

## Data Maturity Framework Questionnaire

Scorecard Category	Question				
<b>Problem Definition</b>					
<b>Problem Definition</b>	What is the problem you are trying to solve? What does success look like/how much does the needle need to move?				
<b>Interventions</b>	What interventions do you have available to solve the problem?				
<b>Impact</b>	If this is successful, what impact will this project have? Will it encourage future projects/goodwill?				
<b>Available Data</b>	What data sets do you have access to relevant to the problem?				
<b>Data Fields</b>	What fields are in each of the data sources? (See Data Sources Worksheet)				
<b>Size</b>	How many people/addresses/facilities/entities does the data contain?				
<b>Target Population</b>	For this problem, what % of entities are at risk or have resources to be intervened?				
<b>Data Governance</b>					
<b>Ownership</b>	For the data sets that you have access to - do you own the data? Do you have permission to use the data? If you do not own the data, do you have the relationships with the data owner?				
<b>Physical Accessibility</b>	Is the data accessible outside the department/agency? Is there a VPN?				
<b>Security Policy</b>	What security policies and considerations need to be in place for each of the data sources? (HIPPA, FERPA)				
<b>Implementation and Maintenance</b>					
<b>Technical Implementation</b>	Do you have people in house who can implement/deploy the solution?				
<b>Data Infrastructure</b>	Do you have the internal tech and data infrastructure to provide a continuous data feed from all the systems, and integrate the results/recommendations back in to the agency systems?				
<b>Maintenance</b>	Can you update, maintain, and support the implemented solution?				
<b>Data Readiness</b>					
		<b>Lagging</b>	<b>Basic</b>	<b>Advanced</b>	<b>Leading</b>
<b>Accessibility</b>	How accessible is the data that's required?				
<b>Storage</b>	How is the data stored?				
<b>Integration</b>	How integrated are the different data sources?				
<b>Relevance and Sufficiency</b>	Do you have data that is both relevant and sufficient to solve the problem?				
<b>Quality</b>	How is the data quality?				
<b>Collection Frequency</b>	How often is the data collected?				
<b>Granularity</b>	What is the level of granularity for the data sources?				
<b>History</b>	How much history is stored and how are updates handled?				
<b>Privacy</b>	What data privacy policies do you have in place?				
<b>Documentation</b>	How well documented are the data?				
<b>ORGANIZATIONAL READINESS</b>					
		<b>Lagging</b>	<b>Basic</b>	<b>Advanced</b>	<b>Leading</b>
<b>Staff Buy In</b>	How bought in are staff throughout the organization? What percentage of the staff are involved in data collection? Data analysis?				
<b>Data Collector Buy In</b>	How bought in are the people on the ground doing the data collection? Do they understand the importance and nuance of data collection? Do they get direct benefit from collection data?				
<b>Leadership Buy In</b>	How does leadership value data? Do they require data to be presented in order to make decisions?				
<b>People Resources</b>	Do the people who will act on the results buy in?				
<b>Data Use Policy</b>	Are there policies in place around who can use data, how they can use data, which parts can they use, and for what purposes?				
<b>Intervenor Buy In</b>	Do the people who will act on the results buy in?				
<b>Funder Buy In</b>	How do your funders consider data? What kind of data do they require? What support for technology and personnel do they give you?				

## Data Maturity Framework Data and Tech Readiness Scorecard

Category	Area	Lagging	Basic	Advanced	Leading
<b>How is Data Stored</b>	<b>Accessibility</b>	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
	<b>Storage</b>	Paper	PDFs or Images	Text Files	Databases
	<b>Integration</b>	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
<b>What is Collected?</b>	<b>Relevance and Sufficiency</b>	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
	<b>Quality</b>	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
	<b>Collection Frequency</b>	Once and never again	yearly	frequently	realtime
	<b>Granularity</b>	City level aggregates	Zipcode/Block level aggregates	Individual level (person or address) level data	Incident/Event level data
	<b>History</b>	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
<b>Other</b>	<b>Privacy</b>	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
	<b>Documentation</b>	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

## Data Maturity Framework Organizational Readiness Scorecard

Area	Lagging	Basic	Advanced	Leading
<b>Staff Buy In</b>	Staff at the organization have some idea that data exists but doesn't understand it is important	There are a few individuals who deeply understand the data available and what can be done with it	Organization has a clear idea of how data can be used to drive business decisions beyond justification of funding	Organization has a culture of data within the organization and demands data to justify all programmatic decisions
<b>Data Collector Buy In</b>	On the ground staff provide data seldomly, sporadically, or incompletely because they are required to but it is seen as a hindrance to their "real job"	On the ground staff regularly provide data because they are required to	On the ground staff provide data on a regular basis and eventually get actionable insights in return	On the ground staff provide data in real time and make decisions based on the data and insights available to them, and offer suggestions on what is collected/what information they could use to improve their job effectiveness
<b>Leadership Buy In</b>	Leaders at this level fundamentally don't know how data can help advance the organization's mission.	Leadership wants to use data but don't have a clear path forward to use data	Leadership has a clear idea of how data can be used to drive business decisions beyond justification of funding	Leadership builds a culture of data within the organization and demands data to justify all programmatic decisions
<b>People Resources</b>	Individual stakeholders maintain siloed data sets	The organization knows how data can help, what data they need, and are able to access it, but lack the in-house data skills, tools, or infrastructure to be able to turn data into meaningful insights that affect human action.	Organizations know how data can help, what data they need, and are able to access it, but lack either the infrastructure or the people to be able to turn data into meaningful insights that affect human action.	The organization has dedicated staff who own data storage AND data content owners who own the cleaning and rigor of the data
<b>Data Use Policy</b>	No policies exist around use, transfer, and sharing of data	Organization has policies in place for the use, transfer, and sharing of data but it does not cover all data that exists within the organization	Organization has policies in place for the use, transfer, and sharing of data internally	Organization has policies in place for the use, transfer, and sharing of data internally and externally
<b>Intervenor Buy In</b>	No partnerships exist	Partnerships exist but data is not shared	Partnerships exist and have policies and technology in place to share data occasionally or through a manual process	Partnerships exist and have policies and technology in place to share data in real-time
<b>Funder Buy In</b>	Funders do not require data other than vanity metrics	Funders ask for key performance metrics	Funders ask for key performance metrics and provide funding for data infrastructure and maintenance	Funders require data driven decision making and provide funding for data infrastructure, maintenance, and usage