# **Tema 3**

## Tratamiento de datos.

Se recopilan datos y se traducen en información usable. Lo analizaremos mediante algoritmos.

Podemos distinguir varios pasos:

* Descubrimiento de datos.
* Integración
* Explotación

**¿Qué es el munging de datos? 🡪** proceso inicial de refinar el contenido de los datos para adecuarlos al consumo de los usuarios y sistemas.

* Exploración de datos.
* Transformación e integración de datos.
* Enriquecimiento de datos e integración.
* Validación de datos.

## Modelos y tipos de datos.

Describen las características de los datos. Distinguimos:

* Operaciones.
* Constricciones.
* Estructuras.

Clasificación de los tipos de datos:

* **Estructurado (20% de los datos)🡪** datos con una longitud y formato definido (números, fechas, direcciones…). Se almacenan en una base de datos SQL. Se usan tablas.

¿Qué es un modelo relacional? 🡪 Los datos se almacenan en tablas

* Semiestructurado 🡪 tienen algunas características consistentes y definidas. No se limita a una base rígida como
* No estructurado 🡪