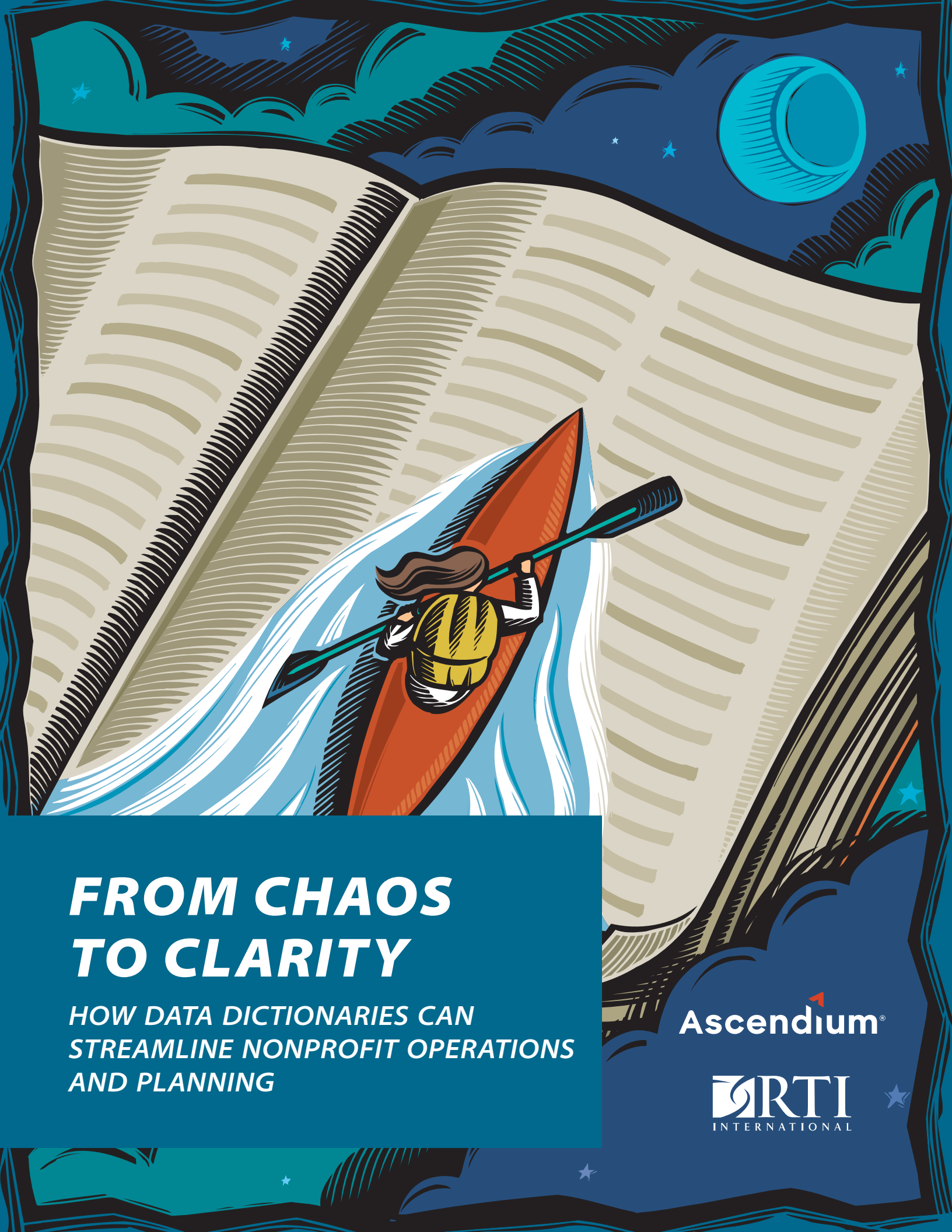
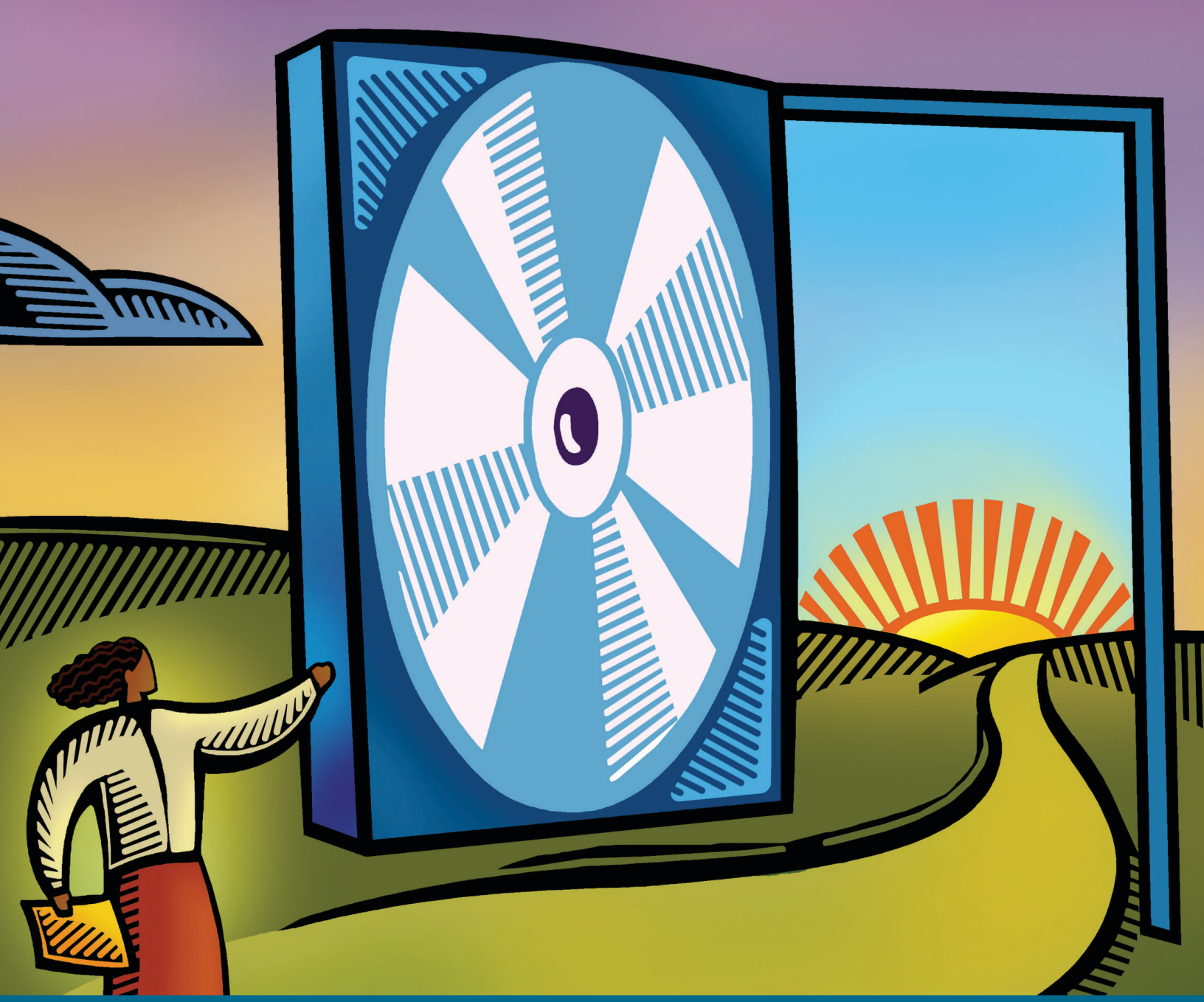


FROM CHAOS TO CLARITY

HOW DATA DICTIONARIES CAN STREAMLINE NONPROFIT OPERATIONS AND PLANNING

Ascendium®

RTI INTERNATIONAL



AUTHORS

Samantha A. Tosto, Hannah G. Cortina, and Alicia McKay, RTI International

SUGGESTED CITATION

Tosto, S. A., Cortina, H. G., & McKay, A. (2025). *From Chaos to Clarity: How Data Dictionaries Can Streamline Nonprofit Operations and Planning.* RTI International.

May 2025

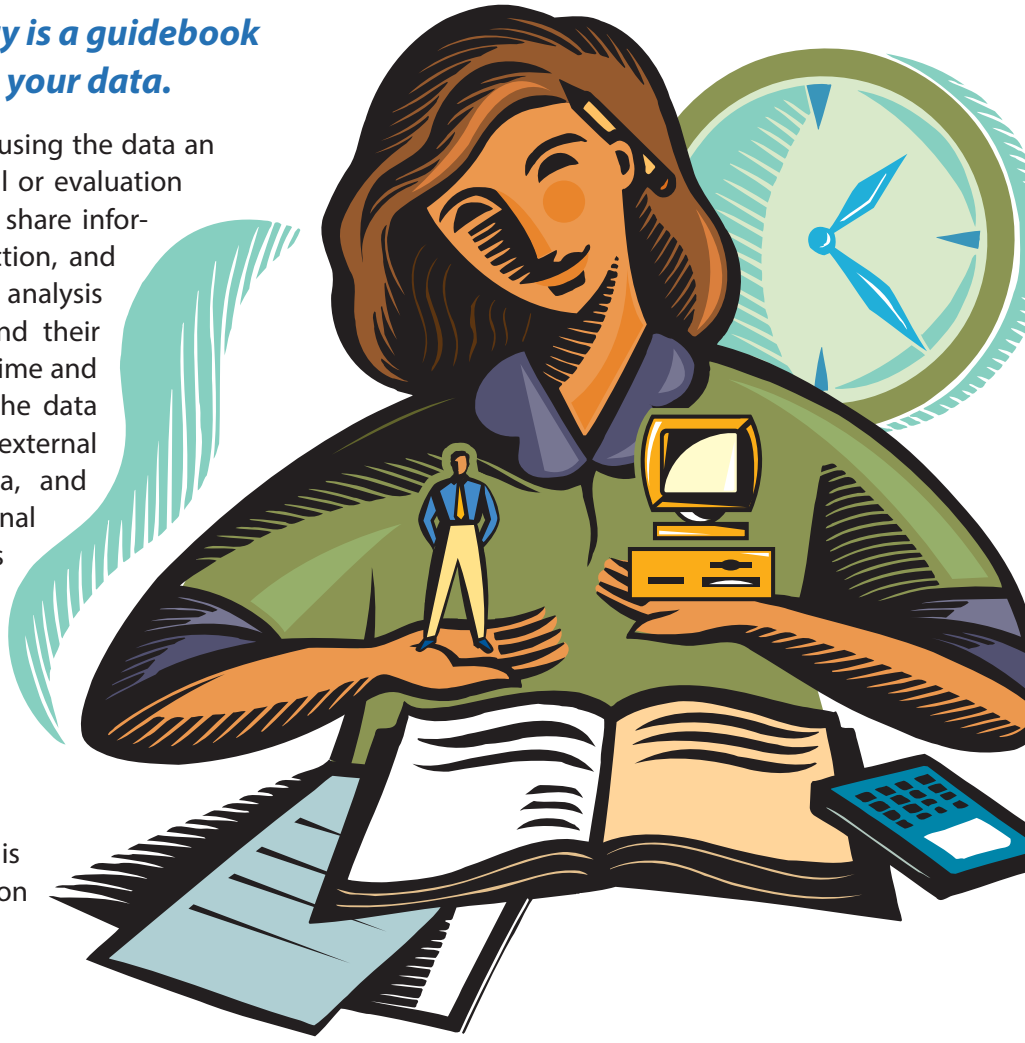
WHAT IS *A DATA DICTIONARY?*

Most organizations collect information, commonly referred to as "data," on the people they are working with and providing services to. Examples of data are participant demographics and background information, amount of time a person has served or number of services they have received, number of clients on a staff member's workload, and much more. The process of collecting, combining, cleaning, storing, managing, and analyzing these data can be both difficult and time-consuming. Given this investment in time and resources, it is important that an organization be able to properly use the available data to further its implementation and sustainability of services and to evaluate overall program impact.

For data to be useful across the organization and across time, staff have to be able to "read" the data. This means that people aiming to use the data to answer questions about the organization's practices need to be able to look at the data and know what they mean. A data dictionary makes this possible by describing each piece of data collected by an agency—including what it is called, what it means, how it is defined or measured, the possible responses or information that it provides, and how it is summarized or quantified by the organization.

Essentially, a data dictionary is a guidebook to understanding and using your data.

It is the road map for entering and using the data an organization collects for operational or evaluation purposes. It is used by agencies to share information about variables, data collection, and data entry, which allows for data analysis and interpretation of your data and their findings in a consistent way across time and staff. Data dictionaries also make the data sharing process easier—they allow external partners to understand your data, and they reduce the need for additional trainings, the number of questions readers have, and the potential for misunderstandings. This research brief provides steps on starting a data dictionary, tips and tricks for best practices, and insight into the various benefits of developing and maintaining a strong data dictionary for any nonprofit organization that is interested in improved data collection and analysis.



BENEFITS
TO YOUR PROGRAM OR AGENCY

Daily Organizational Needs

In terms of daily use, a data dictionary assists with data continuity, analysis, and replication. That is, it helps convey what the data actually mean, ensures that each data point maintains the same meaning over time, aids in identifying and tracking necessary changes to data collection instruments or questionnaires, and makes data entry or interpretation easier and less prone to inaccuracies. Data dictionaries also reduce the likelihood of user error, duplication of staff work, overcomplication of the entry process, and coding inaccuracies. Additionally, when other organizational staff need to look at the data—which may occur for a variety of reasons (e.g., reporting out to funders, summarizing client outcomes, conducting research)—a data dictionary simplifies the process by creating clear and transparent descriptions of the dataset that are easily interpretable, even if a user has minimal prior knowledge of the data.

Data Sharing, Research, and Evaluation

A data dictionary can also help other organizations, external evaluators, and researchers understand your data in a snapshot. Outside users will be able to see where your data go and how they are measured, as well as how both external and internal users will be able to track success. If your program is being evaluated, a data dictionary can reduce the burden of the evaluation process on your team and budget by requiring less involvement from staff who may otherwise be needed to explain each variable. Moreover, the dictionary can streamline the evaluation process, as the evaluators will have a data "map" to follow. A data dictionary allows outside agencies (both evaluators and possible partners) to understand how your data management works. Strengthening this evaluation process can help foster these partnerships, which may subsequently bring forth new ideas or collaborative ways for serving clients and measuring their outcomes.

Essentially, a data dictionary will contribute to creating stronger, clearer results and may ultimately help your program test its effectiveness and contribute to sustainability.

WHAT GOES IN *A DATA DICTIONARY?*

A data dictionary serves as a key or guide that allows someone to be able to read and analyze a dataset, even if they are not already familiar with the information included.

Datasets may include information from intake forms, participant surveys, demographic information, or outcome assessments. The information will often be converted into numerical or coded values in the data file. A data dictionary helps connect those values with understandable terms.

Here are some fields that might be included in your data dictionary.

1. Question Text/Data Element—Outward-facing language of the variable (i.e., what is the client asked about?)

- **Ex.** "Please select the option that best describes your highest level of education."
- **Ex.** "Please indicate your agreement with the following statement: I have strong family relationships and friendships."

2. Variable Name—Abbreviated name of the variable for internal purposes

- **Ex.** Education
- **Ex.** Relationships

3. Variable Label—Brief description of the variable

- **Ex.** Highest education achieved
- **Ex.** Strength of relationships

4. Variable Measurement—Classification of how the item is measured. There are four options: nominal (qualitative categories), ordinal (ranks), interval (known, equal intervals), or ratio (has a true zero point).

5. Response Options—Specify whether responses to this item were open-ended; if not, list the choices that were made available to the respondent

- **Ex.** HS diploma/GED; Some college/associate's degree; Bachelor's degree; Graduate degree; Prefer not to respond
- **Ex.** Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree

6. Value Label—Value or numerical indicator assigned to each response option. In other words, what numerical value is assigned to each answer response option that will be tracked in the data system? If the data element is open-ended, value labels will not be necessary.

- **Ex.** HS diploma/GED = 1; Some college/associate's degree = 2; Bachelor's degree = 3; Graduate degree or higher = 4; Prefer not to respond = 5; Missing data/blank response = 999
- **Ex.** Strongly disagree = 1; Somewhat disagree = 2; Neither agree nor disagree = 3; Somewhat agree = 4; Strongly agree =5; Missing data/blank response = 999

EXAMPLE FIELDS

The tables below emphasize of the importance of having a data dictionary. **Table 1** is a sample dataset that includes intake assessment information on six of an organization's clients. Without a data dictionary or previous exposure to these data, understanding and analyzing this dataset would be difficult, as some item values might not be intuitive. For example, a staff member unfamiliar with the data collection instruments or data storage would be unable either to interpret what each variable indicates in terms of what was asked of the client or to understand the numerical values listed for each variable represented.

Table 1.
Sample Dataset of Client Intake Information

IDNumber	Education	Reentry Challenges	Relationships
2025_001	1	Finding a job	2
2025_002	4	N/A	5
2025_003	3	999	4
2025_004	999	Employment; family issues	2
2025_005	2	Money	999
2025_006	3	Driver's license	1

Table 2 is a sample data dictionary that provides the codes or definitions for the items included in Table 1. The Question Text/Data Element column lists some questions that might be included on a client intake assessment. The subsequent columns provide defining information of the variable name, label, response options, and values that a person reading these data would need to know to understand what the content of the dataset means. With this data dictionary, a staff member would more easily be able to understand and analyze the data in Table 1.

Table 2. Sample Data Dictionary Fields for Client Intake Information

Question Text / Data Element	Client identification	Please select the option that best describes your highest level of education	What are the biggest barriers/challenges related to your reentry goals?	Please indicate your agreement with the following statement: I have strong family relationships and friendships.
Variable Name	IDNumber[1]	Education	ReentryChallenges[2]	Relationships
Variable Label	Client anonymized identification	Highest education status	Challenges related to reentry	Strength of relationships
Variable Measure-ment	Numerical	Categorical	Text	Categorical
Response Options	Assigned at intake by clinician	Select options: • HS diploma/GED • Some college/associate's degree • Bachelor's degree • Graduate degree or higher • Prefer not to respond	Open-ended	Select scale options: • Strongly disagree • Somewhat disagree • Neither agree nor disagree • Somewhat agree • Strongly agree
Value Labels	N/A	(1) HS diploma/GED (2) Some college/associate's (3) Bachelor's (4) Graduate degree or higher (5) Prefer not to respond (999) No response	(999) No response	(1) Strongly disagree (2) Somewhat disagree (3) Neither agree nor disagree (4) Somewhat agree (5) Strongly agree (999) No response

[1] It is important to note that tracking client ID for each response is critical to being able to link client data over time or across datasets, as some information may be stored in separate files.

[2] Note that the variable names do not have any spaces. Many data software applications require that there be no spaces in these fields.

HOW TO **GET STARTED**

Creating a data dictionary may feel like a daunting task at first, especially if you are not well experienced with data collection and analysis. The following steps can help you get on the right track to creating a data dictionary.

1 *Take an inventory of all the data you are collecting.*

Ideally, all the data will be stored in one data management system. In some cases, there may be multiple data systems or a split between computer and paper forms. Either way, create a list of every piece of information you collect on clients and the response options for each component.

> **Some data management systems can export a list for you so that you do not have to collect the information manually.**

2 *Enlist your team.*

Ask the staff members who are entering the data to define the variables and response categories. If the proposed definitions vary, meet as a team to decide on the correct definitions.

> **This activity may also lead you to add variables to your data collection if staff identify gaps in information. Conversely, you may decide to remove variables if staff identify duplicative information.**

3 *Identify the response value labels.*

As explained above, the response value labels are important to interpret the data during analysis. To find these, you will want to confirm the coding scheme with whomever in your organization designed the data collection measure. If no explicit values were assigned during design, data management systems often automatically code response options in numerical order beginning with 1 (i.e., 1, 2, 3, …. n). If it seems like this is the case, make sure you review the order in which response types are offered on the original data collection measure and track these codes in your data dictionary.

4 *Flag missing data.*

Sometimes, data are missing. A client may have skipped a question on the data collection instrument, or an instrument may not have been available or used yet at some point in the program's history. As such, it is important to flag and create identifiers for missing data in a way that is different from legitimate responses that may indicate a lack or absence of the variable. For example, there should be a clearly identifiable difference between a respondent who reports that they have been incarcerated 0 times and one who skips the question entirely. A common practice (as shown in the tables above) is to list missing numerical values as 999 (as this is not a number that is likely to indicate a person's age, highest grade level, or other common numerical items) and to list missing categorical variables as N/A (not available) or NR (not reported).

5 *Organize the data dictionary.*

It may make sense to list the variables alphabetically or sequentially (i.e., the order in which they are entered into the system), depending on your organization's needs. For example, if your staff must often look up specific client information, it may be easier for them if the data are organized alphabetically. Conversely, if staff want to review how client information has changed over time, organizing the variables sequentially would make it easier to identify when each data point was collected.

6 *Test it!*

Create a pretend client and go through your intake and service processes. As you are entering data, check to make sure each component is listed in the data dictionary accurately.

Additionally, have data entry staff test the dictionary in real time and note any difficulties or confusing points.

Export the data from whatever data management tool is being used (e.g., Excel, Salesforce) and conduct a quality check on how the data transfer across platforms.

HELPFUL
TIPS & TRICKS

There are many uses for a data dictionary. However, it is only as useful as it is *accurate*. You can maintain its fidelity and efficiency by following these recommendations.

Keep separate data dictionaries for each dataset in an organization, if applicable.

Make sure all questions/data options are both exhaustive (include all possible options) and mutually exclusive (do not overlap in their meaning).

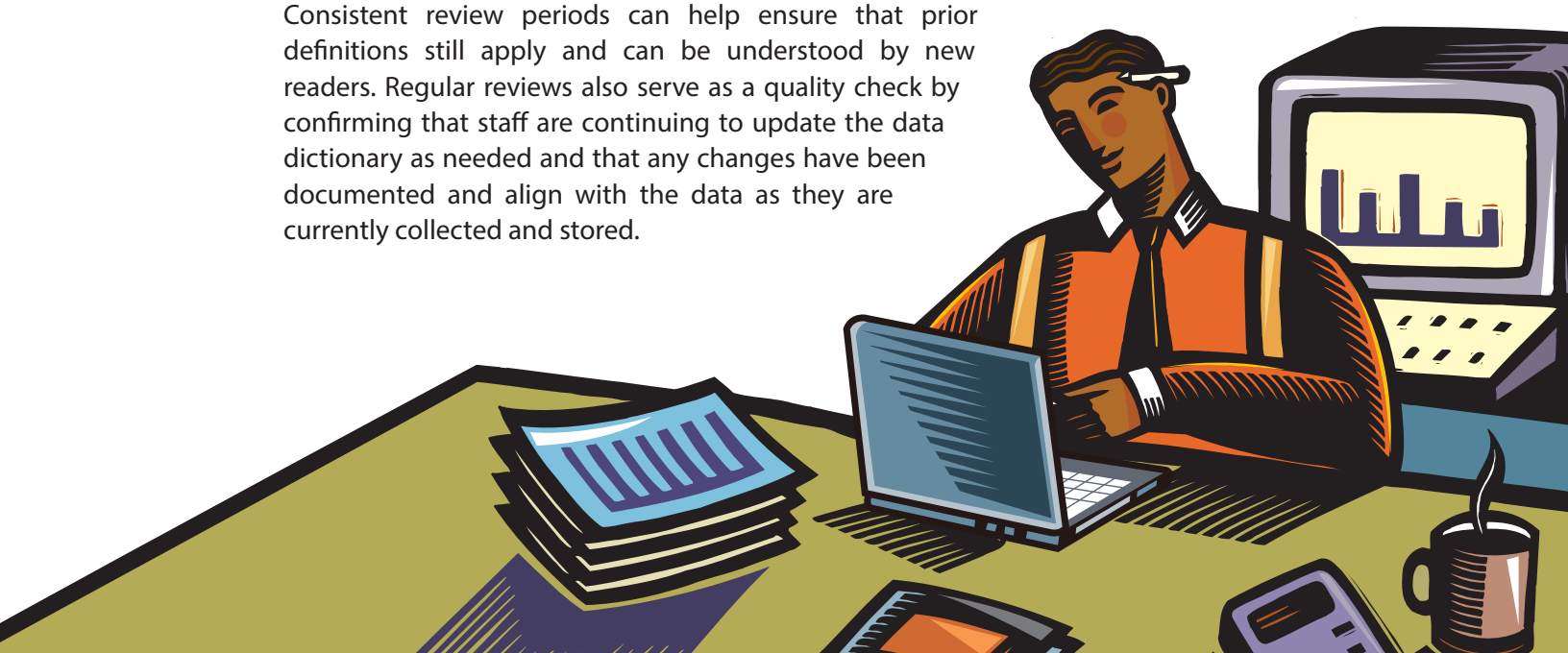
Create indicators of missing data (prevents blank fields in your dataset). See #4 in the How to Get Started section.

If *anything* changes in your data collection (e.g., adding variables, changing answers), update it in your data dictionary.

- Archive previous versions in a separate folder with the last used date or in a manner that makes sense to your organization. To avoid any confusion among staff, make it obvious which data dictionary is current.
- Archiving older versions of the dictionary allows future staff or organizational leaders to see when changes were made to data practices.
- Keep a modification log in which all staff enter any changes they make to the data dictionary. This log should include the change made, rationale for the change, date the change was made, and whether that change is retrospective (changes all old data) or should be used only prospectively (for new data).

Set a regular review period for your data dictionary.

- It may make sense to review annually or semiannually, or more or less frequently depending on your organization. Consistent review periods can help ensure that prior definitions still apply and can be understood by new readers. Regular reviews also serve as a quality check by confirming that staff are continuing to update the data dictionary as needed and that any changes have been documented and align with the data as they are currently collected and stored.



CONCLUSION

Data collection, management, and analysis can be a lot of work, placing additional resource burdens on an organization's designated staff. However, clean and interpretable data are vital to measuring outcomes and program progress. Having a data dictionary will benefit your agency's ability to understand, report on, and share your data. Using a data dictionary will also improve your program's daily organizational abilities, and it will improve your program's capacity for data sharing, research, and evaluation. Although the creation of a data dictionary is a meticulous task and maintaining one requires upkeep, the benefits far outweigh the effort. These benefits include better positioning your organization to understand the impact it is having and to support sustainability efforts. Ultimately, the use of a data dictionary is an essential component of fostering your agency's strength and achieving desired outcomes for your organization, your staff, and the community members you serve.


Ascendium®
RTI
INTERNATIONAL