**TIPS FOR DATA COLLECTION & REPORTING**

GUIDANCE: The following is offered as a distillation of a set of reasonable research practices and considerations that will inform what data you want to collect, and how you collect it. It is not intended to be an exhaustive overview of standard social science research methods. Careful attention to these matters will increase the likelihood that ultimately you have enough reasonably high-quality data to satisfy your reporting requirements, to have an understanding of your accomplishments and to provide data in service of this sector.

Contemplating, to the extent that you can, all the things you/your funder wants to know (or report) will help ensure that the information you collect will be reasonably complete. This is often a challenge. First, people who are not researchers by nature do not ordinarily thing about questions in terms of “data”. Second, in constructing a data collection effort you are managing competing interests. Those interests are, on the one hand, learning as much as possible about this particular moment as it relates to the sector. And on the other hand, you don’t want to overburden the providers collecting information that you may never use. Think about it this way: Every piece of data you collect comes at a cost. Will that piece of data provide funder-required information or meaningful information for the sector?

Following some of the “tips” below will help to ensure that the data you collect will be sound and can provide a foundation to both understand what you’re accomplishing and to refine ongoing program activities. The bigger picture is that we are at a very critical moment for a vulnerable sector. High-quality and reasonably comprehensive data on emergency funds for the home-based childcare sector will allow for a report of accomplishments and learnings back out to the sector, strengthening the system in the long-term.

Lastly, this material should be reviewed before you launch your program. That is because once the program launches, the level of effort necessary to go back and gather various pieces of data will be much greater – and likely yield less valuable results – than if you had a good strategy at launch.

*Data Collection Tools / Instruments*

* Today, there are many online vehicles to gather data (e.g., online surveys such as Google Forms, Survey Monkey). They vary in cost from free to relatively expensive and you should take care in selecting that which will meet your collection and analysis needs. But none of the tools ensure that the questions you ask are the right ones or are properly formed; that is something you need to do.
* There are important differences between data collected in a process that has the respondent completing a survey themselves (i.e., self-administered) versus one in which you have a staff member asking questions of respondents and recording their answers.
	+ A self-administered data collection tool should be simple, clean, and inclusive of clearly written instructions.
	+ Self-administered data collection tools should also include plain-language definitions of terms, as necessary – so that all respondents have the same understanding of the survey/question’s purposes.
* Do make sure that there is enough identifying information collected (e.g., respondent name, provider, address) to:
	+ Know who provided the data.
	+ Associate the survey response with other data you are collecting (e.g., financial information as part of a grant application).
		- If you can give each provider a unique identification number (i.e., no two providers have the same number) that will enable you to connect other data you may have (e.g., licensing information) or will collect in the future (e.g., a survey after funds were disbursed).
		- Some data collection tools may allow you to catalog names of centers, for example, into a pull-down menu for future data collection.
* Do record the date the survey was completed. The world is changing quite rapidly, especially in certain communities. Knowing when the data was collected against a chronology of how the COVID-19 pandemic is unfolding is important context.
* Do not ask too many questions (i.e., prioritize what you need to know over that which you’d like to know) – 15 or 20 questions that the respondent can answer in a few minutes is better than a large number of questions that will take a long time to answer.
	+ Only ask questions that are germane to the specific objectives you set out for the data collection effort.
* Do review your questions to ensure that you are not using jargon, acronyms or other terms that the respondent won’t understand (or could understand differently than you could expect).
* Do make sure that each question is just one question.
* Experts agree that questions that ask people what they *will do* are generally less reliable than those that ask what they *have done*. Minimize the number / importance of any questions that ask respondents to project into the future – unless they are required by the funding program.
* Do test out your questions with a few people – perhaps coworkers – to:
	+ See how long the survey takes (shorter is better).
	+ Make certain that the questions objectively elicit information and do not “lead” the questionnaire respondent to a particular answer (perhaps an answer they might think you want to hear).
	+ Ensure that the questions elicit the information you think you are eliciting with the question.
* Do use mostly closed-ended questions (that is, there is a fixed number of multiple-choice responses).
	+ Pay attention to ensure that you have all the possible categories among your multiple-choice responses (good way to do that is through a pre-test, as noted above).
* Use open-ended questions only when you are looking for a “narrative” from the respondent. Then ensure that you take notes on that narrative or allow the respondent to write the narrative themselves.
* Do not ask questions that the respondent likely cannot answer given what they know about the circumstance.
	+ Unless for a required compliance/eligibility purpose, do not ask questions that require “research” on the part of the respondent.

*Collection Points: Baseline (pre-) and post-event analysis*

* The world is changing quickly, especially for childcare providers who are financially fragile. Therefore, “pre-” and “post-” measurement may not be too far apart in time.
* The “pre-” state of the provider would be, logically, at the time that they are filling out the survey. However, some “pre-” measures (e.g., size of staff, number of children served, financial standing) should suggest that the respondent reflect back to a month or so prior to the large-scale onset of the COVID-19 pandemic.
* The “post-” measurement should take place long enough after assistance has been provided (not just approved) so that its uses and impacts can be reported. Depending upon how quickly funds are disbursed, 30-45 days after may be the appropriate time for “post-” measurement.
* If there are subsequent disbursements of funds, there should be subsequent follow-up surveys.

*Data Integrity*

* If the data collection tool is administered by a funder staff member or some party other than the respondent, ensure that everyone taking information from a respondent has the same understanding of the purpose of the questions, the meanings of the potential responses and how to “code” responses.
	+ This is enhanced by have a clear set of instructions.
	+ As noted previously, if the survey is self-administered (i.e., the respondent takes the survey on their own without the personal assistance of a surveyor, staff member, etc.) make sure that there are clear instructions on how to complete the survey and any definitions.
* Make sure that the database into which responses are collected saves each response (i.e., do not wait days to save the database as you add responses).
	+ This is managed by several survey tools that save automatically at the completion of the survey.
* The information you are collecting is likely to includes some data elements that are personal or confidential. Take steps to ensure that only those very few people who need to see the data can see it in a way that would allow them to connect an answer to a particular respondent (e.g., home-based provider X has $2,275 / month in revenue). This is achieved through: (1) making sure staff understand the confidential nature of the data and the importance of maintaining that privacy and (2) not leaving the database in a location (e.g., on a wide accessible server or a thumb drive that anyone can pick up and read) that is accessible to those who have no legitimate purpose to access the data and have not been instructed as to the privacy concerns.

*Reporting and Analysis*

* Most data collection tools will have an ability to generate “frequencies” or “marginals” for each question.
	+ Percentages of respondents for each question option is always useful.
		- Percentages based on the number of valid responses to a question (i.e., the number of respondents that completed the question) – rather than the total number of respondents to the survey – is a proper way to account for less than complete information.
		- Bar charts and pie charts are good visual ways of displaying your data – again, many data collection tools will do that for you.
	+ You may not be able to do much more than frequencies and percentages, especially if the number of surveys is small. That said, you might be able to do frequencies/percentages across multiple variables (e.g., purpose of funding (salary, rent/utilities, supplies) against, for example, the amount of funding requested.
		- Statistics such as means or standard deviations require the underlying data to be continuous in nature (e.g., data elements like the undifferentiated number of children served by the provider or annual provider income).
		- Some statistics, like means, are influenced by extreme values (e.g., 9 centers, each with between 5 and 10 children and 1 center with 250 children will produce a mean that does not represent a typical center). In that instance, use measures such as medians (i.e., if you had 11 respondents who gave you information about the number of children they serve and you put them in order from smallest to largest, the median is the middle value) or modes (i.e., the most frequently occurring value).
	+ Do not overemphasize small differences when there are relatively few respondents. Consider that with 20 respondents, each respondent’s answer is worth 5 percentage points on any given question option.
	+ This kind of assessment will be most fruitfully done if there is both qualitative data (e.g., analysis of narratives from respondents) and quantitative data (e.g., tabulating the number of responses to questions, connecting one question to another). Social scientists call this “mixed-methods research.”
		- To derive maximum value from the qualitative data, best to read all responses and develop a set of “concepts” or “themes” that appear to be repeated in at least a few of the responses
		- Then associate those concepts or themes to the survey responses.
			* Don’t be afraid to use de-identified respondent quotes that illustrate a critical point in any write-up you produce.

*Last Piece of Advice*

Your ability to learn and inform is only as good as the weakest piece of your data collection / analysis process. Do the best that you can to construct a clear and sufficiently complete data collection tool that will yield the fundamental information you need to understand your accomplishments and inform the field. Questions that are poorly constructed, collecting data you won’t (or can’t) use or collecting the data in an inconsistent manner is counterproductive. Your guiding principle here is exemplified in a quote from Albert Einstein: “Everything should be made as simple as possible, but not simpler.”