

# Proyecto TRAFair: Generación y publicación de datos de calidad del aire en las ciudades de Zaragoza y Santiago de Compostela<sup>\*</sup>

José R.R. Viqueira<sup>1</sup>, Raquel Trillo-Lado<sup>2</sup>, Sebastián Villarroya<sup>1</sup>, Lorena Marrodán<sup>2</sup>, José M. Cotos<sup>1</sup>, Sergio Ilarri<sup>2</sup>, José A. Taboada<sup>1</sup>, and Enrique Torres-Moreno<sup>2</sup>

<sup>1</sup> Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS),  
Universidade de Santiago de Compostela (USC),  
15782, Santiago de Compostela, Spain  
{jrr.viqueira, sebastian.villarroya, manel.cotos,  
joseangel.taboada}@usc.es

<sup>2</sup> Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza  
{raquel1, marrodan, silarri, ktm}@unizar.es

**Resumen** En este artículo se describen brevemente los trabajos en marcha relacionados con la generación y publicación de datos acerca de la calidad del aire en el ámbito del proyecto Europeo TRAFair. En concreto, se describe la solución adoptada para la adquisición de datos de sensores, los estándares utilizados para la publicación de datos en abierto y las aplicaciones de usuario final que serán desarrolladas, concluyendo el artículo con la identificación de retos técnicos relacionados con la heterogeneidad de los datos y con la generalización de soluciones basada en la asunción de modelos de datos estandarizados.

**Keywords:** Datos de sensores, Datos abiertos, Datos geoespaciales, Modelado ambiental, Calidad del aire

## 1. Introducción

La contaminación atmosférica es, hoy en día, un problema de salud pública de primer orden, que genera gran preocupación en la ciudadanía. La mala calidad del aire es ya una de las principales causas de mortalidad prematura en la Unión Europea. Por ello, en 2013, la Comisión Europea puso en marcha el paquete Aire Puro<sup>3</sup>, que incluye políticas específicas para que los estados miembros garanticen

---

<sup>\*</sup> Este trabajo ha sido financiado por la Comisión Europea a través del programa CEF-TELECOM de INEA (INEA/CEF/ICT/A2017/1566782), el proyecto TIN2016-78011-C4-3-R (AEI/FEDER, UE), el Gobierno de Aragón (Referencia Grupo T35 17D, grupo COSMOS) cofinanciado con Feder 2014-2020 “Construyendo Europa desde Aragón” y la Consellería de Educación, Universidade e Formación Profesional de la Xunta de Galicia (ED431B 2018/28).

<sup>3</sup> <https://www.consilium.europa.eu/es/policies/clean-air/>.

una buena calidad del aire para los ciudadanos, y establece límites de emisión para compuestos como el  $\text{NO}_2$ , procedente en gran parte del tráfico.

El objetivo general del proyecto TRAF AIR (Understanding Traffic Flows to Improve Air Quality) es el desarrollo de un servicio que permita a los ciudadanos y autoridades municipales analizar el impacto que distintas condiciones sobre el tráfico tienen en la calidad del aire. El proyecto, iniciado en noviembre de 2018, se está ejecutando en 6 ciudades de Italia y España: Zaragoza, Florencia, Módena, Livorno, Santiago de Compostela y Pisa.

Para conseguir el objetivo descrito arriba, las actividades definidas en el plan de trabajo del proyecto apuntan al diseño e implementación de distintos componentes de este servicio. En concreto, se está diseñando un almacén de datos de entrada, que debe incorporar tanto los datos generados por los sistemas de sensorización de tráfico y calidad del aire, como el resto de datos de entrada necesarios en los diferentes modelos considerados. Se está diseñando e implementando un sistema de adquisición de datos para los sensores de calidad del aire. Se ejecutará un modelo de estimación de los flujos de tráfico a partir de sensores, un modelo de estimación de generación de polución a partir del tráfico, un modelo de dispersión de contaminantes y un modelo de estimación de polución en tiempo real a partir de las medidas de los sensores. Los datos de los sensores y los generados por los modelos serán publicados a través de los estándares apropiados en portales de datos abiertos de las administraciones involucradas, que serán recolectados por el Portal de Datos Europeo. Finalmente, se implementarán algunas aplicaciones de usuario final tanto para los ciudadanos como para las diferentes administraciones públicas involucradas en el proyecto.

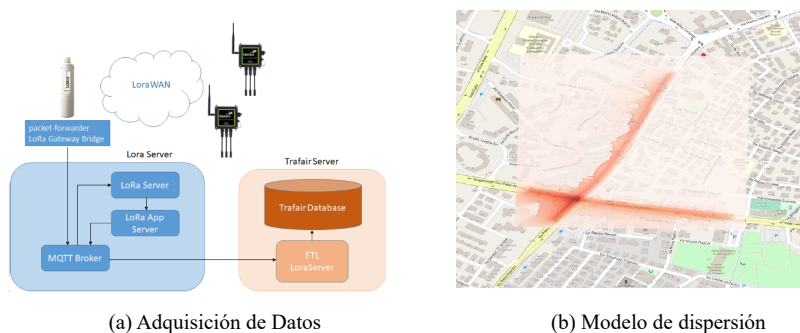
En la Sección 2 se describe brevemente el sistema de adquisición actualmente en uso. La descripción de los estándares y soluciones adoptadas inicialmente para la publicación de datos en abierto se describe en la Sección 3. La Sección 4 proporciona información sobre las aplicaciones a desarrollar y el artículo concluye, en la Sección 5, con una breve discusión sobre algunos retos a abordar.

## 2. Generación de datos

Los datos gestionados y publicados en el ámbito del proyecto provienen de tres grandes tipos de fuentes: i) datos de entrada a los modelos considerados, ii) datos de sensorización y iii) resultados obtenidos tras la ejecución de los modelos.

Los datos de entrada a los modelos incluyen datos de tipo convencional, como la descripción de la flota de vehículos del municipio, datos de tipo geoespacial, como la descripción de la red de calles y de la edificación, y datos de tipo científico, como la predicción meteorológica. Estos datos no se generan en el ámbito del proyecto, y se incorporan al almacén directamente de fuentes externas.

Los datos de sensorización incluyen medidas de sensores de tráfico, cuyo sistema de adquisición está fuera del ámbito del proyecto, y medidas de observación de calidad del aire, generadas por la infraestructura de sensorización del proyecto. La arquitectura de la primera versión del sistema de adquisición desplegado por la Universidade de Santiago de Compostela puede verse en la Fig. 1(a). Los



**Figura 1.** Generación de datos.

sensores utilizan una red LoRaWAN para enviar valores de voltaje relacionados con las concentraciones de CO, NO, NO<sub>2</sub> y O<sub>3</sub> a un Gateway. Este Gateway envía las tramas de datos a un broker MQTT, para que puedan ser accedidas por un servidor de red LoRaWAN y una vez procesadas sean enviadas a un servidor de aplicaciones. Dado que ya se dispone en la arquitectura de un broker MQTT, este método de transferencia es utilizado para enviar los datos al proceso ETL que carga los datos en la base de datos de TRAF AIR.

En TRAF AIR se usan diferentes modelos para generar datos de distinto tipo. Los datos de los sensores de tráfico son utilizados por un modelo de flujo de tráfico para generar series temporales de valores de flujo de tráfico vinculados a cada sección de la red de calles de la ciudad. Estos datos se combinan con los datos descriptivos de la flota de vehículos, para estimar los valores de concentración de los distintos gases en cada sección de calle. En el siguiente paso, un modelo de dispersión combina los datos anteriores de contaminación debidos al tráfico, con datos de predicción meteorológica y con datos de la geometría de la edificación para generar una predicción de la concentración de cada gas en cada punto de la ciudad para cada hora del siguiente día. La Fig. 2(b) proporciona un ejemplo de dicha predicción para una zona de la ciudad de Módena en un instante concreto.

### 3. Datos abiertos

Uno de los objetivos de TRAF AIR es la publicación en abierto de datos de observación y predicción de calidad del aire en las 6 ciudades, a través del portal de datos Europeo. Para ello se seleccionarán estándares relevantes tanto relacionados con la directiva INSPIRE para la creación de la Infraestructura de Datos Espaciales (IDE) de Europa, como con la Web de Datos. En esta primera fase del proyecto se tendrán en cuenta estándares propuestos por el Open Geospatial Consortium (OGC) y adoptados, o en vía de adopción, por INSPIRE. En concreto, para el acceso a metadatos se considerarán los estándares Catalog Service del OGC y Data Catalog Vocabulary del W3C. Para los datos geoespaciales, tanto vectoriales como raster, se usarán los estándares de servicio

Web Map Service, Web Feature Service y Web Coverage Service del OGC. Por último, para el acceso a datos de sensores, se usará el estándar SOS (Sensor Observation Service) de la iniciativa Sensor Web Enablement del OGC, que utiliza para la codificación de datos el modelo Observations and Measurements, incorporado también en la ontología Semantic Sensor Network del W3C.

## 4. Aplicaciones

Las aplicaciones a desarrollar destinadas a los ciudadanos tienen como objetivo aumentar la conciencia pública sobre la contaminación del aire en las zonas urbanas, mientras que las aplicaciones dirigidas a las administraciones públicas persiguen ofrecer sistemas de monitorización y simulación de la calidad del aire y tráfico para facilitar la definición de estrategias y medidas a implantar relativas a la gestión del tráfico. En concreto, se desarrollarán móviles dirigidas a los ciudadanos que permitirán obtener alertas y recomendaciones personalizadas relativas al estado de la calidad del aire y analizar rutas específicas dentro de la ciudad indicando: la estimación de la exposición a contaminantes, su huella de carbono y cómo podría contribuir a disminuirla. Estas aplicaciones también incorporarán un mecanismo de retroalimentación permitiendo a los usuarios informar sobre la conveniencia y precisión de las medidas y recomendaciones recibidas. Respecto a las aplicaciones dirigidas a la administración, además de las herramientas de monitorización se proporcionarán sistemas para el análisis y la estimación de la difusión de contaminantes en el área urbana. Estas herramientas también permitirán simular nuevas hipótesis de circulación como, por ejemplo, cambios en la flota urbana de vehículos (aumento de vehículos con bajas emisiones, transporte público, etc.) y estudiar sus impactos en la calidad del aire.

## 5. Retos de investigación

Durante la ejecución de TRAFAIR surgen retos tecnológicos que demandan soluciones más allá del estado actual de la tecnología en diversos ámbitos. A continuación, se muestran algunos ejemplos relacionados con la gestión de datos.

Un primer reto es el análisis integrado de datos en almacenes heterogéneos de datos geoespaciales, es decir, almacenes de datos que combinan series temporales (generadas por sensores y algunos modelos), datos geoespaciales de tipo vectorial (red de calles, edificación, etc.) y series temporales de coberturas geoespaciales raster (generadas por modelos). Soluciones iniciales de análisis integrado de datos vectoriales y raster ya han sido propuestas [5,4]. Sin embargo, su rendimiento todavía está lejos del deseado, sobre todo en entornos de exploración interactiva.

Un segundo reto está relacionado con la adopción de un metamodelo basado en el estándar Observations and Measurements para habilitar la adopción de soluciones de ingeniería basadas en modelos para el desarrollo de componentes de la arquitectura de adquisición, almacenamiento y publicación de datos en abierto, que puedan incorporar capacidades de integración semántica de datos [3].

Por último, el tercer reto que se plantea es el empleo de técnicas de minería de procesos para detectar nuevas tendencias de uso en las aplicaciones desarrolladas no previstas en su diseño [2]. Además, estas técnicas y variantes de la restricción de coetaneidad [1] se emplearán para determinar la calidad de datos proporcionados por los ciudadanos, en concreto, para analizar la reputación o confianza de los datos proporcionados por cada uno de los ciudadanos.

## Referencias

1. Abián, D., Bernad, J., Trillo-Lado, R.: Using contemporary constraints to ensure data consistency. In: Actas de la 34th ACM Symposium on Applied Computing, SAC 2019, Limassol, Chipre, 08-12 abril (2019)
2. Bernardi, S., Alastuey, R.P., Bonilla, A.S., Trillo-Lado, R.: Towards a systematic model-driven approach for the detection of web threats and use cases. In: Actas del Int. WS on Petri Nets and SW Engineering, en la 38th Int. Conf. on Application and Theory of Petri Nets, Zaragoza, España, 25-30 de junio. pp. 225–226 (2017)
3. Regueiro, M.A., Viqueira, J.R., Stasch, C., Taboada, J.A.: Semantic mediation of observation datasets through sensor observation services. *Future Generation Computer Systems* 67, 47 – 56 (2017)
4. Villarroya, S.: Acquisition and Declarative Analytical Processing of Spatio-Temporal Observation Data. Ph.D. thesis (2018)
5. Villarroya, S., Viqueira, J.R., Regueiro, M.A., Taboada, J.A., Cotos, J.M.: Soda: A framework for spatial observation data analysis. *Distrib. Parallel Databases* 34(1), 65–99 (Mar 2016)