### **Attribution ShareAlike (CC BY-SA)image**

**Matriz de madurez de datos[[1]](#footnote-1)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Categoría** | **Áreas** | **Deficiente** | **Básico** | **Intermedio** | **Avanzado** |
| **Cómo se almacena la información** | **Acceso** | Solo se accede desde la aplicación que recolecta los datos | Se puede acceder a los datos, pero con un software especial y específico. | Los datos están en formatos accesibles como CSV, JSON, XML o una base de datos accesible de forma remota. | Se puede acceder a los datos mediante una API. |
| **Almacenamiento** | Papel | PDF o imágenes | Archivos de texto | Bases de datos |
| **Integración** | Los datos solo están en el sistema donde se recolecta la información. | Se exporta ocasionalmente los datos y se integra de forma ad hoc | Data warehouse con actualización de datos automática. | Tanto datos internos como externos están integrados en la base de datos. |
| **Qué información se recolecta** | **Relevancia** | Los datos son irrelevantes para el problema. Por ejemplo, quieres ver probabilidad de terminar la universidad pero no tienes los datos de quien se gradúa. | Algunos de los datos son relevantes, pero son insuficientes porque faltan componentes claves. Por ejemplo, ocupando el caso anterior, solo tienes asistencia a clases. | Se tiene información útil y relevante, pero no completa. Por ejemplo, en el caso anterior hay datos demográficos y notas anuales, pero no hay información de apoyo extra curricular. | Existe toda la información relevante y es suficiente para resolver el problema sin hacer transformaciones relevantes. |
| **Calidad** | Faltan filas | Faltan columnas (variables claves) | Están todos los datos, solo hay errores menores como de tipeo. | No hay problemas de falta de datos o errores de tipeo, las bases están sanitizadas. |
| **Frecuencia** | Solo una vez | Anual | Frecuente | Tiempo real |
| **Granularidad** | Agregado a nivel de ciudad | A nivel de cuadra o código postal | A nivel individual (persona o dirección) | Detalle a nivel de evento/hitos |
| **Historia** | Se borran los datos históricos. | Se guardan los datos históricos pero las actualizaciones se sobreescriben. | Se guarda la información histórica ocupando timestamp. | Todos los datos se mantienen y se relacionan con los datos anteriores con un modelo integrado. |
| **Otros** | **Privacidad** | No hay política de privacidad | La política no permite ocupar ningún dato | Se necesita aprobación ad-hoc para el uso de los datos. | Está definido el acceso a los datos y se controla la privacidad de los datos para preservar la privacidad de los individuos. |
| **Documentación** | No existe documentación digital o metadata. Los códigos de las variables no están documentadas. | Se define un diccionario de datos que explica las variables y categorías. | Existe un diccionario de datos y disponibilidad de metadata. | Existe diccionario de datos, metadata, y adicionalmente los supuestos, sesgos y los datos que no se están obteniendo. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Categoría** | **Áreas** | **Deficiente** | **Básico** | **Intermedio** | **Avanzado** |
| **Cómo se almacena la información** | **Acceso** |  |  |  |  |
| **Almacenamiento** |  |  |  |  |
| **Integración** |  |  |  |  |
| **Qué información se recolecta** | **Relevancia** |  |  |  |  |
| **Calidad** |  |  |  |  |
| **Frecuencia** |  |  |  |  |
| **Granularidad** |  |  |  |  |
| **Historia** |  |  |  |  |
| **Otros** | **Privacidad** |  |  |  |  |
| **Documentación** |  |  | . |  |

1. Adaptada del Data Maturity Framework de la U. de Chicago <http://dsapp.uchicago.edu/resources/datamaturity/> [↑](#footnote-ref-1)