finalizado el período del contrato, Aclima dejará de participar en estas revisiones generadas por PML. Sin embargo, Aclima recomienda que los PML envíen las actualizaciones directamente a CARB y, para mantener la coherencia, sigan el mismo protocolo de nomenclatura de revisiones (véase [Sección 8.2.3](#)).

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

8. Transferencia de datos

Esta sección describe cómo Aclima formateará y entregará *finalizado* datos a CARB. Los datos finalizados se refieren a los datos verificados correspondientes al nivel más alto disponible de datos no modelados para cada flujo de medición, generalmente Nivel 2a, aunque puede ser Nivel 1 en algunos casos. Se centra en la descripción del esquema de datos y su aplicación típica, y en cómo las herramientas de almacenamiento en la nube facilitarán la transferencia de datos, junto con consideraciones de usabilidad y seguridad.

8.1. Esquema de datos de nivel 2a

8.1.1. Descripción general del esquema

Una subtarea crítica del SMMI es la entrega de datos finales (es decir, series temporales de concentración de 1 Hz geolocalizadas) para todos los contaminantes medidos por Aclima o los Laboratorios Móviles Asociados (PML). Los datos entregados deben estar formateados en un esquema que contenga la información necesaria para describir la concentración de una ubicación en un momento determinado y comprender cualquier información contextual importante, como indicadores de calidad de datos y metadatos de vehículos. Esta información permitirá a los reguladores de la calidad del aire hacer un uso eficaz de estos datos.

El esquema que se muestra en [Tabla 8.1](#) Contiene toda esta información en un formato mínimo que intenta lograr un equilibrio entre limitar la información redundante y permanecer accesible para los analistas que no deberían necesitar aplicar múltiples operaciones de unión para preprocesar los datos.

Tabla 8.1: esquema de datos para datos de nivel 2a

Nombre dla categoría	Tipo de campo	Descripción dla categoría
marca de tiempo	CADENA	Marca de tiempo UTC con codificación ISO 8601 en la que se tomó la medición, con información de zona horaria. La cadena de formato es AAAA-MM-DDTHH:mm:ssZ (p.ej. 03/09/2024 14:46:39)
parámetro	CADENA	El mensurando descrito por la fila, generalmente un contaminante (p. ej., no, no2, co) o un subtipo de contaminante (pm2.5_ch1). Para cada valor de parámetro, una tabla de búsqueda independiente

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

		vincula este campo con los códigos de parámetro del SCA (relación 1:1) y una descripción textual más extensa (véase Tabla 8.2 ; ver Sección 8.1.5 (para una nota general sobre los códigos AQS).
método	CADENA	Un identificador de método para el sensor utilizado para obtener la medición. Véase la explicación en Sección 8.1.4A continuación. Para cada valor de método, una tabla de búsqueda independiente vincula este campo con el código de método AQS correspondiente (relación 1:1) y una descripción de texto más larga (consulte Tabla 8.2 ; ver Sección 8.1.5 (para una nota general sobre los códigos AQS).
duración	ENTERO	Duración de la muestra para el mensurando asociado. Esto es relevante en particular para ciertos métodos que requieren tiempos de muestreo superiores a 1 segundo o que pueden tener tiempos de muestreo variables a lo largo del período de recopilación de datos. En estos casos, marca de tiempo la categoría representa el inicio del período de muestreo.
valor	FLOTAR	El valor de medición del mensurando parámetro En el momento marca de tiempo en la ubicación ubicación .
unidad	CADENA	La unidad de medida que valor representa.
latitud	FLOTAR	La latitud en la que se tomó la medición. Se utiliza junto con la longitud para describir completamente la posición corregida de la medición.
longitud	FLOTAR	La longitud en la que se tomó la medición. Se utiliza junto con la latitud para describir completamente la posición corregida de la medición.
indicador de	ENTERO	Un valor entero que representa un booleano (0 → falso o 1 → verdadero). Un valor de 1 indica que

estado		esta medición ha sido marcada, lo que significa que hay una anotación relevante (en códigos calificadores) que contextualiza la calidad o condiciones de los datos.
códigos calificadores	CADENA	Un conjunto de códigos AQS separados por comas que documenta el motivo de la exclusión de los datos o señala cualquier condición notable anotada mediante códigos AQS. Tenga en cuenta que los datos solo se reportan cuando los instrumentos están activos dentro de un área de recolección de datos definida. Tenga en cuenta que cualquier dato recopilado mientras los instrumentos están activos y en el área de recolección de datos, pero que se invalida por otra razón, <i>voluntad</i> ser incluido.
id del vehículo	CADENA	Una cadena que identifica de forma única el vehículo que recopiló la medición descrita por la fila.

8.1.2. categorías de esquema y utilidad

Hay dos categorías principales de categorías en este esquema simple: primarios y metadatos. Las categorías primarios son `marca de tiempo, parámetro, longitud, latitud, unidades, y valor`. Estos categorías describen la hora en que se tomó la medición, el contaminante medido, la ubicación, el valor y las unidades de medida.

Las categorías de metadatos son `indicador de estado, códigos calificadores, método y id del vehículo`. Estos categorías proporcionan un contexto importante que ayuda en la interpretación de la medición descrita en Las categorías primarios. `indicador de estado, códigos calificadores, y id del vehículo` describir la calidad de la medición a través de `indicador de estado y códigos calificadores` y la plataforma que recopiló la medición a través de `id del vehículo`. la categoría `método` se utiliza para identificar el tipo de sensor o el método de detección utilizado para la recopilación de la medición.

8.1.3. Tabla de búsqueda de códigos AQS

La tabla de búsqueda de códigos AQS se utiliza para establecer un puente entre el esquema Aclima y los códigos de parámetros y métodos AQS, como se describe en [Tabla 8.2](#).

Tabla 8.2: mapeo entre las claves del esquema Aclima y Las categorías de la tabla de búsqueda para permitir la conexión con los códigos de parámetros y métodos de AQS

Aclima schema key	categorías de la tabla de búsqueda	
<i>parámetro, por ejemplo, "no2"</i>	<i>AQS código de parámetro, p.ej. "42602"</i>	<i>Descripción, por ejemplo, "Dióxido de nitrógeno (NO2)"</i>
<i>método, por ejemplo, "amn-no2"</i>	<i>AQS método código, p.ej. "010"</i>	<i>Descripción, por ejemplo, "Monitor pasivo integrado"</i>

8.1.4. Rellenar la categoría de método

El método de detección utilizado para recopilar la medición descrita por una fila constituye un metadato importante que no puede descartarse, especialmente dada la diversidad de mediciones entre Aclima y los tres PML. Proponemos incluir esta información directamente en el esquema de datos, como se indica a continuación:

1. Utilizar y ampliar, cuando sea aplicable, los códigos de método descritos en el diccionario de codificación AQS.
2. Diseñar identificadores de cadena únicos para diferentes sensores e incluirlos como categorías de identificación.

8.1.5. Nota general sobre los códigos AQS

La selección exacta de parámetros, métodos y códigos de calificación se desarrollará en coordinación con CARB y los equipos de PML. Habrá cierta superposición con los parámetros, métodos y códigos de calificación regulatorios existentes de AQS, pero probablemente también se incluirán nuevos códigos para mediciones únicas realizadas como parte de este proyecto.

8.2. Transferencia de datos de nivel 2a

8.2.1. Alcance y objetivos de la transferencia de datos finalizada

Aclima ha diseñado su enfoque de transferencia de datos principalmente para garantizar que CARB tenga acceso a los datos finalizados de la forma más rápida y fluida posible. Aclima pondrá los datos a disposición a través del almacenamiento en la nube periódicamente. Aclima espera que CARB extraiga los datos del almacenamiento en la nube cuando le resulte conveniente.

8.2.2. Disponibilidad de los datos finalizados a través de Google Cloud Storage

Aclima utilizará Google Cloud Storage (GCS) para almacenar instantáneas de los datos finalizados. GCS es un servicio gestionado por Google que permite almacenar datos no estructurados de cualquier tamaño y recuperarlos en cualquier momento. GCS ofrece diversas interfaces de usuario, bibliotecas de cliente y herramientas de línea de comandos para acceder a los datos o integrarse con otros proveedores de servicios en la nube. [Figura 8.1](#) A continuación se muestran flujos opcionales para la transferencia de datos, que se describen con más detalle en las siguientes secciones.

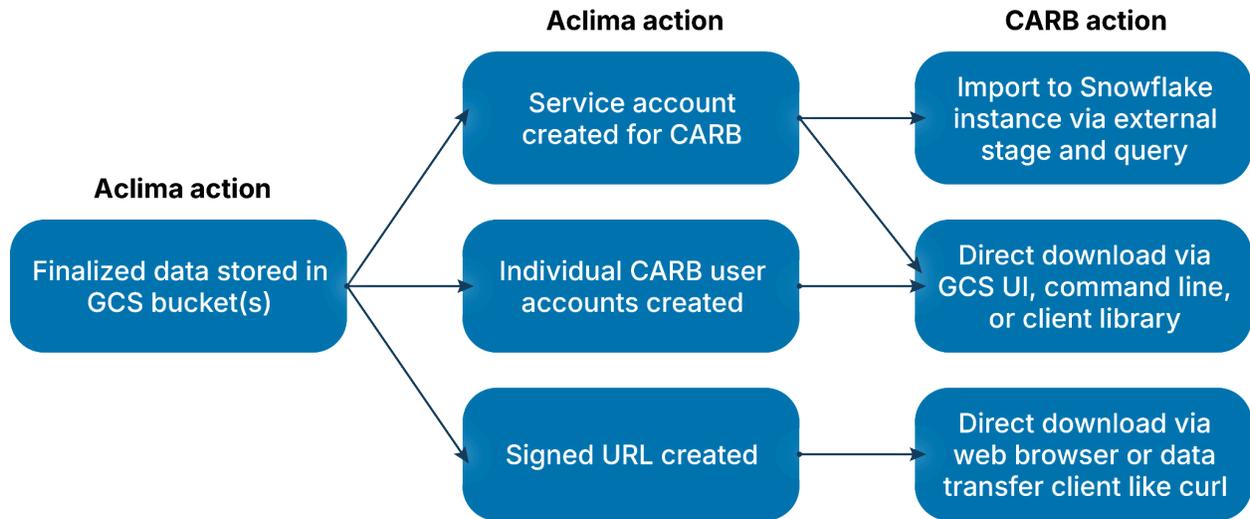


Figura 8.1: opciones para acceder a datos de Nivel 2a desde los depósitos GCS, utilizando una variedad de medios según la preferencia de CARB

El bucket de GCS se configurará inicialmente como multirregional para un acceso con la menor latencia. Su clase de almacenamiento predeterminada será estándar, optimizada para un acceso frecuente a bajo costo y con tarifas de almacenamiento en reposo más altas. Si los patrones de acceso indican que se accede a un archivo con poca frecuencia (menos de una vez al mes), la clase de almacenamiento podría migrarse a nearline o coldline, posiblemente mediante el mecanismo de clasificación automática de GCP.

8.2.3. Formato de archivos de datos

Los datos brutos de 1 Hz se proporcionarán en archivos comprimidos con valores separados por comas (csv), uno por `método de organización nombre por mes`.

Por ejemplo, todos los $PM_{2.5}$ Los datos del instrumento AMN PM para el mes 1 de recopilación de datos estarían en un solo archivo csv; todos los $PM_{2.5}$ del instrumento AMN PM para el mes 2, estaría en un archivo csv separado y único.

Se seguirá la siguiente convención de nomenclatura, que se describe en detalle en [Tabla 8.3](#):

```
gs://[bucket]/[organización]_[método]_[fecha de medición]_[número de revisión]_[fecha de revisión]
```

P.ej.

gs://bucket123/aclima_AMN-bc_202605_r1_20261111

Además, se incluirá un archivo Léame en el nivel superior del almacén de datos en la nube para documentar la naturaleza de cualquier revisión realizada, comenzando con la primera versión de cualquier archivo cargado ("r1").

Tabla 8.3: definición de elementos de la convención de nomenclatura

gs://	El dominio GCS
balde	El contenedor GCS que contiene todos los datos; será el mismo para todas las entregas
organización	Indica la organización que realizó el seguimiento, por ejemplo, "aclima"
método	Nombre del instrumento de recolección, p. ej., "amn-bc" para el sensor de carbono negro del nodo móvil Aclima. Se garantizará que la organización y el método sean únicos.
fecha de medición	AAAAMM fecha de recolección, p. ej. "202605"
número de revisión	La primera versión del archivo cargada será "r1". Las versiones revisadas posteriores, si las hubiera, serán r2, r3, etc. Los motivos y la naturaleza de la revisión se incluirán en el archivo Léame del contenedor de datos.
fecha de revisión	AAAAMMDD Fecha de revisión, si la hubiera, p. ej., "20261111". Para la primera versión del archivo cargada, la fecha de revisión será la fecha de carga.

8.2.4. Frecuencia de entrega y tamaño estimado de archivo

Los datos finales se transferirán a CARB mensualmente, comenzando cuatro meses después del inicio del monitoreo. Los datos recibidos de los sensores se transfieren a todos los niveles de procesamiento de datos y a las instantáneas de datos finalizadas en un contenedor GCS, donde estarán disponibles para CARB. La última entrega de datos se realizará al final del contrato. [Figura 8.2](#) A continuación se describe una serie hipotética de entregas de datos, suponiendo un período de monitoreo hipotético que comienza el 1 de junio de 2025 y termina el 30 de octubre de 2025. Nota: esto es solo a modo de ilustración; el período de monitoreo propuesto real es de 9 meses.

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

En este escenario, la primera entrega de datos ocurre 4 meses después del inicio (1 de octubre de 2025, a más tardar) e incluye todos los datos recopilados durante junio de 2025. Las entregas de datos continúan mensualmente de la misma manera hasta la entrega final de datos 3 meses después del final del monitoreo (1 de febrero de 2026, a más tardar), que cubre el monitoreo realizado durante octubre de 2025. Los nuevos meses de datos se agregarán al depósito como nuevos archivos (consulte la convención de nomenclatura de archivos, [Sección 8.2.3](#)), mientras que las entregas mensuales anteriores siguen disponibles. Para el momento de la entrega del mes pasado, los datos de todos los meses están disponibles como archivos individuales.

Todos los datos de cada entrega se entregarán simultáneamente en el contenedor GCS. Una vez disponibles, se notificará automáticamente a CARB por correo electrónico al gestor de proyectos de CARB. CARB prevé que el tamaño de las entregas mensuales de datos será de decenas a cientos de GB.

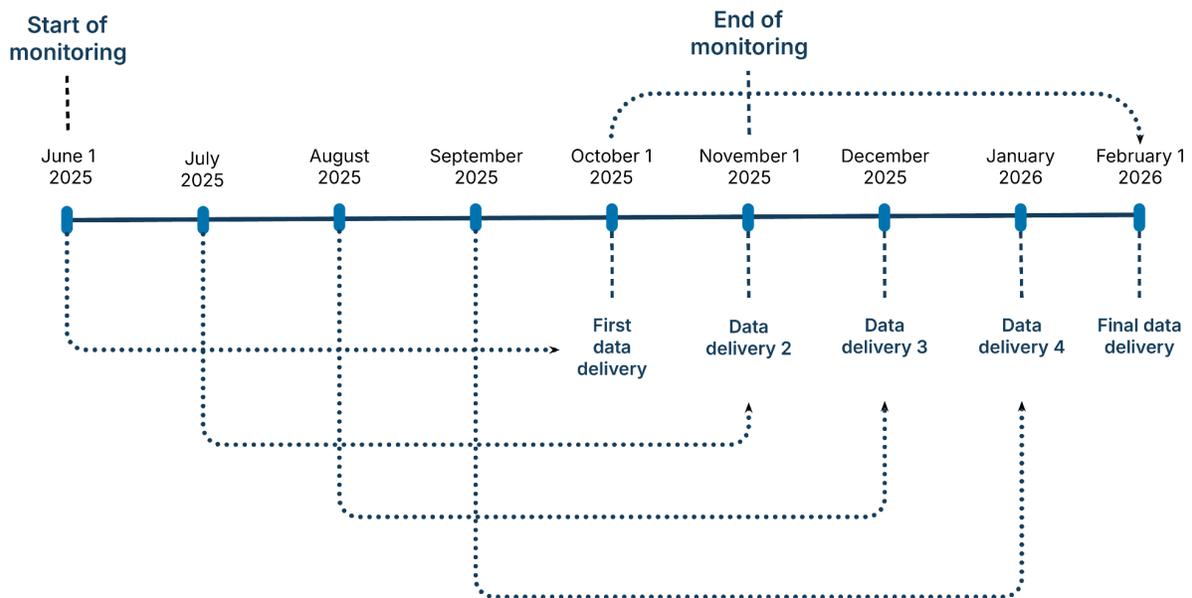


Figura 8.2: Ejemplo de cadencia de entrega de datos, asumiendo un período de monitoreo hipotético solo del 1 de junio de 2025 al 1 de noviembre de 2025. *Tenga en cuenta que esto es sólo para fines de ilustración sencilla; el período de seguimiento real propuesto es de 9 meses.* En este ejemplo hipotético, la primera entrega de datos, el 1 de octubre de 2025, abarca todos los datos recopilados entre el 1 y el 30 de junio de 2025. La última entrega, el 1 de febrero de 2026, abarca todos los datos recopilados entre el 1 y el 31 de octubre de 2025. Las entregas de datos siempre son acumulativas dentro del contenedor, ya que se suman al conjunto de archivos existente.

8.2.5. Obligación de alimentos

Aclima proporcionará a CARB acceso a los datos almacenados durante tres meses tras la finalización del contrato. Posteriormente, Aclima retirará el acceso a CARB y desaprovisionará el contenedor de almacenamiento.

8.3. Vías de transferencia propuestas

Aclima entiende que CARB prefiere transferir directamente de GCS a Snowflake mediante métodos programáticos. Esto se describe a continuación, junto con otras posibles vías.

8.3.1. Ruta preferida: Integración de Snowflake

Aclima colaborará con el departamento de TI de CARB para establecer una integración con GCS de Aclima y la instancia Snowflake de CARB. El departamento de TI de CARB en Snowflake creará una cuenta de servicio a la que Aclima podrá otorgar permisos para acceder a los buckets de GCS de Aclima que almacenan los datos finalizados. Aclima otorgará a la cuenta de servicio creada por Snowflake permisos para acceder a los objetos del bucket mediante Google Cloud IAM. Este enfoque tiene la ventaja de que no requiere que ningún usuario individual de CARB tenga cuentas de GCP: solo se autentica la cuenta de servicio.

Los pasos técnicos clave en esta integración se describen a continuación y en [Figura 8.3](#), y en detalle en el documento oficial [Tutorial de integración de GCS<>Snowflake](#) de Copo de nieve.

1. **(Acción CARB) Crear una integración de almacenamiento en Snowflake:** Use CREAM INTEGRACIÓN DE ALMACENAMIENTO para crear un objeto de integración que Snowflake utiliza para autenticarse con GCS, estableciendo ubicaciones de depósito permitidas (y opcionalmente, bloqueadas) para el control de acceso.
2. **(Acción CARB) Recuperar el ID de la cuenta de servicio GCS:** Ejecute DESC STORAGE INTEGRATION <integration_name> para obtener el ID de la cuenta de servicio GCS administrada por Snowflake.
3. **(Acción Aclima) Otorgar permisos en Google Cloud:** En GCS, asigne la función de Visor de objetos de almacenamiento a esta cuenta de servicio para los depósitos especificados.

4. **(Acción de CARB) Crear un escenario externo:** Defina una etapa externa en Snowflake que haga referencia a la integración de almacenamiento y especifique la URL del depósito de GCS.
5. **(Acción CARB) Cargar o descargar datos:** Utilice COPY INTO u otros comandos para interactuar con datos en GCS a través de la etapa Snowflake.

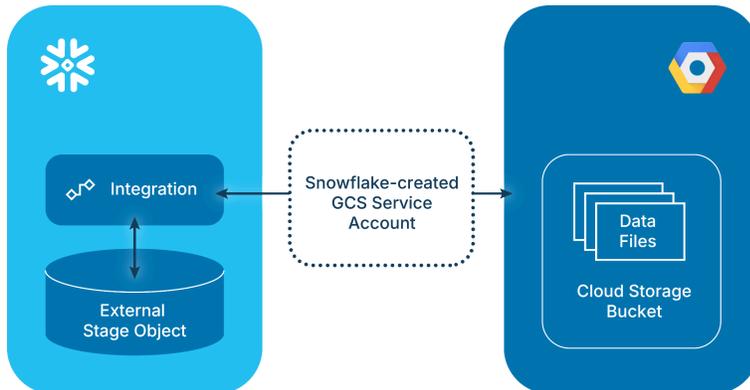


Figura 8.3: flujo de integración para una etapa de Snowflake Cloud Storage.

8.3.2. Ruta alternativa: Interfaz de usuario y funciones de las herramientas de la consola de Google Cloud

Una forma alternativa de acceder a los datos es mediante las herramientas integradas en Google Cloud Platform (GCP), como se describe a continuación. Estas vías de acceso requieren que cualquier persona que interactúe con los datos a través de GCS tenga una cuenta de GCP.

- Interfaz de usuario de la consola de Google Cloud: esta interfaz permite a CARB, cuando cuenta con las credenciales adecuadas, navegar visualmente al depósito de almacenamiento específico que contiene los datos de Nivel 2a y seleccionar carpetas u objetos de datos específicos dentro de ese depósito para descargar.

← Bucket details REFRESH LEARN

bucket-test-1399

Location: us (multiple regions in United States) | Storage class: Standard | Public access: Not public | Protection: None

OBJECTS | CONFIGURATION | PERMISSIONS | PROTECTION | LIFECYCLE | OBSERVABILITY | INVENTORY REPORTS

Buckets > bucket-test-1399

UPLOAD FILES | UPLOAD FOLDER | CREATE FOLDER | TRANSFER DATA | MANAGE HOLDS | DOWNLOAD | DELETE

Filter by name prefix only | Filter Filter objects and folders | Show deleted data

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last modified	Public	Actions
<input type="checkbox"/>	test_data.csv	3.2 KB	text/csv	Dec 21, 2023, 10:09:02 AM	Standard	Dec 21, 2023, 10:09:02 AM	No	Download

Figura 8.4: captura de pantalla de la interfaz de usuario de la consola GCS, que resalta la función de descarga disponible para un conjunto de datos específico dentro de un depósito.

- Bibliotecas de cliente: GCS admite diversos lenguajes para la descarga programática. Por ejemplo, para usar...[Biblioteca cliente de Python](#) CARB haría lo siguiente:
 - a. Instale la biblioteca cliente localmente, si aún no está instalada.
 - b. Configurar la autenticación, que requiere una clave de cuenta de servicio con los permisos adecuados y la configuración de la variable de entorno local apropiada.
 - c. Descargue los datos utilizando funciones de Python desde la biblioteca del cliente, como


```
blob.download_to_filename(nombre_de_archivo_de_destino)
```
- Utilidad de línea de comandos: La interfaz de línea de comandos (CLI) de Google Cloud está disponible a través de un shell de línea de comandos local como Bash, Zsh o Fish. CARB podría usar la CLI para realizar una descarga de la siguiente manera:
 - a. Instalar el [Interfaz de línea de comandos de Google Cloud](#), si aún no está instalado.
 - b. [Autenticar](#) con Google Cloud con una cuenta de usuario o servicio que tenga permisos para acceder al bucket.
 - c. Dada la ruta del archivo al depósito, utilice el `gsutil` Función para descargar uno o más archivos de la ruta del bucket. Varios `gsutil` Los subcomandos se pueden usar para descargar archivos, como `cp` y `sincronización r` Las descargas se pueden reanudar. Al descargar una gran cantidad de archivos, la descarga se puede paralelizar mediante...-m opción.

- d. Una utilidad de línea de comandos alternativa que ofrece una funcionalidad similar a `gsutil` con un rendimiento mejorado es `almacenamiento en la nube`, que también forma parte de la CLI de Google Cloud.

Más información e instrucciones completas sobre el uso de las bibliotecas de cliente y la utilidad de línea de comandos están disponibles a través de [Sitio web de Google Cloud](#).

8.3.3. Seguridad de la vía de transferencia

Aclima utiliza las herramientas de Gestión de Identidad y Acceso (IAM) de Google Cloud y las mejores prácticas de acceso mínimo para proteger todos los datos de monitoreo móvil que recopila. Estas prácticas rigen el acceso a los datos almacenados en Cloud Storage, BigQuery, bases de datos de código abierto que se ejecutan en GCP y cualquier otra infraestructura utilizada para procesar datos.

Para la ruta de integración preferida por CARB a través de Snowflake, Aclima asignará acceso IAM a una cuenta de servicio. Esto permite a CARB cargar una clave de acceso JSON a Snowflake y usar las etapas externas de Snowflake y las herramientas de consulta SQL para importar datos directamente a las instancias de Snowflake.

Según la necesidad de CARB, Aclima puede considerar el uso de una URL firmada para acceso temporal: esto le otorgaría a un usuario de CARB acceso limitado en el tiempo a objetos individuales dentro del depósito de almacenamiento, sin requerir roles de IAM a largo plazo.

Como alternativa, Aclima puede asignar un rol de IAM predefinido a usuarios designados de CARB, probablemente a nivel del Visor de Objetos de Almacenamiento (es decir, acceso de solo lectura a los objetos del bucket). Esto permitirá a Aclima otorgar acceso específico a los permisos del bucket de almacenamiento.

Aclima permitirá el registro de auditoría, el monitoreo y las notificaciones dentro de Google Cloud para rastrear el acceso y las acciones.

9. Conjuntos de datos y visualizaciones de StoryMap

Esta sección describe cómo se utilizarán los conjuntos de datos de Aclima (incluidos los datos finalizados y los fenómenos modelados, es decir, los niveles 2a y superiores) para respaldar las visualizaciones utilizando la herramienta preferida de CARB, ESRI StoryMaps, y cómo se desarrollarán y entregarán estas visualizaciones a CARB.

Los datos de monitoreo móvil recopilados como parte de SMMI tienen como objetivo facilitar acciones focalizadas por parte de las comunidades y CARB, incluido cualquier trabajo futuro para identificar y priorizar ubicaciones para un monitoreo del aire a escala comunitaria más integral, o desarrollar Programas de Reducción de Emisiones Comunitarias (CERPs).

Para respaldar este posible trabajo futuro, Aclima proporcionará un conjunto definido de conjuntos de datos y visualizaciones de StoryMaps asociadas, para los cuales los datos de monitoreo móvil son idóneos y que pueden ayudar a las comunidades a comprender e interpretar mejor los datos en el contexto de sus objetivos. Estos datos estarán disponibles para su consideración por parte del Grupo de Expertos del Proyecto (PEG) y para su uso durante el proceso CAMP.

Estos conjuntos de datos y visualizaciones están diseñados para respaldar a) b) identificar las fuentes y b) identificar los lugares de impacto desproporcionado, como se describe a continuación y se enumera con más detalle en [Tabla 9.1](#).

9.1. Conjuntos de datos para ayudar a identificar fuentes

9.1.1. Conjuntos de datos basados en mejoras

Estos utilizan una combinación de medidas que son *indicativo* de una fuente específica para sugerir la ubicación de la fuente (y, por lo tanto, a menudo se denominan "indicadores"). Aclima define una mejora como una elevación localizada en la concentración de un contaminante que se puede distinguir mediblemente del fondo ambiental. Estos suelen representar la detección de emisiones en forma de una columna coherente en las proximidades de la fuente. Las mejoras se identifican generalmente en los datos con resolución temporal, y las observaciones repetidas de una mejora en una ubicación determinada sugieren cierto grado de persistencia en la señal y, por lo tanto, una mayor confianza en un punto crítico y la fuente subyacente. La caracterización temporal también puede ser un aspecto importante de los análisis basados en mejoras.

Algunos ejemplos de tipos de fuentes que pueden identificarse mediante este tipo de análisis incluyen partículas de diésel, fugas de gas natural y partículas gruesas (PM) procedentes de sitios de construcción u otras instalaciones industriales.

Los conjuntos de datos basados en mejoras se pueden visualizar espacialmente de distintas maneras, desde mapas de calor hasta puntos individuales.

9.1.2. Conjuntos de datos basados en la concentración ambiental

La presencia de gradientes espaciales en las concentraciones ambientales observadas para contaminantes atmosféricos específicos (o combinaciones de diferentes contaminantes medidos) puede indicar la presencia de una fuente y la extensión espacial de las emisiones detectables de dicha fuente. Estos gradientes espaciales pueden observarse en las concentraciones promediadas temporalmente, agregadas a partir de múltiples mediciones, o en series temporales individuales de un mismo recorrido. Mientras que los conjuntos de datos basados en mejoras pueden indicar con mayor precisión la ubicación de una fuente y el grado de impacto de la contaminación en las zonas inmediatamente a sotavento, los conjuntos de datos basados en la concentración ambiental pueden proporcionar una visión espacial más amplia de ciertas fuentes y las zonas que impactan de forma persistente.

Un ejemplo de este tipo de señal es la observación de trazadores de fuentes móviles, como el NO₂o carbono negro en un vecindario colindante con una autopista y la disminución de las concentraciones con la distancia a la autopista. En algunos casos, un solo contaminante puede estar asociado con una fuente específica de preocupación, como el arsénico de las cementeras. Para muchos tipos de fuentes, se utiliza una combinación de mediciones para identificar la fuente, como la combinación de metano y compuestos que contienen azufre que identifican los corrales de engorde.

Los conjuntos de datos basados en la concentración ambiental se pueden visualizar como gradientes en una variedad de agregaciones espaciales, como hexbins o segmentos de carreteras en diferentes niveles de resolución.

9.2. Conjuntos de datos para ayudar a identificar ubicaciones con un impacto desproporcionado

Estos conjuntos de datos se visualizan mejor como gradientes espaciales (mediante diferentes agregaciones espaciales, como hexbins o segmentos de carretera) y marcadores de puntos críticos derivados estadísticamente (puntos geográficos). Identificar ubicaciones con un impacto desproporcionado implica evaluar visualmente la proximidad de los gradientes espaciales y los mercados críticos a una zona geográfica de

interés. Estos conjuntos de datos pueden ser útiles al combinarse con otros conjuntos de datos geospaciales para generar medidas cuantitativas de equidad de impacto.

Tabla 9.1: Descripción y clasificación de conjuntos de datos y visualizaciones apropiadas (posibles) en StoryMaps

Identificar fuentes		
Conjuntos de datos basados en mejoras		Visualizaciones apropiadas
Indicadores ²		
Carbono negro (BC)	Análisis de datos a nivel de bloque de contaminantes individuales para detectar firmas.	Punto, mapa de calor, tabla
P.M _{2.5}		Punto, mapa de calor, tabla
TVOC (contaminante único, sin tipo)		Punto, mapa de calor, tabla
Fuga de gas natural	Indicación clara de la fuente, caracterizada como gas natural, con su localización. Un cronograma probabilístico de cuándo comenzó la fuga y cuándo se resolvió.	Punto, mapa de calor, tabla
Múltiples modalidades de detección de grado de referencia	Los instrumentos de grado de referencia operados por los PML (y, en medida limitada, la flota Aclima) admitirán una variedad de modalidades adicionales como productos de datos indicadores, como TVOC especiados y metales tóxicos.	Punto, mapa de calor, tabla
Identificación de puntos críticos a partir de indicadores	Puntos críticos localizados y resueltos en el tiempo para indicadores etiquetados con la dirección del viento en el momento de ocurrencia.	Punto, mapa de calor, tabla

²Es decir, indica un tipo de fuente particular y se puede utilizar para identificar la ubicación de una fuente.

Conjuntos de datos basados en la concentración ambiental		Visualizaciones apropiadas
PM _{2.5} , NO ₂ , O ₃ , y BC como gradientes espaciales	Promedios agregados geoespacial y temporalmente, que respaldan la identificación visual de gradientes claros desde una fuente particular hasta áreas impactadas	Hexbin (por ejemplo, niveles de hexágono H3 10 y 11) ³); tramos de carretera. Ver figuras 9.1 y 9.2 para ejemplos
Identificación de puntos críticos a partir de concentraciones ambientales para PM _{2.5} , NO ₂ , O ₃ , y antes de Cristo	Puntos críticos persistentes a escala de barrio	Hexbin (por ejemplo, hexágono H3 niveles 10 y 11) o punto

Identificar lugares de impacto desproporcionado		
P.M _{2.5} , NO ₂ , O ₃ , y BC como gradientes espaciales	Estos conjuntos de datos basados en la concentración ambiental, como se describió anteriormente, también son útiles para identificar visualmente ubicaciones de impacto desproporcionado o para realizar trabajos correlativos posteriores con otros conjuntos de datos geoespaciales.	Hexbin (por ejemplo, niveles de hexágono H3 10 y 11) ⁴); segmentos de carretera

³Los hexágonos de nivel H3 10 tienen en promedio ~0,01 km² en área y con longitudes de borde promedio de ~0,075 km; los hexágonos de nivel H3 11 tienen en promedio ~0,002 km² en área y con longitudes de borde promedio de ~0,029 km. Ver [esta referencia](#) para obtener estadísticas más calculadas para hexágonos bajo el sistema H3.

⁴Los hexágonos de nivel 10 H3 tienen en promedio ~0,01 km² en área y con longitudes de borde promedio de ~0,075 km; los hexágonos de nivel H3 11 tienen en promedio ~0,002 km² en área y con longitudes de borde promedio de ~0,029 km. Ver [esta referencia](#) para obtener estadísticas más calculadas para hexágonos bajo el sistema H3.

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)
Iniciativa de monitoreo móvil a nivel estatal

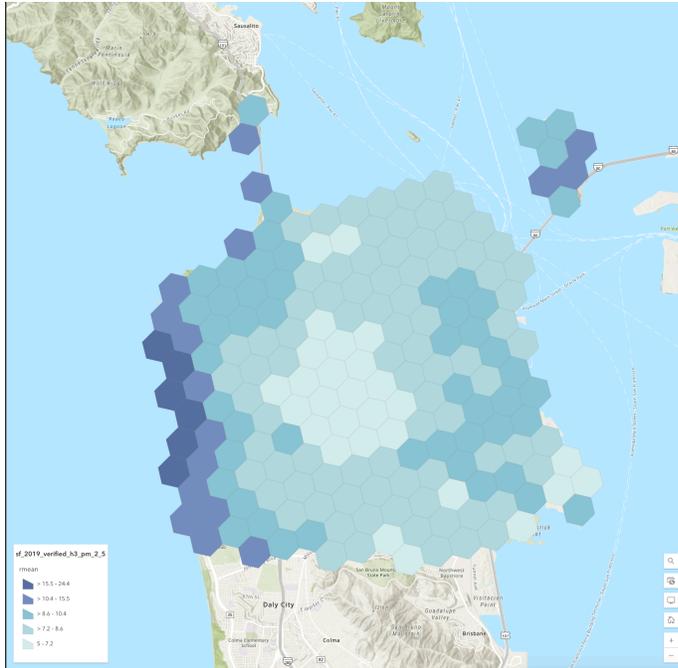


Figura 9.1: Ejemplo de un conjunto de datos de concentración ambiental ($PM_{2.5}$) se grafica mediante hexbins, como un elemento de StoryMaps. Esta visualización se puede realizar con los niveles de datos L3 o L4.

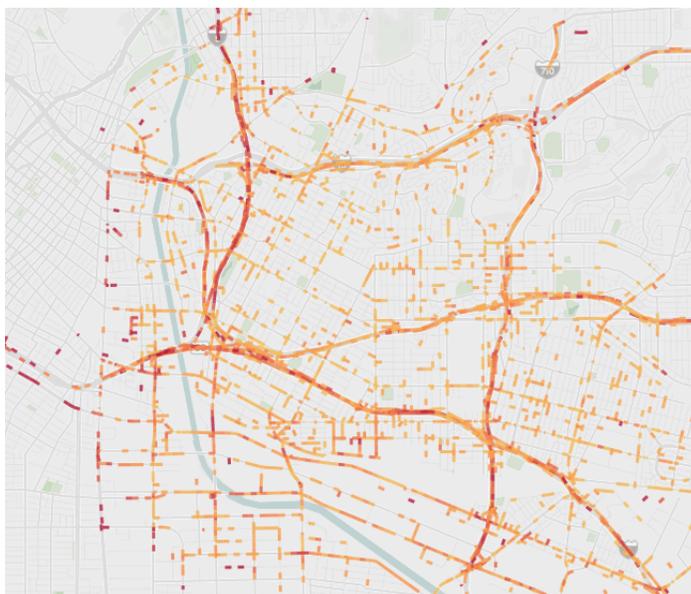


Figura 9.2 Ejemplo de un conjunto de datos basado en mejoras (Multicontaminante: Carbono negro y óxido nítrico) graficado mediante segmentos de carretera. Esta visualización se realiza con el nivel de datos L3.

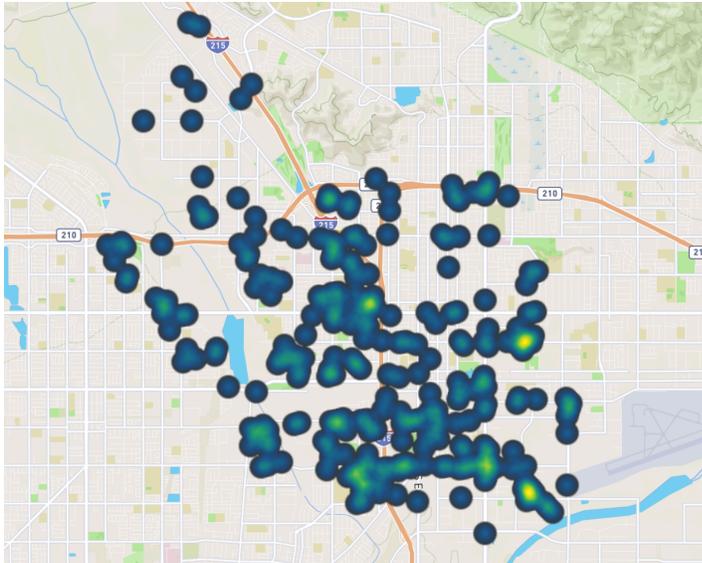


Figura 9.3 Ejemplo de representación gráfica de un conjunto de datos basado en mejoras (TVOC) como mapa de calor. Esta y otras visualizaciones también pueden mostrarse en formato tabulado. Esta visualización se realiza con el nivel de datos L3.



Figura 9.4 Ejemplo de representación gráfica de un conjunto de datos mediante una visualización de puntos (eventos de mejora de TVOC). Esta visualización se realiza con el nivel de datos L3.

9.3. Enfoque de visualización de conjuntos de datos

Para que estos conjuntos de datos sean útiles para las comunidades, es necesario visualizarlos de forma accesible. CARB ha seleccionado ESRI StoryMaps como plataforma de visualización. Puede haber hasta 64 StoryMaps individuales, dependiendo de las decisiones de PEG y CAMP. Para que esto sea posible, dadas las limitaciones de tiempo y

recursos, Aclima propone adoptar un enfoque simplificado basado en plantillas para la creación de estas visualizaciones. Esto garantizará que los conjuntos de datos se presenten únicamente en formatos adecuados y que se disponga de una experiencia común para todas las comunidades, a la vez que se conserva la posibilidad de adaptarlos a las necesidades de la comunidad, PEG y CARB. Cada visualización se implementará utilizando las herramientas de StoryMaps, idealmente en cuentas alojadas por CARB, para que la propiedad se pueda transferir sin problemas.

Aclima propone dos componentes principales disponibles para su inclusión en el StoryMap de cualquier comunidad:

1. Una plantilla de StoryMap común disponible para todas las comunidades:
 - a. Explica los antecedentes del proyecto y el contexto científico a través de elementos narrativos.
 - b. Visualiza todos los conjuntos de datos basados en la concentración ambiental utilizando diferentes elementos espaciales de StoryMap
 - c. Proporciona una estructura estándar para el flujo de StoryMap
2. Una capa de datos opcional para visualizar conjuntos de datos basados en mejoras:
 - a. Visualiza datos provenientes de sensores Aclima o instrumentos PML
 - b. Disponible para integración con cualquier elemento espacial existente, como un elemento espacial separado o en forma tabulada

Estos componentes se describen con más detalle en las siguientes secciones.

9.3.1. Plantilla común de StoryMap

Aclima proporcionará una plantilla común para todas las instancias de StoryMap de todas las comunidades. Esta plantilla incluirá:

1. **Módulos narrativos** que presentan el proyecto, brindan información de fondo, comparten el contexto científico (por ejemplo, información sobre contaminantes monitoreados, diseño de muestreo para monitoreo móvil, preguntas y respuestas, etc.) y pueden intercalarse con visualizaciones espaciales o tabuladas específicas para brindar información más detallada.
2. A **superposición de gradiente espacial**(agregados especialmente en hexágonos H3 y temporalmente a lo largo de todo el período de monitoreo), sobre uno o más mapas base estándar. Esto proporcionará acceso principalmente a conjuntos de datos basados en la concentración ambiental. Puede estar disponible como elemento cartográfico independiente o como capa base en mapas más complejos.
3. A **superposición de puntos** Datos L2a de todos los contaminantes detectados por los sensores Aclima, es decir, cada medición individual como un único punto

geográfico. Normalmente, esto se superpone a una capa de gradiente espacial, pero también está disponible como elemento de mapa independiente.

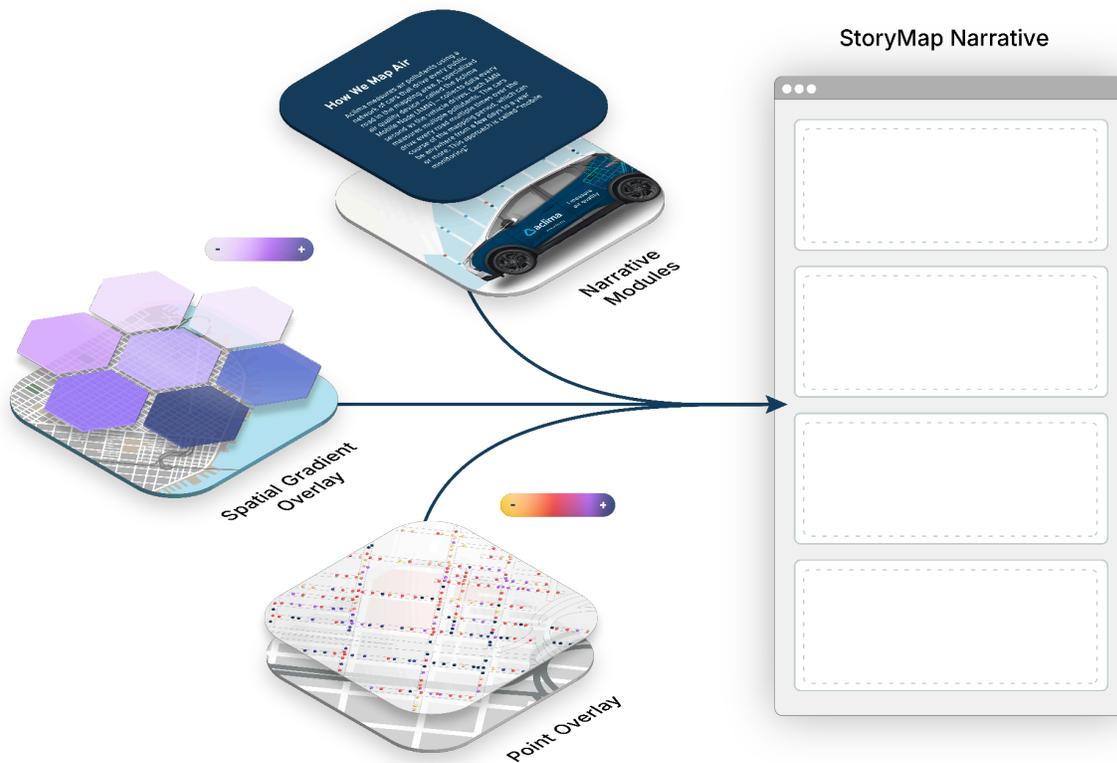


Figura 9.5: Resumen del diseño de la plantilla StoryMap. Los módulos narrativos, la superposición de gradientes espaciales y las superposiciones de puntos se entrelazan y combinan para construir una narrativa lineal de StoryMap.

9.3.2. Capa de datos basada en mejoras, derivada de los sensores Aclima

Aclima trazará conjuntos de datos basados en mejoras (productos de datos "indicadores", como carbono negro o $PM_{2.5}$; ver [Tabla 9.1](#)(arriba) para contaminantes detectados con las Plataformas Móviles Aclima (AMP) mediante un diseño de muestra de conducción de área amplia, como superposición sobre el mapa base. Los datos solo serán visibles con niveles de zoom más altos y podrán presentarse como datos puntuales, concentraciones de segmentos de carretera o una visualización de picos en 3D, según el contaminante y la entrada PEG/CAMP. Esta superposición incluirá contenido narrativo para proporcionar contexto y orientación.

Esta superposición estará disponible para todas las visualizaciones de todas las comunidades, si se detectan indicadores basados en mejoras.

9.3.3. Capa de datos basada en mejoras, derivada de instrumentos PML

Aclima representará conjuntos de datos basados en mejoras para contaminantes detectados con instrumentos de referencia PML mediante una variedad de diseños de muestreo de áreas específicas, como una superposición sobre el mapa base. Estos datos solo serán visibles con niveles de zoom más altos y podrán presentarse como datos puntuales, concentraciones en segmentos de carretera o una visualización de picos en 3D, según el contaminante y la entrada PEG/CAMP. Esta superposición se acompañará de contenido narrativo para proporcionar contexto y orientación, adaptados a cada comunidad específica.

Dado que la recopilación de datos de PML utiliza diseños de muestreo por área específica que se basan en diversas fuentes objetivo (p. ej., datos de la flota de Aclima; fuentes conocidas; inquietudes de la comunidad, etc.), y con una capacidad limitada para campañas ejecutadas con estos diseños, las superposiciones de este tipo serán exclusivas de comunidades específicas, no estarán disponibles para todas las comunidades y podrían no distribuirse uniformemente entre ellas. Aclima prevé entre 15 y 20 campañas de muestreo específicas que podrían distribuirse en cualquier parte del área del proyecto y, por lo tanto, entre 15 y 20 superposiciones diferentes de este tipo.

Plantillas para superposiciones de conjuntos de datos derivados de PML

Para los datos recopilados por los PML *solo* Aclima y los PML han desarrollado una serie de plantillas de Planes de Muestreo, Análisis y Presentación (SAPP) que guiarán la recopilación de datos de PML e influirán en la visualización del contenido de StoryMap generado por PML. Un único SAPP:

- Capta una preocupación típica de la comunidad
- Enmarca esto como una pregunta científica o un objetivo de seguimiento.
- Identifica los contaminantes implicados y los tipos de fuentes
- Sugiere un diseño de muestreo apropiado
- Identifica los análisis que se ejecutarán en los datos recopilados
- Sugiere resultados estadísticos y/o visuales apropiados
- Indica posibles acciones regulatorias solo como referencia

En el siguiente documento se resume un conjunto inicial de preocupaciones de la comunidad que se están considerando en forma de borrador: [Tabla 9.2](#), y el contenido del borrador para una preocupación específica se describe en [Tabla 9.3](#). Los SAPP se finalizarán mediante aportes adicionales de PEG, CAMP y CARB.

Tabla 9.2: Resumen de las inquietudes de la comunidad en una lista preliminar de plantillas del Plan de Muestreo, Análisis y Presentación (SAPP) para respaldar los debates de PEG y CAMP y, en última instancia, guiar la presentación de fuentes de PML (y contaminantes específicos provenientes de áreas objetivo impulsadas por AMP) en StoryMaps

SAPP	Preocupación de la comunidad
1	¿Las refinerías cercanas están afectando mi vecindario?
2	Hay una planta de almacenamiento de combustible en mi comunidad. ¿Debería preocuparme por el benceno u otros contaminantes en mi vecindario?
3	¿Las instalaciones están emitiendo contaminantes distintos a los que permite su permiso?
4	¿Cuándo las fuentes impactan mi vecindario?
5	Mi barrio fue objeto de discriminación. ¿Sigue sufriendo un impacto desproporcionado?
6	¿Las autopistas están afectando mi vecindario, especialmente las rampas de entrada y salida?
7	En mi comunidad hay un almacén donde se almacena equipo médico esterilizado. ¿Debería preocuparme por el óxido de etileno (EtO)?
8	Veo polvo que se aleja de una cementera en mi barrio. ¿Hay algo tóxico en el polvo que deba preocuparme?
9	Hay tantas fuentes potenciales de contaminación en mi vecindario. ¿Cómo puedo saber cuáles son las más peligrosas y cuáles deben priorizarse?
10	¿Cómo afecta el tráfico de camiones a mi comunidad?

Tabla 9.3: Contenido del borrador del SAPP n.º 7, relacionado con las preocupaciones sobre el óxido de etileno (EtO)

Preocupación de la comunidad	En mi comunidad hay un almacén donde se almacena equipo médico esterilizado. ¿Debería preocuparme por el óxido de etileno (EtO)?
Posibles preguntas científicas	<ul style="list-style-type: none"> • ¿Existe una mejora de la persistencia del EtO en las proximidades del almacén? • ¿Con qué frecuencia se encuentra y hasta dónde se propaga? • ¿Cuál es el trasfondo del EtO en ausencia de fuente? • ¿Qué eventos se correlacionan con la detección de una columna de preocupación? • ¿En qué otro lugar de la comunidad hay

	preocupaciones similares?
Contaminante implicado	Póngase en fila
Tipos de fuentes implicadas	<ul style="list-style-type: none"> ● Instalaciones petroquímicas ● Proveedores de gas comprimido ● Instalaciones de esterilización
Diseño de muestreo	<ul style="list-style-type: none"> ● Implementar instrumentos PML (TILDAS, AROMA-ETO) ● Conducir dentro y fuera de la columna ● Permanecer estacionario en la columna ● Conducir transectos a favor y en contra del viento
Análisis	<ul style="list-style-type: none"> ● Concentración media y máxima de EtO en la columna ● Derivar la frecuencia de detecciones de EtO ● Calcular la concentración media de fondo de EtO en ausencia de columna ● Estimaciones de la concentración ambiental de EtO en zonas a sotavento
Salidas	<ul style="list-style-type: none"> ● Mapa de la extensión de la columna utilizando datos de 1 Hz ● Mapas de calor alrededor de una instalación ● Ubicación de la fuente ● Frecuencia visualizada de detección ● Superposiciones de dirección y fuerza del viento ● Histogramas con superposiciones de límites agudos y de largo plazo ● Gráfico de barras y bigotes para promedios estacionarios o de vecindad
Acciones regulatorias	<ul style="list-style-type: none"> ● Iniciar un estudio o inspección sobre las instalaciones en cuestión

9.4. Suministro de datos y transferencia de propiedad de StoryMaps

Aclima desarrollará todos los StoryMaps y todos los conjuntos de datos y productos de datos de apoyo, según una lista final desarrollada a través de consultas con PEG, CAMP y CARB, dentro de los límites de razonabilidad y viabilidad.

Aclima pondrá a disposición de CARB los conjuntos de datos y productos de soporte en un bucket GCS, para que CARB pueda acceder a ellos y transferirlos a Snowflake o a los servidores Esri de CARB. Las visualizaciones de StoryMap serán diseñadas y

Plan de Monitoreo del Aire de la Comunidad: Apéndice F (v3.0)

Iniciativa de monitoreo móvil a nivel estatal

desarrolladas por Aclima, ya sea mediante cuentas Esri dedicadas proporcionadas por CARB o en una cuenta Esri proporcionada por Aclima, que posteriormente serán transferidas a CARB. Aclima recomienda que los miembros de su equipo tengan acceso a la cuenta Esri de CARB para optimizar este trabajo y reducir los riesgos posteriores al contrato.

Aclima no instalará ni gestionará el alojamiento de StoryMap durante ni después del período del contrato. Todos los productos estarán disponibles y se podrán mantener una vez finalizado el período del contrato (Aclima no tendrá ninguna otra participación después de esa fecha).

Aclima entiende que CARB no le proporcionará ninguna vía directa al producto AQview y no exige que Aclima realice ningún trabajo específico en AQview.