

# CIPFP Mislata

Luis García Bonifaz - l.garciabonifaz@edu.gva.es

## BDA T01 - Fundamentos Big Data

---

### Actividad 1

---

#### Smart City - Tarea 1: Diseño del ADN de los Sensores

**Objetivo:** Entender la importancia de modelar los datos antes de empezar. Un buen diseño inicial nos ahorrará muchos problemas.

**Instrucciones:**

1. Crea un Repositorio Git Hub para ir añadiendo todo el trabajo del proyecto
2. Abre un editor de texto simple (Bloc de Notas, VS Code) o una hoja de cálculo.
3. Crea una entidad para el sensor de temperatura definiendo su estructura.
4. Crea una estructura similar para:
  - El sensor de CO2 (con un atributo `co2` en `ppm`)
  - El sensor de calidad del agua (con `ph`, `chlorine` en `mgL`, etc.).
5. No te preocupes por la sintaxis perfecta, céntrate en definir los **atributos, sus tipos de dato y sus unidades**.

#### Análisis Académico - Tarea 1: La Primera Inspección

**Objetivo:** Desarrollar un "ojo crítico" para la calidad de los datos. Es una de las habilidades más importantes de un profesional del dato.

**Instrucciones:**

1. Crea un Repositorio Git Hub para ir añadiendo todo el trabajo del proyecto
2. Descarga los ficheros CSV del proyecto académico.
3. Ábrelos con Microsoft Excel, Google Sheets o un visor de CSV. **No modifiques nada todavía.**
4. Conviértete en un detective y responde a estas preguntas para cada fichero:
  - ¿Cuál es el separador de columnas (coma `,` o punto y coma `;`)?
  - ¿La primera fila contiene los nombres de las columnas (encabezados)? ¿Son claros?
  - Inspecciona visualmente las primeras 20-30 filas. ¿Ves valores que te parezcan extraños o que faltan (celdas vacías, "N/A", "s/d")?

- ¿Los formatos son consistentes? Por ejemplo, ¿las fechas están siempre como `DD/MM/AAAA` o a veces cambian?
- Identifica las "claves" o "IDs" que podrían servir para relacionar unos ficheros con otros (ej: `id_alumno` en el fichero de `calificaciones.csv` y también en `alumnos.csv`).

5. Apunta tus hallazgos en un documento de texto. Este será nuestro punto de partida para la limpieza de datos en módulos posteriores.