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

**Data Maturity Framework[[1]](#footnote-1)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **Area** | **Lagging** | **Basic** | **Advanced** | **Leading** |
| **How is Data Stored** | **Accessibility** | Only accessible within the application  where it is collected | Can be accessible outside the application  but proprietary format, requiring  specialized analysis software | All machine readable in standard open  format (CSV, JSON, XML, database) | All machine readable in standard open format and available through an API |
| **Storage** | Paper | PDFs or Images | Text Files | Databases |
| **Integration** | Data sits in the source systems | Data is exported occasionally and  integrated in ad hoc manner | Central data warehouse - realtime  aggregation and linking (Automatic) | External data also integrated |
| **What is Collected?** | **Relevance and Sufficiency** | The data you are collecting on subjects of interest is irrelevant to the problem you want to solve: ie you want to do predict which students need extra support to graduate on-time but don't have data on graduation outcomes | Some of the data you have is relevant,  but it is insufficient because key fields are  missing, ie no data on academic behavior  or attendance history, etc. | You have data that is helpful and relevant  for solving the problem but not sufficient to  solve it well. ie you have yearly academic  and demographic information but are  missing extra-curricular activities, or  interventions they were targeted with | You have all the relevant data  about all the entities being  analyzed and it's sufficient to  solve the problem you are  tackling |
| **Quality** | Missing rows (people/address level  entities missing in the data) | Missing columns (variables missing) | No missing data but errors in data  collection such as typos | No missing data and no errors  in data collection |
| **Collection Frequency** | Once and never again | yearly | frequently | realtime |
| **Granularity** | City level aggregates | Zipcode/Block level aggregates | Individual level (person or address) level data | Incident/Event level data |
| **History** | No History Kept - old data is deleted | Historical data is stored but updates overwrite existing data | Historical data is stored and new data gets appended with timestamp, preserving old values | All history is kept and new data schema gets mapped to old schema so older data can be used |
| **Other** | **Privacy** | No privacy policy in place | no PII can be used for anything | ad-hoc approval process in place that allows selected PII data to be used for selected/approved projects | Software defined/controlled privacy protection that allows analytics to be done while preserving privacy based on predefined policies |
| **Documentation** | no digital documentation or metadata: data exists but field descriptions or coded variables are not documented | data dictionary exists (variables and categories defined) | data dictionary plus full metadata available (including conditions under which the data were captured) | data dictionary plus full metadata available including collection assumptions, what's not collected, and potential biases |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **Area** | **Lagging** | **Basic** | **Advanced** | **Leading** |
| **How is Data Stored** | **Accessibility** |  |  |  |  |
| **Storage** |  |  |  |  |
| **Integration** |  |  |  |  |
| **What is Collected?** | **Relevance and Sufficiency** |  |  |  |  |
| **Quality** |  |  |  |  |
| **Collection Frequency** |  |  |  |  |
| **Granularity** |  |  |  |  |
| **History** |  |  |  |  |
| **Other** | **Privacy** |  |  |  |  |
| **Documentation** |  |  | . |  |

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