

OSMazinger

Un motor de cálculo de indicadores geográficos para OpenStreetMap y otras fuentes de datos



Javier Paniagua es egresado del DEI, UC (Ing. Informática) y PhD en ICT del DISI, Universidad de Trento.

Ha participado en varios proyectos de los programas FP7 y H2020 de la Unión Europea en las áreas de multimedia, GIS, NLP y Knowledge Representation. Actualmente trabaja en R&I en SpazioDati, una joven empresa italiana especializada en integración, análisis e interpretación de datos empresariales y es Lead Engineer y Project Manager para proyectos de investigación de la Comunidad Europea, y contraparte UniTrento por el proyecto SmartTraffic de Conacyt.

Hoy Lunes 25 de Marzo – 16:00hs – Sala Manfred Stark



Contenido

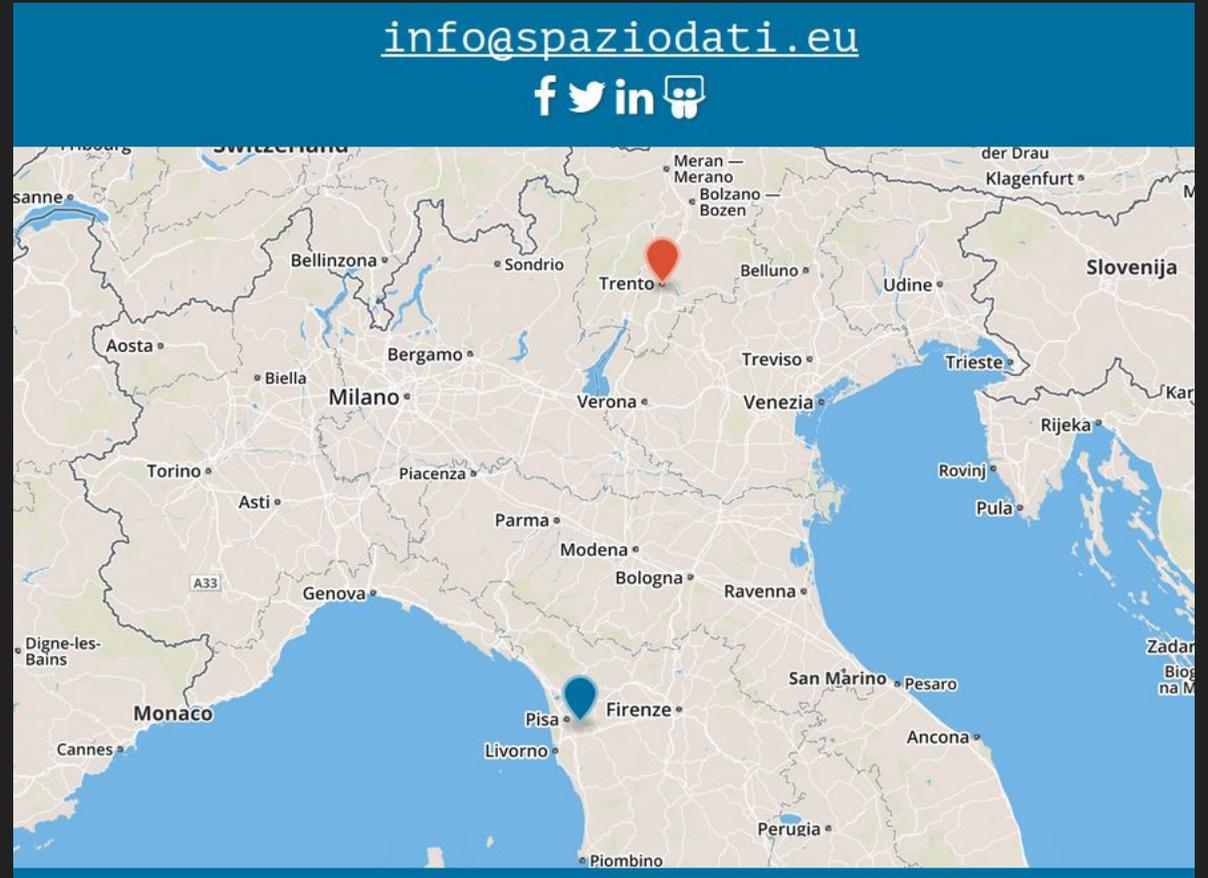
- Bio
- El contexto de OSMazing
- OSMazing
 - Indicadores geográficos
 - Pipeline
 - Desafíos
 - Caso de Uso: estimar el valor de las propiedades
- Shameless plug: Proyecto SmartTraffic

Bio

- Egresado de la UC Asunción en el 2005 (Ing. Informática)
- Algunos años de experiencia freelance
- Sistemas de tasación para telefonía celular
- Pasantía + PhD en UniTN en 2015 (ICT: personal photography)
- R+I en SpazioDati
- Contraparte UniTn en el Proyecto SmartTraffic de CONACYT

All you need is lo^H^HDATA

- Una (ex)start-up, operativa desde 2012
- ~30 personas en 2 sedes en Italia (Trento y Pisa)
- Especialista en (big)data collection/integration/interpretation
- “Business intelligence”



El contexto de OSMazinger: Atoka

El directorio más grande de información empresarial en Italia:

- IT: 6M de empresas, RU: 4,5M, UK: 4,1M
- Info freshness: 10m - 5d

Key value proposition:

Integrar información oficial *y no oficial* sobre las empresas

- [Demo time](#)



OSMazinger: Indicadores ¿qué son?

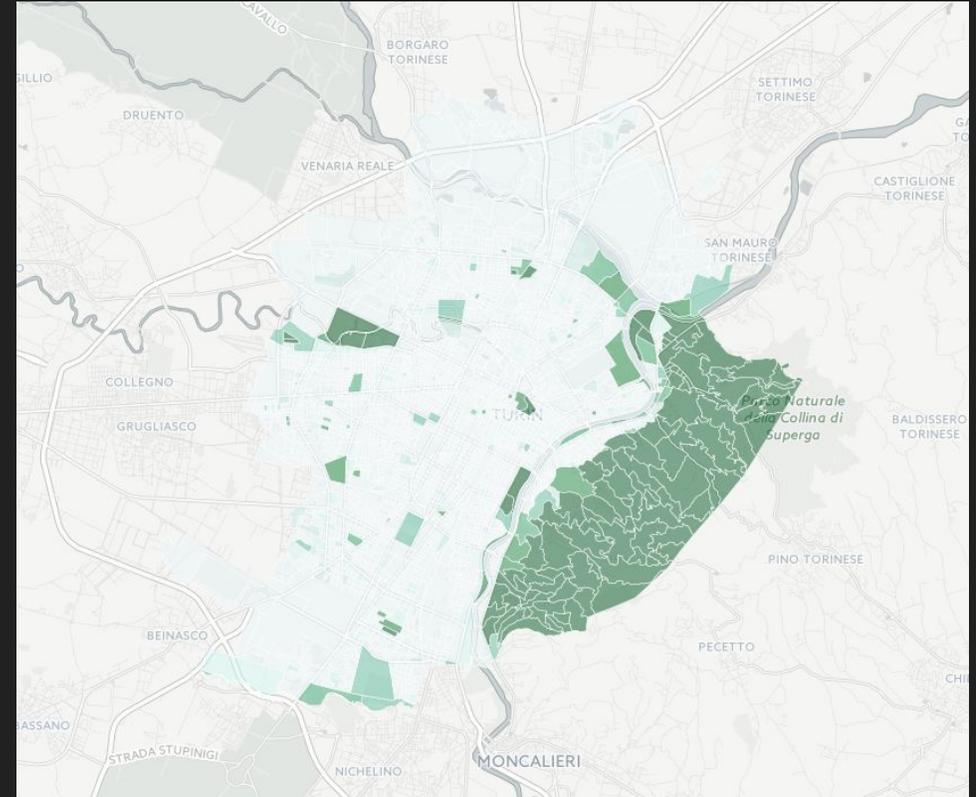
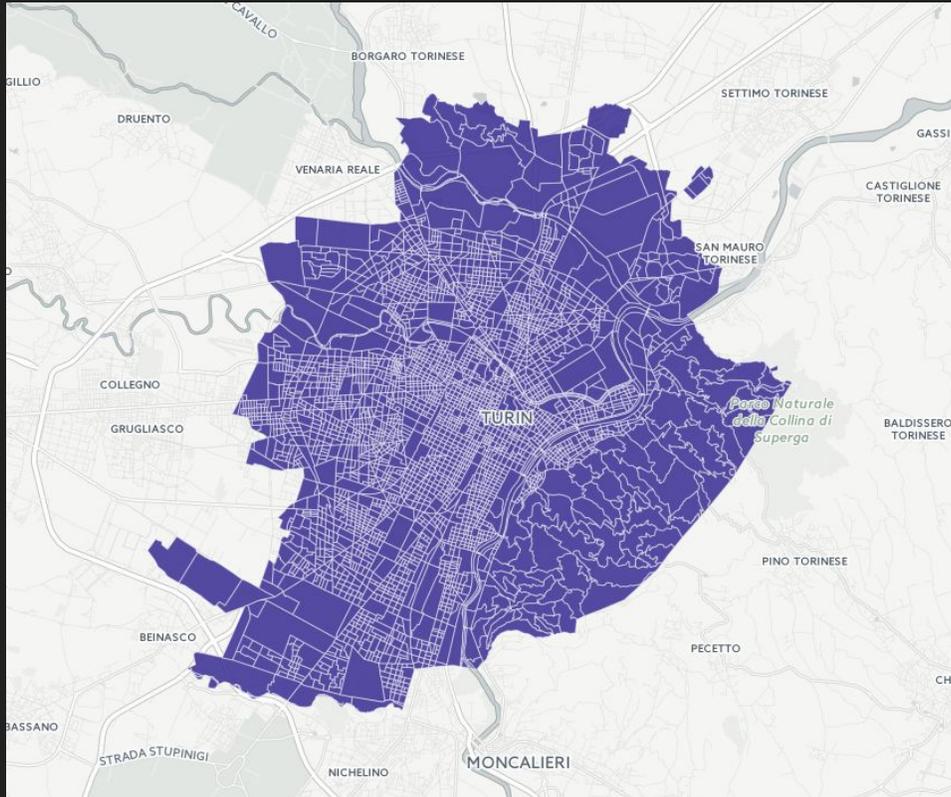
Un (geo-)indicador es una descripción numérica de algún fenómeno en un área geográfica

Ejemplos:

- Cantidad de “kindergartens” cercanos
- Distancia al hospital más cercano
- Porcentaje de áreas verdes del lugar

OSMazinger: Indicadores ¿qué son?

Ejemplo: la ciudad de Turín descrita según el porcentaje de áreas verdes en cada zona



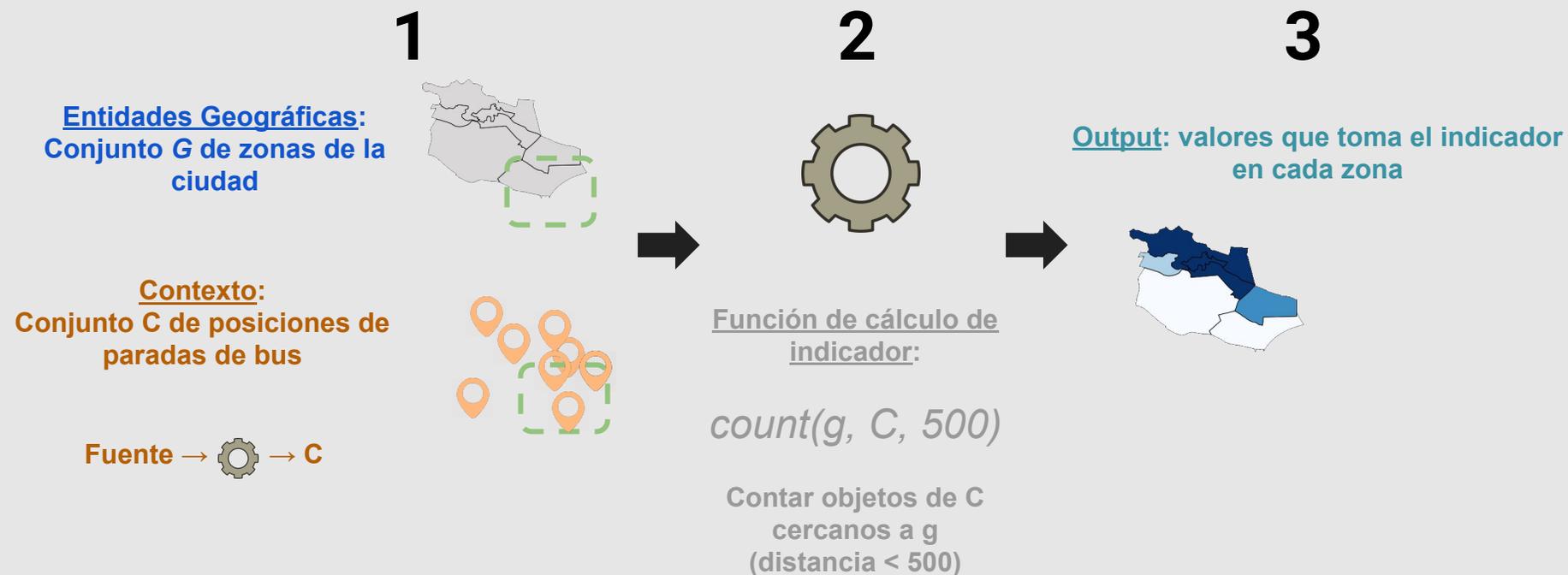
OSMazinger: elementos a definir

Motor de cálculo de indicadores geográficos donde se define:

- Conjunto G de entidades geográficas a describir por medio del indicador
- Conjunto C de entidades geográficas de contexto que ayudarán a describir el indicador.
 - Un mapeo entre vocabulario de la fuente y C
 - Un algoritmo de transformación para eliminar inconsistencias y duplicados
- $f(g, C) = i_g$ que relaciona una entidad $g \in G$ con el contexto C para dar el valor i_g del indicador i para la entidad g

OSMazinger: pipeline

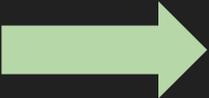
Objetivo: calcular # de paradas de bus cercanas en cada **zona de una ciudad**



OSMazinger: desafíos

- Heterogeneidad semántica
 - Alineación del vocabulario de la fuente con el de la aplicación deseada
 - Variación de expresividad de lugar a lugar
- Heterogeneidad estructural
 - Variación de la riqueza de la representación para un mismo tipo de entidad geográfica
- Datos sucios:
 - Existencia de duplicados

OSMazinger: heterogeneidad semántica

escuelas  [amenity=school] Simple!

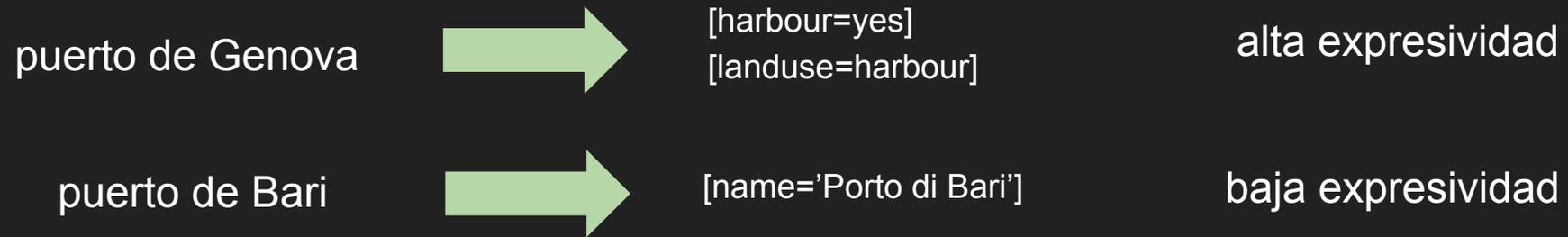
aeropuertos 

```
[
  "aeroway=aerodrome][iata][aerodrome:type~'public',i",
  "aeroway=aerodrome][iata][aerodrome:type~'joint',i",
  "aeroway=aerodrome][iata][type~'public',i",
  "aeroway=aerodrome][iata][type~'joint',i",
  "aeroway=aerodrome][iata][aerodrome~'public',i",
  "aeroway=aerodrome][iata][aerodrome~'joint',i",
  "aeroway=aerodrome][iata][aerodrome~'international',i",
]
```

 No tan simple

Solución: soportar el mapeo entre vocabulario origen y vocabulario destino

OSMazinger: heterogeneidad semántica

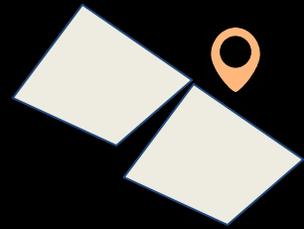


Solución: location-aware queries

OSMazinger: heterogeneidad estructural

Problema:

- Kindergartens a veces representados como puntos y a veces representados como polígonos. ¿Cómo uniformar el criterio de representación?



nodos y polígonos → “centroidify” → sólo nodos

Solución:

Proveer funciones de uniformación de representación

OSMazinger: datos sucios

Problema:

- El mismo kindergarten puede estar representado como nodo y como polígono. ¿Cómo evitar contarlo varias veces en el indicador “número de kinders cercanos”?



2 annotations → “dedup-by-vicinity” → 1 entity

Solución:

Proveer funciones de deduplicación

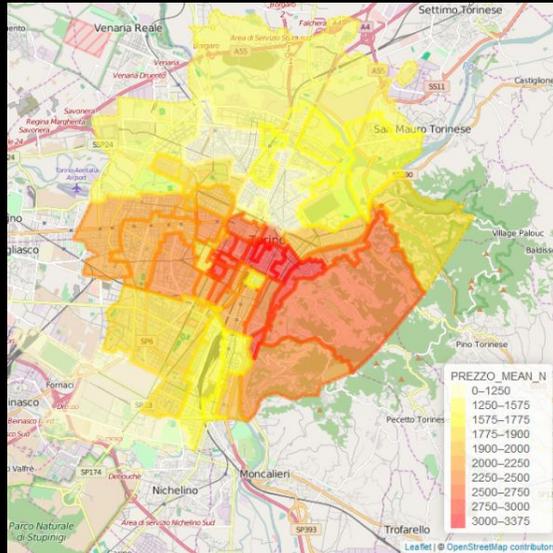
OSMazinger: recap

Motor de cálculo de indicadores geográficos

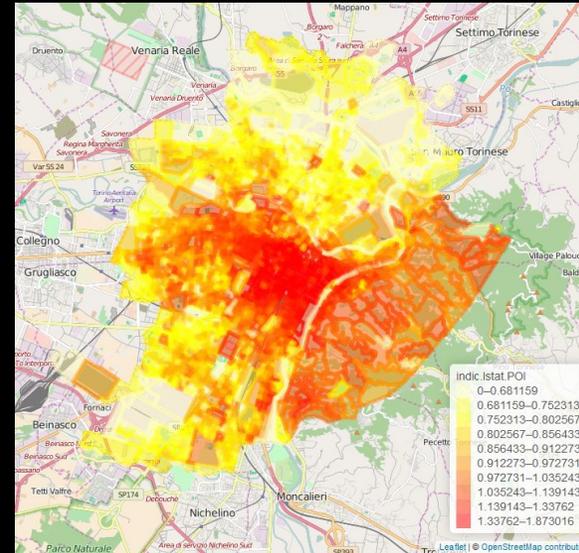
- Conjunto de elementos geográficos G a describir
- Conjunto de elementos contextuales C
 - Mapeo de vocabularios
 - Funciones de limpieza e uniformación
- Función de cálculo de indicador

OSMazinger: estimación de precios de inmuebles

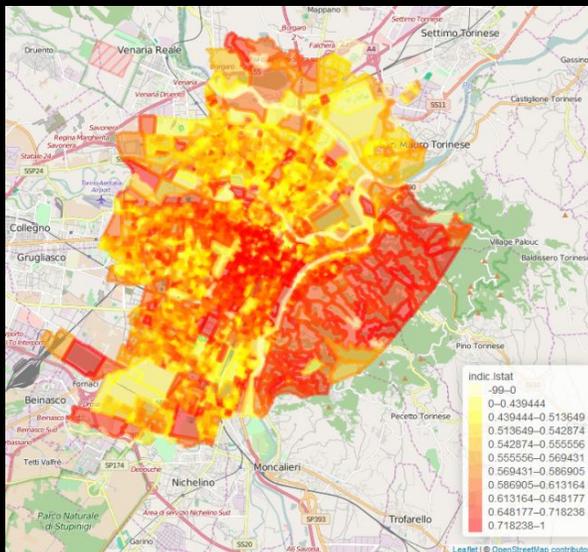
Datos oficiales por zonas de catastro



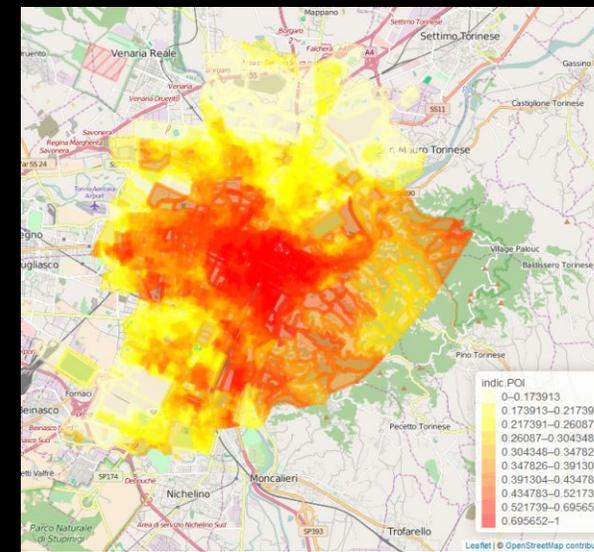
Ponderado OSMazinger e Istat



Estimación con datos oficiales Istat

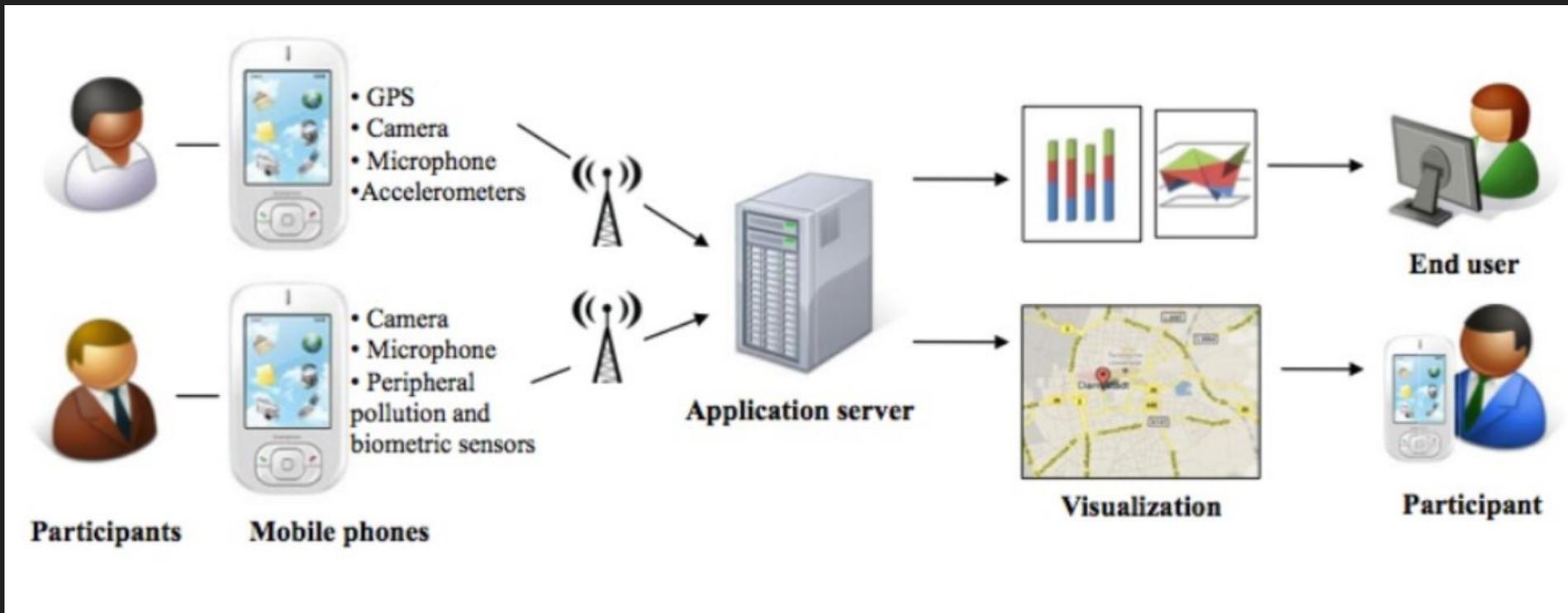


Estimación OSMazinger

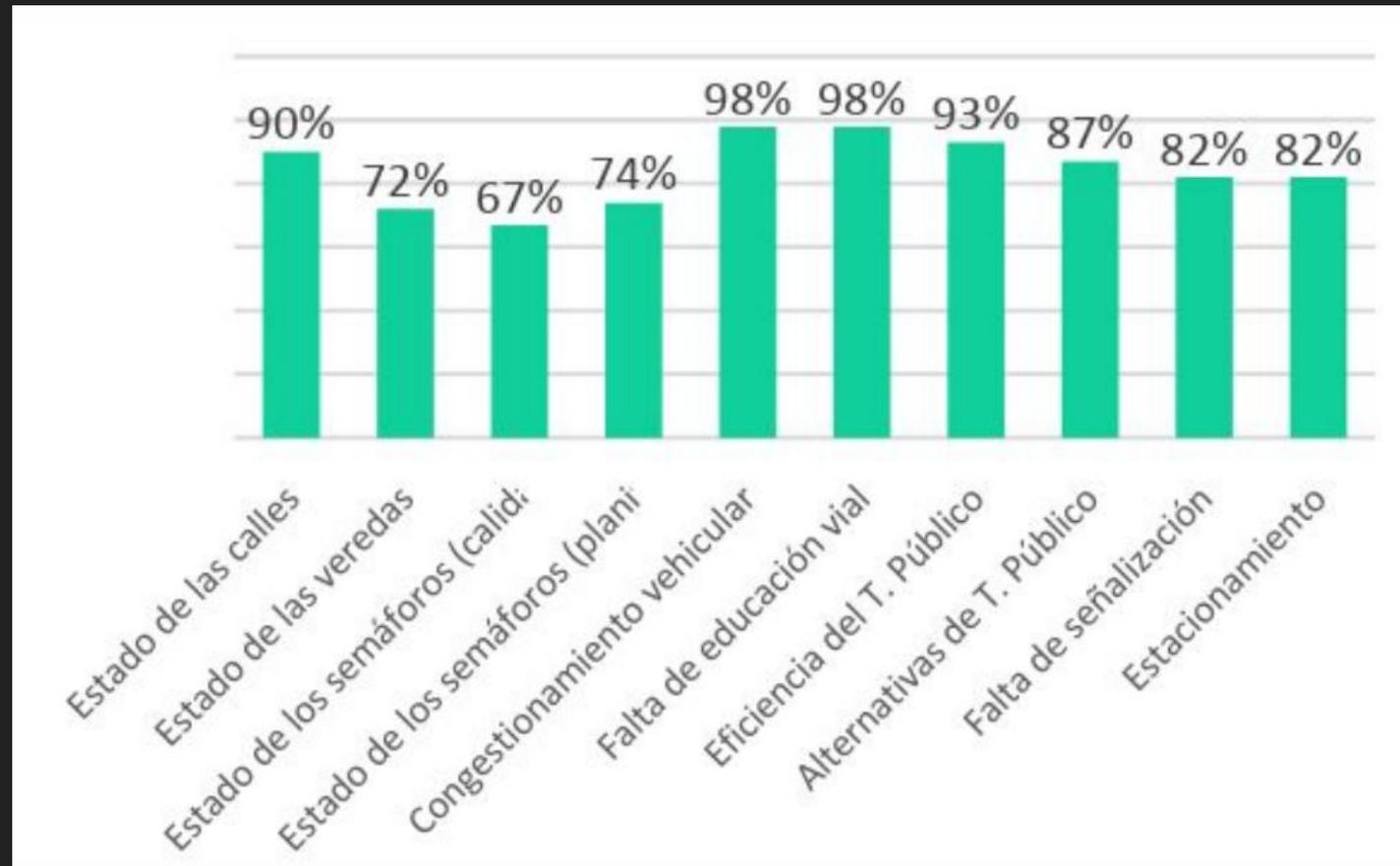


El Proyecto SmartTraffic

Crear soluciones **mobile crowdsensing** para problemas de movilidad urbana en Asunción.



SmartTraffic: Encuesta de Percepción Ciudadana



Problemas identificados como **GRAVES** o **MUY GRAVES** y por qué
porcentaje de los encuestados

SmartTraffic: Aplicaciones

- **SmartParking**

Reducir el tiempo necesario para encontrar lugares libres en una zona particular de estacionamiento

- **UCArpooling**

Optimizar la tasa de ocupación de vehículos para comunidades cerradas mediante el ride sharing

- **SmartMoving**

Facilitar el uso de veredas por parte de personas con capacidad motora reducida

¡Gracias!

¿Preguntas?

El estudio sobre la percepción ciudadana aquí:

<http://bit.ly/asuprobs>