

## 2023 New York Statewide Civil Legal Aid Technology Conference

### 1A Legal Aid Data and Evaluation Toolkit

Tuesday, April 18, 2023

3:30 PM – 4:20 PM

Live Virtual Presentation

CLE Credits: 1.0 Practice Management

### PRESENTERS

*Moderator: [Shellie S. Reid](#), Manager, LSNTAP*

*Speaker: [Caitlin McTiernan](#), Just-Tech, LLC; [Shellie S. Reid](#), Manager, LSNTAP; [Ellen Samuel](#), Director of Consulting, Just-Tech; [Holly Stevens](#), Ph.D., Chief Data Officer, Legal Service Corporation*

*Description: This session will introduce conference attendees to the purpose and structure of the Legal Services National Technology Assistance Project's (LSNTAP) 2023 Data and Evaluation Toolkit. The panelists will discuss how to use the toolkit to build a strong data culture, including how organizations can demonstrate a commitment to improved use of data, can improve the quality of the organization's data, and can secure staffing for data work.*

### CLE RESOURCES

Legal Services National Technology Assistance Project  
Data and Evaluation Toolkit 2023



Legal Services National  
Technology Assistance Project

# Data and Evaluation Toolkit

## 2023



# Table of Contents

## Introduction

What is data analysis and why is it relevant for legal aid organizations?

What does this toolkit offer to legal aid organizations?

## Overview of Data Analysis Projects

How is this toolkit organized?

What resources are needed for a data analysis project?

When should an organization engage in a project involving data analysis?

What is the lifecycle of a data analysis project?

How does equity relate to data analysis?

## Defining the Analysis

What is the key question behind the analysis?

How and by whom will results be used?

What comes next?

## Collecting the Data

How is data categorized?

How is data collected?

How is data measured?

Why do data types and levels of measurement matter?

## Preparing and Analyzing the Data

What is data preparation and analysis?

How is data prepared and why does it matter?

How is data analyzed?

## Presenting and Learning from the Data

What are additional examples and resources for data visualization?

How should the audience approach data analysis results?

What's next?

## Building a Strong Data Culture

How can organizations improve their ongoing use of data?

How can organizations maintain high quality data?

How can organizations ensure they have adequate staffing for data analysis work?

## Overview of Evaluations

How do evaluations fit in to data analysis work?

What is program evaluation?

Why would an organization do a program evaluation and how do they start it?

What are Technology Innovation Grant (TIG) evaluations?

## Data Analysis Project Examples

Project Example 1 - Website Example

Project Example 2 - Language Access Example

Project Example 3 - Performance Review Example

## Acknowledgments

# Introduction

**Read This Section To:** Understand the scope and framework of this toolkit.

## **What is data analysis and why is it relevant for legal aid organizations?**

Data analysis is a process of examining data to produce useful information. Analyzed data may shed light on trends, challenge or validate existing assumptions, inform cost saving efforts, and promote greater efficiency. Legal aid organizations use data to make better-informed decisions on all kinds of questions, from operations and fundraising to staff supervision and program design.

Data analysis projects can help address questions such as the following:

- To what extent did this training impact participant knowledge of a legal topic?
- Which communities would benefit from a new outreach campaign?
- Where should a new office be located?
- Is this website user-friendly?
- Is a new program cost effective?
- How are caseloads distributed across attorneys?

A whole host of data sources could be used to answer the questions listed above. These sources include end-user surveys, client demographic data, program finance documentation, and participant interviews. This toolkit will go into detail on selecting appropriate data and where to go once data sources are identified.

## **What does this toolkit offer to legal aid organizations?**

This toolkit provides an overview of data analysis projects, tailored to legal aid providers with field-specific examples. This toolkit will be particularly helpful to those who:

- Want to use data to answer a specific question but are unsure how to get started,
- Are unfamiliar with data analysis processes and terms,
- Have an interest in improving data literacy at their organization, and/or
- Are interested in learning about evaluations and how they make use of data analysis.

Beyond a basic overview of common types of data analysis, this toolkit does NOT provide in-depth information on specific data analysis methods. Although this toolkit may provide enough context for providers to undertake simple data analysis efforts, more complex data projects will likely need trained data analysts or researchers. This toolkit provides a framework and shared vocabulary for engaging in data analysis projects but does not provide technical instructions on how to conduct data analysis.

There are many actions that legal aid providers of all sizes can take today to start improving their organization's relationship with data. The goal of this toolkit is for readers to walk away with a better understanding of data analysis in order to make greater use of data at their organization.

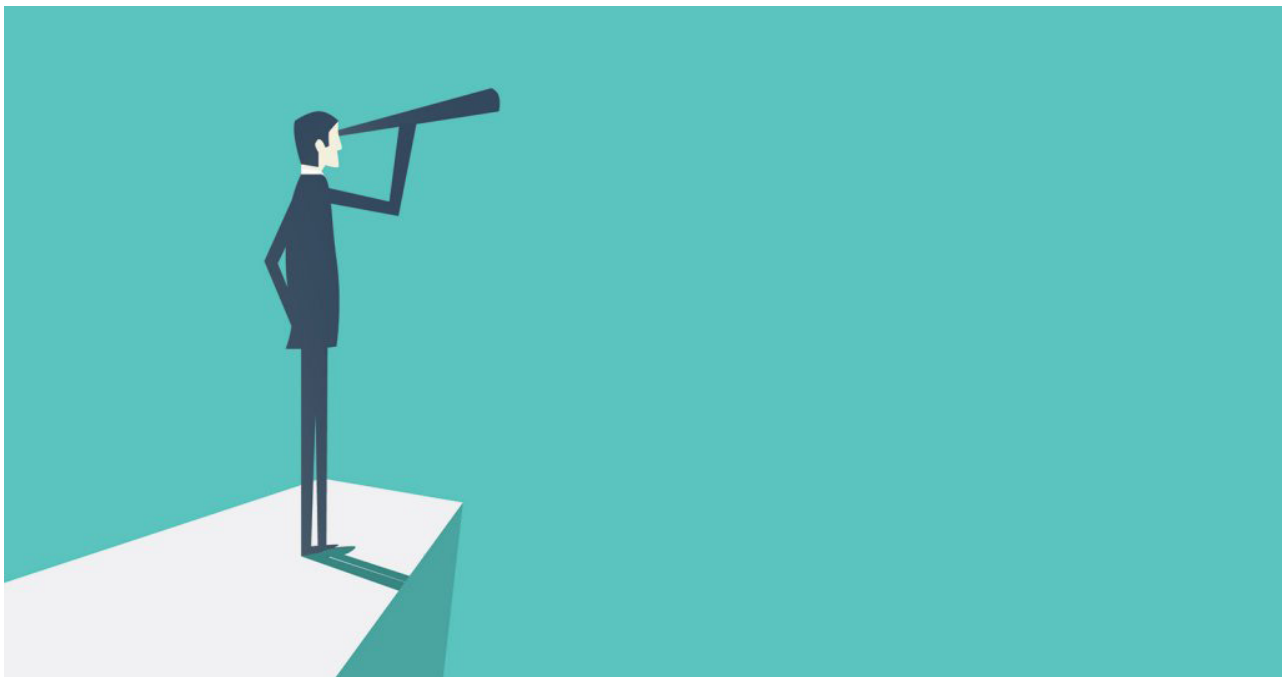


## How is this toolkit organized?

The toolkit is broken down into the following sections:

- 1. Overview of Data Analysis Projects:** Gain familiarity with the inputs and lifecycle of a data analysis project.
- 2. Defining the Analysis:** Walk through the questions, audiences, and logistical considerations that should be addressed at the start of a project.
- 3. Collecting the Data:** Review different data types, sources, and collection strategies, and understand why data type matters for analysis purposes.
- 4. Preparing and Analyzing the Data:** Explore concepts, vocabulary, and approaches relating to preparing and analyzing data.
- 5. Presenting and Learning from the Data:** Consider strategies for presenting data and processing analysis results.
- 6. Building a Strong Data Culture:** Review tips for developing a robust data culture and staffing for data analysis.
- 7. Overview of Evaluations:** Learn about common use cases and examples of evaluation in legal aid.
- 8. Data Analysis Project Examples:** Explore examples of data analysis projects on online intake systems, language access, and staff workloads.

## Overview of Data Analysis Projects



**How To Use This Section:** Gain familiarity with the inputs and lifecycle of a data analysis project.

### What resources are needed for a data analysis project?

Data sources, analysis techniques, costs, and resource needs will vary depending on the project. But at its core, every data analysis effort boils down to just a few components:

- **A Key Question:** This is the heart of the project, establishing what the learning objectives are from the data analysis effort. It is important that the question is actionable; at the project onset, there should already be an idea of how results may be used to impact the organization.
- **A Dataset:** An analysis can utilize existing data from public sources, case management systems, free web analytic tools, and more. New data can be collected by adding questions fields to existing processes, surveying users, conducting interviews, and other methods.
- **Tools to Prepare, Analyze, and Display Data:** There are countless free and low-cost tools to use for these purposes, such as Google Data Studio, Microsoft Excel and Power Bi, and Tableau. Some entities, such as Tech Soup, provide discounts on related software to nonprofit organizations.
- **An Individual or Team to Process the Data:** This work does not always require a formally trained data analyst or a paid external consultant. Depending on complexity, an existing staff member, local volunteer, college intern, or academic institution might be able to carry out the analysis.
- **An Engaged Audience:** Keep the audience in mind throughout every step of the project. Be sure to consider the audience's level of data literacy and intended use of results when presenting results.

A data analysis project can be made more complex by incorporating multiple questions, datasets, tools, or audiences. It is not always necessary to do so – even the simplest of data analysis projects can yield insightful findings that have an outsized impact on a decision. No matter the scale of the effort though, it is important to always make sure the data analysis project is accurate, ethical, and actionable.

### When should an organization engage in a project involving data analysis?

Put simply, an organization is ready to start engaging in data analysis work once it has identified a key question and an audience. Organizations should not begin a data analysis project until they have defined their key question, as this helps to narrow the scope of the project and prevent collecting data with no clear use. This question informs what data will be needed and what analysis and presentation strategies will be deployed.

Identifying the intended audience at an early stage will also help to limit the scope of the project and inform how results should be presented. Depending on the project, data analysis may be the sole source of information used to address a question, or it may be just one of many inputs (e.g., conversations with experts, historic precedents, availability of funds, etc.) that factor into a particular decision.

After selecting the key question and audience, an organization may need additional time to acquire funding, staff, and/or data to complete the data analysis effort. But when it comes to using data to inform a decision, this work should not be written off as too expensive or technically complex until the key question and audience are identified. The related data and resources needed to compile insightful results may already be on hand or easily attainable.

## What is the lifecycle of a data analysis project?

Data analysis efforts follow these five steps:

**Define the Analysis:** Determine what question(s) to answer, who wants the answer, and why it needs answering. This will shape the goals, methods, and presentation of the analysis.

**Collect Data:** Identify accurate, relevant data points that will help answer the key question. Implement sound collection practices to acquire any new data needed for the analysis.

**Prepare Data:** Review the dataset to ensure correct, complete data. Reformat data as needed to make for easier analysis and visualization.

**Analyze Data:** Make use of the numerous techniques available to describe and analyze the data, to detect patterns, trends, and insights that relate back to the key question.

**Present and Learn from the Analysis:** Display insights and trends that relate directly to the key question. Make sure results are clear, self-explanatory, and tailored to audience needs.

Data analysis projects do not always end with this last step. Some projects may result in ongoing monitoring or further analysis. Results from other projects may spur new key questions that can be addressed using additional data analysis. Rather than a straight line with a clear start and stop, the data analysis lifecycle is often more a continuous circle, as illustrated below:

Each of the five steps are reviewed in more detail in separate sections of this toolkit. Continue to the next section to learn more about identifying key questions and how they influence all parts of the data analysis project.



**Project Example:** Legal Aid of ABC (LAA) is looking to use data analysis to inform their end-of-year performance review process. Managers from each of their three internal teams (housing, family law, and immigration) have requested data from LAA's case management system to help them review staff performance in the past calendar year.

This example project will be referenced throughout the toolkit, in alignment with the lifecycle of a data analysis project. This example, as well as additional examples found in the final Data Analysis Project Examples section of the toolkit, will present various use cases, data collection methods, and analysis approaches that might be used by a legal aid provider for a data analysis project.

### How does equity relate to data analysis?

When there is data, there are equity implications. Equity and ethics are important to consider at every stage of the data analysis project lifecycle, such as:

- Will data collection or analysis impact historically marginalized and vulnerable populations? If yes, what efforts will be taken to reduce harm and address burden?
- How will demographic data be collected and used?
- Which methods and languages will be used for data collection?
- How will data collections methods be accessible (e.g., surveys, focus groups, interviews, etc.)?
- How will results be accessibly presented?
- How will results be disseminated back to involved communities and individuals?

Fortunately, there are many organizations that are doing great work on building data equity literacy, such as:



The [Data Equity Framework](#) and other tools provide a robust approach to equity-focused data analysis. The [Talking Data Equity](#) series provides ongoing opportunities to learn more.



Check out the [Why Am I Always Being Researched guidebook](#) to learn more on an equity-based, community-driven approach to research and analysis.



This group maintains a comprehensive list of resources for [developing](#) and [elevating data equity practices](#).

These resources are highly recommended to build a more robust understanding of what it means to have an equity-focused approach to data analysis work.

## Defining the Analysis



**How To Use This Section:** Walk through the questions, audiences, and logistical considerations that should be addressed at the start of a project.

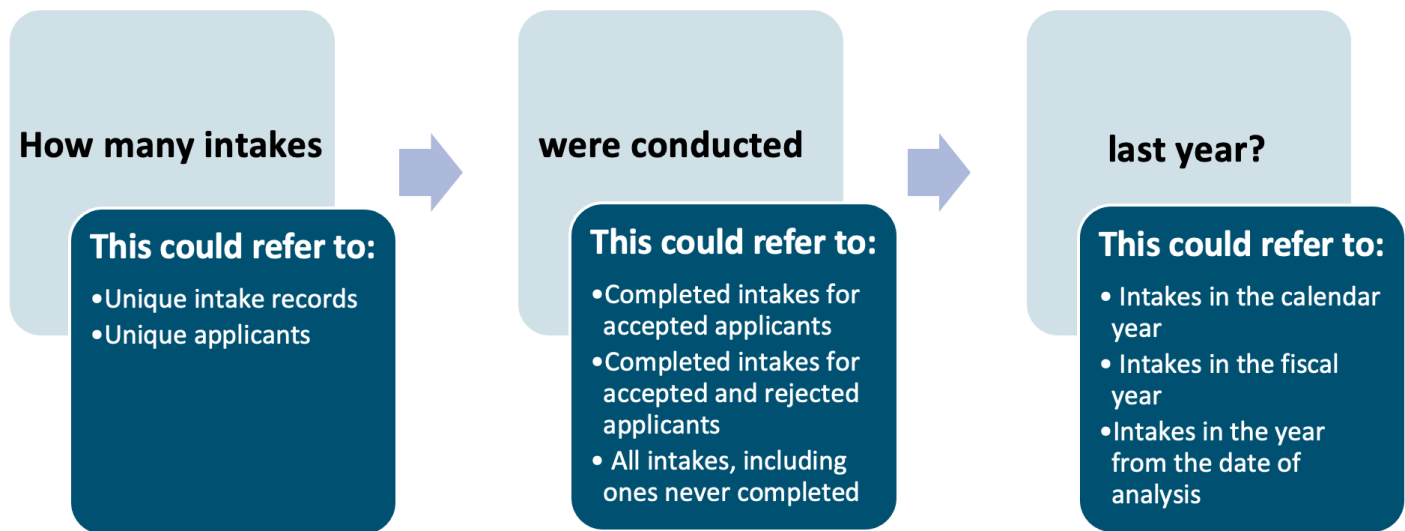
### **What is the key question behind the analysis?**

Data analysis is used to address a particular knowledge gap. The key question articulates this gap by communicating what is hoped to be learned from the data. Data analysis results may be used as the sole basis for answering the key question, or results may be just one input among many to fill this knowledge gap.

The key question needs to be established at the start because it anchors the analysis and informs what datasets, metrics, and analysis techniques will be used. A well-thought-out key question helps avoid unnecessary work, as all other steps of the data analysis project will be crafted specifically to answer it. While this toolkit refers to the “key question” as singular, note that many projects will have more than one key question that will be addressed through the analysis.

It is also important to think through different interpretations of the key question to make sure there is a common understanding among those involved with the project. For example, the question “How many intakes were conducted last year?” could be interpreted differently, as highlighted:





When crafting the key question, make sure to establish clarity around what exactly is being asked to ensure the analysis is tailored appropriately. A clearer version of this same key question could be: “How many unique intake records were started in calendar year 2021?”.

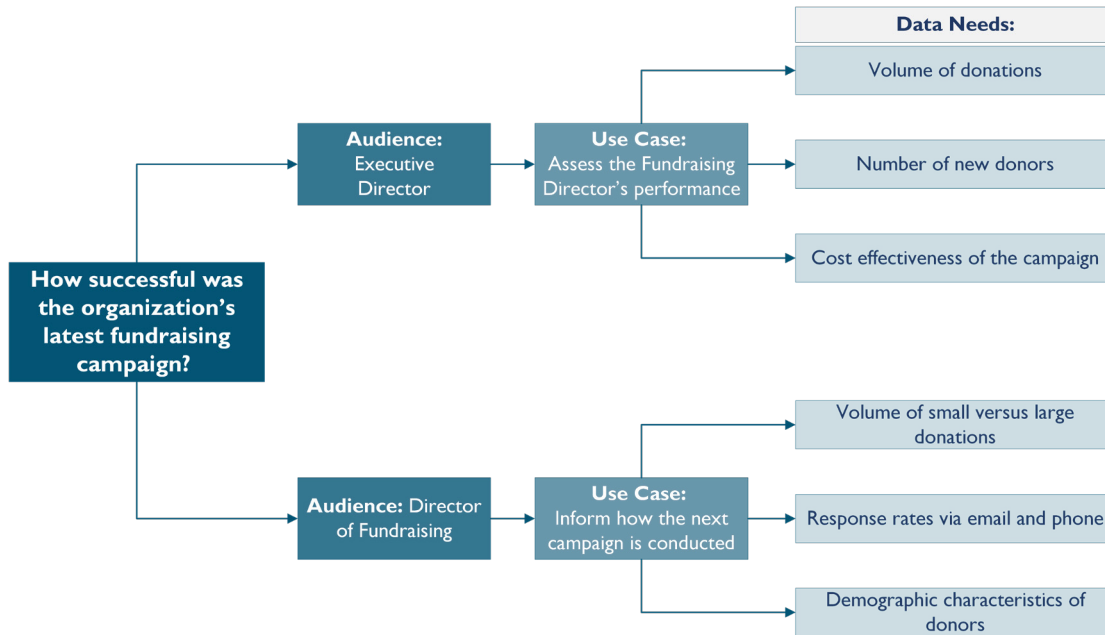
### How and by whom will results be used?

**How:** Data can provide insights into a particular question, but it takes additional planning to decide how to translate insights into actions. Once the purpose of the data analysis project is defined through its key question, consider how the answers to this question will be used. The goal here is to define actionable steps, so the knowledge gained through the analysis can be put to good use.

Data analysis projects come with different costs, though there are plenty of free tools, resources, and volunteers that can help to offset these costs. Examples of project costs may include the salary of staff conducting the analysis, the purchase of new software, and the time it takes for participants to complete data collection activities. Defining both the key question AND intended use of the data analysis project helps ensure that investment put into this work pays off. It may also help justify the costs of the project.

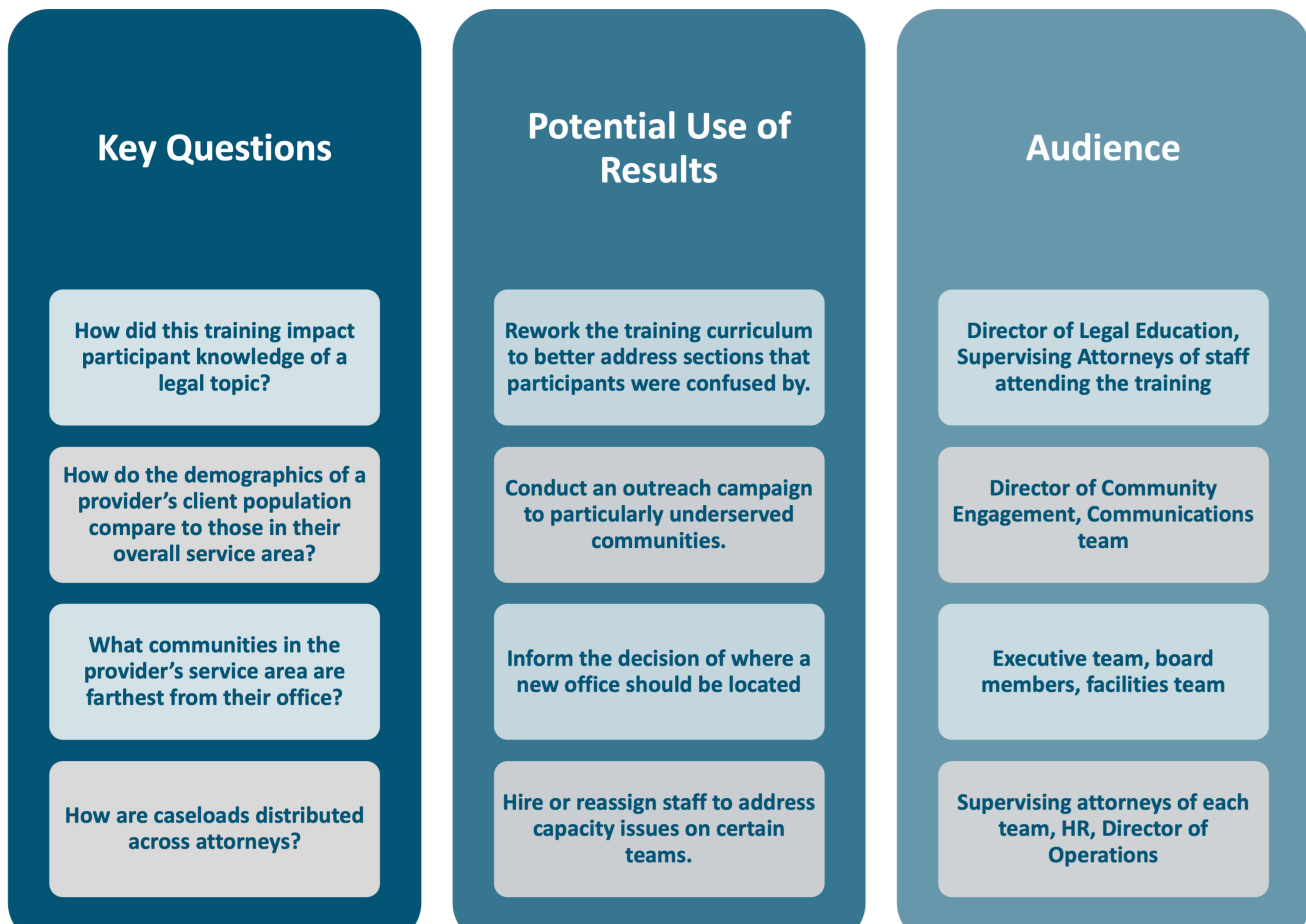
**By Whom:** It is also important to identify the audience for the analysis. Different audiences may have their own expectations around the metrics and analysis needed to answer identical questions, as expressed in the example below.

An organization’s Director of Fundraising and Executive Director (ED) both ask the same question: How successful was the organization’s latest fundraising campaign? The metrics and use case for this analysis will vary by the intended audience, as displayed here:



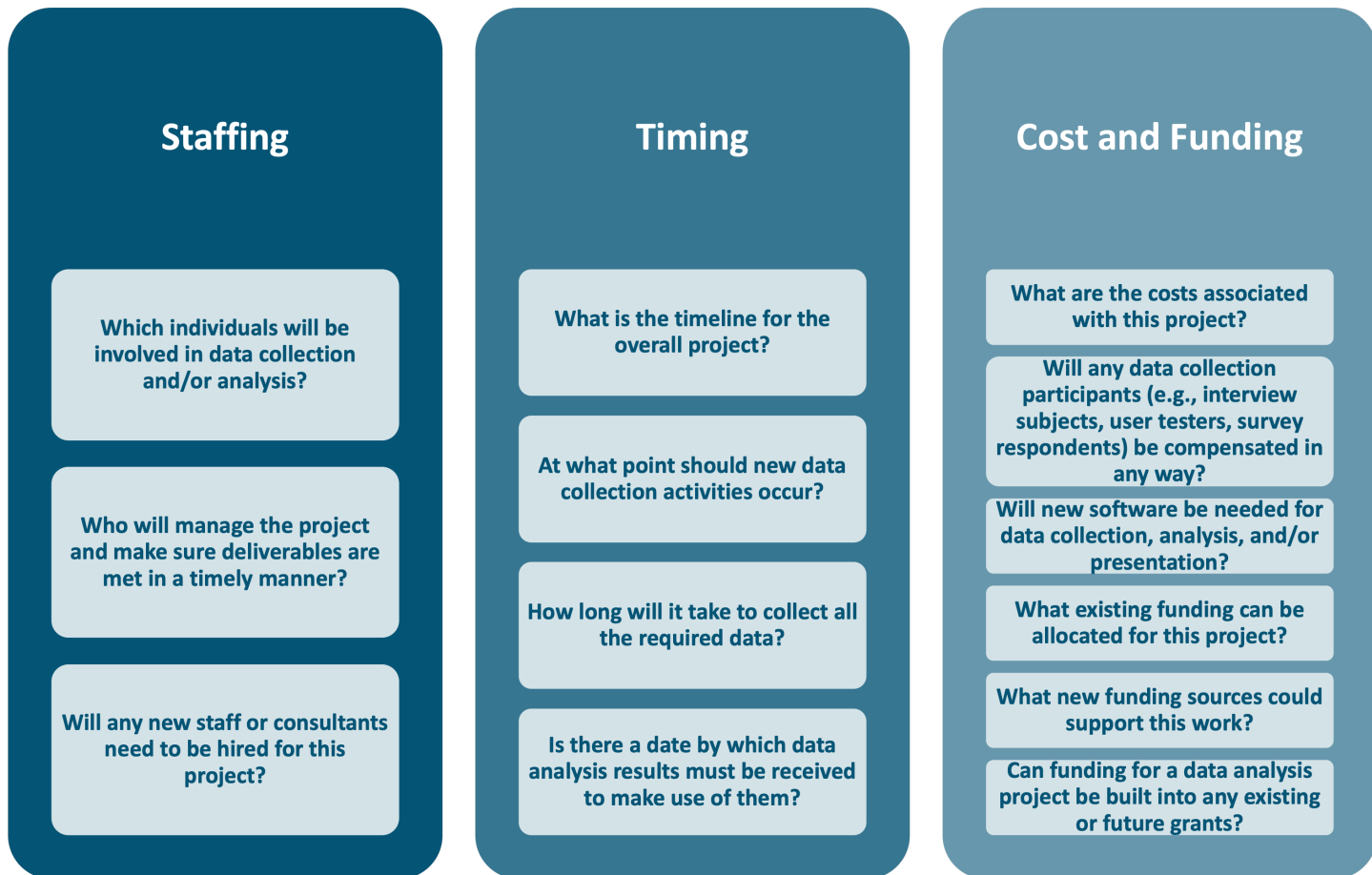
Additionally, the person raising the key question and the person who will put results into action are not always the same. For example, a Director of Operations may ask “Is this program cost effective?” but the program’s manager will be the one to adjust the program based on results. This is another reason to take time at the project start to identify the audience; if there are multiple audiences, the analysis will need to be comprehensive enough to be informative for all.

The following chart puts the above pieces together and shows possible uses and audiences for example key questions:

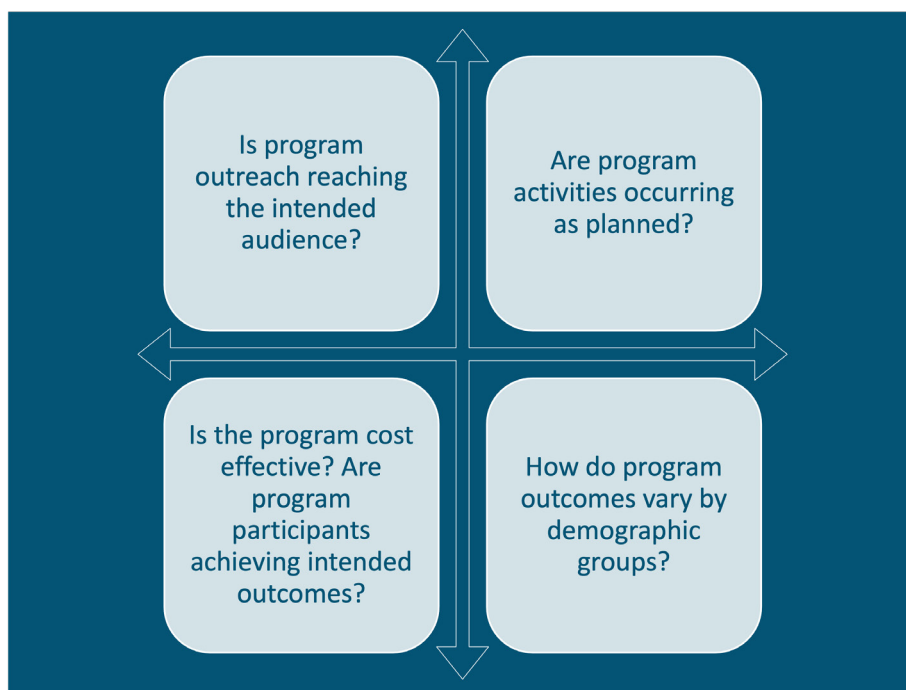


### What comes next?

Depending on the project and resources available to the organization, additional logistical questions may need to be answered before the next step, data collection, can begin. Questions to consider include:



Data analysis is often used to answer questions about a particular program or intervention, such as:



For a data analysis project focused on a particular program, consider creating a “logic model” to establish a common understanding of the program and its associated goals. A logic model documents the inputs, activities, and intended outputs and outcomes of the program. To learn more about logic models, check out the [LSNTAP Project Management Toolkit](#).

**Project Example:** Legal Aid of ABC (LAA) is looking to use data analysis to inform their end-of-year performance review process. Managers from each of their three internal teams (housing, family law, and immigration) have requested data from LAA’s case management system to help them review staff performance in the past calendar year.

**Key Question:** How did staff performance vary in the past calendar year, across the organization and within teams?

**Intended Use of Data Analysis:** LAA will use results to assess performance and inform decisions around salary increases, promotions, and bonuses for staff.

**Audience:** The supervisors for each of the three teams (housing, family law, and immigration) will review results and use them to inform their end-of-year performance reviews with supervisees. The ED will also review the results to ensure consistency across the organization in terms of compensation and promotion.

## Collecting the Data



**How To Use This Section:** Review different data types, sources, and collection strategies, and understand why data type matters for analysis purposes.

### How is data categorized?

After establishing the key question, use cases, and audiences of a project, the next step is to identify (1) what data will be used in the analysis and (2) how the data will be collected. This section provides background on selecting types of data, levels of measurement, and sources of data, to provide context for answering these two questions. It is important to note that data collection strategies, measurement options, and analysis techniques vary depending on the types of data selected for the analysis.

To start, there are two types of data: quantitative and qualitative.

**Quantitative data** is information that can be counted or measured and is expressed numerically. Examples of quantitative data include:

- A client's age
- A client's total income
- The date a case was closed



**Qualitative data** is information that is expressed through text, rather than numbers. Examples of qualitative data include:

- A client's gender
- The types of income a client receives
- The types of service provided on a closed case
- A client's comments on the service they received

Both quantitative and qualitative data can be analyzed, but the available analysis techniques will vary. Analysis may rely on quantitative data, qualitative data, or both (i.e., mixed methods). There is a large body of research around the different strengths and trade-offs for each type of data, which is beyond the scope of this toolkit. To learn more, both [McGill University](#) and [Eastern Michigan University](#) have resources that dive deeper into the differences around quantitative and qualitative research.

### How is data collected?

Legal aid organizations often already have a wealth of data at their disposal, such as:

- Case Data
- Client Data
- Timekeeping Data
- Payroll Data
- Grant Data

These data sources may be on-hand and useful for some data analysis projects. For other projects, new data may be needed to be able to address the key question. New data might be needed because this information was never collected previously, is outdated or incomplete, or was not collected in a way that aligns with the needs for the current project.

There are many strategies for collecting new data, including though:



**Surveys:** A survey is a series of questions designed to collect targeted information from a designated sample.



**Interviews:** An interview involves asking a respondent a series of in-depth questions, typically in a one-on-one conversation.



**Focus Groups:** A focus group is a moderated, planned discussion with multiple individuals who share perspectives and feedback on designated topics.



**Observations:** An observation effort involves watching and gathering details on the environment, activities, and/or individuals who are involved in a particular program or phenomenon.

For example, say a legal aid provider is interested in collecting feedback from applicants who use an online intake form to request assistance. Feedback was not collected previously, so the provider may choose to use any or all of the above strategies to gather this new data, such as by:

- Adding a survey after online intakes are submitted for applicants to share feedback on their experience,
- Conducting one-on-one interviews with applicants to discuss pain points with the application process,
- Facilitating discussions with a focus group of applicants on the online intake process, or
- Observing applicants as they navigate through the online intake process

Both quantitative and qualitative data can be collected through the above methods. However, the way data collection is structured may change depending on whether the aim is to produce quantitative or qualitative data. For example, the online application feedback survey might contain a question asking the applicant to estimate how many minutes it took for them to complete the application (quantitative data). That same survey might also have an open-text question asking the client for comments or suggestions about the application process (qualitative data).

When selecting data collection strategies for a project, stakeholders involved in the decision will usually have to balance a need for high-quality data with cost, timing, and other resource constraints. The [Center for Disease Control and Prevention](#) offers guidance on various strengths and weaknesses for each of these data collection methods. Expanded definitions of these data collection strategies can be found in the [International Organization for Migration's](#) Methodologies for Data Collection and Analysis guide and the [Urban Institute's](#) Data Collection project.

In addition to new data and data that organizations have on-hand, data analysis projects might use publicly available datasets from government agencies, courts, published research, and other external sources. As an example, [Census Bureau datasets](#) might be used to better understand the poverty levels and demographics of the communities in which a legal aid provider operates. There is a vast array of datasets accessible to the public, from sources such as:

- Research and data resources from the [Legal Services Corporation](#)
- Federal agencies like the [Bureau of Economic Analysis](#) and the [National Center for Education Statistics](#)
- State-specific data portals like the [California Open Data Portal](#), [Iowa Data](#), [Michigan's Open Data Portal](#), and the [Texas Open Data Portal](#)
- City-specific data portals like [NYC Open Data](#), [Los Angeles Open Data](#), and [Chicago Open Data Portal](#)

As part of the Justice in Government Project Toolkit, the [National Legal Aid and Defender Association](#) maintains a resource on many of these public datasets that may be useful for legal aid providers and researchers.

Whether new or existing data (or both) is appropriate to use will vary based on the purpose of the analysis. Collecting new data may be relatively more expensive and time consuming than using existing data, but new data is often necessary to gather the information needed to properly address a project's key question. This

step in the project lifecycle could take days, months, or even years depending on what data is needed and the selected data collection processes.

**How is data measured?**

Levels of measurement (i.e., Nominal, Ordinal, Interval, and Ratio) are different types of scales that explain how values for a particular data point relate to one another. A data point can be sorted into one of the four level of measurement scales, which determines the available options for how that data point can be analyzed. Listed

**Measurement Scale**

<b>Nominal Scale</b>	<b>Ordinal Scale</b>	<b>Interval Scale</b>	<b>Ratio Scale</b>
----------------------	----------------------	-----------------------	--------------------

**Definition**

<b>Nominal scales</b> are categories. These categories do not have numeric significance and the order of the categories has no real meaning.	<b>Ordinal scales</b> are also categories. Unlike nominal scales, these categories can be placed in an order that has meaning. However, the difference between categories cannot be quantified.	<b>Interval scales</b> are numeric and have a meaningful order. The difference between two values can be quantified, but there is no real "0" on the scale.	<b>Ratio scales</b> are numeric and have a meaningful order. The difference between two values can be quantified and there is a real "0" on the scale.
--	---	---	--

**Example**

<ul style="list-style-type: none"> <li>• Zip Code</li> <li>• Country of Origin</li> </ul>	<ul style="list-style-type: none"> <li>• Level of Service Provided (Advice, Brief Service, etc.)</li> <li>• Client's Satisfaction with Service Provided (from 1-Unsatisfied to 5-Satisfied)</li> </ul>	<ul style="list-style-type: none"> <li>• Date Case Closed</li> <li>• Temperature of the office</li> <li>• Intake Start Time</li> </ul>	<ul style="list-style-type: none"> <li>• Client's Age</li> <li>• Client's Total Annual Income (in \$)</li> <li>• Duration from Start to End of Intake Process (in Minutes)</li> </ul>
---	--	--	---

**Relevant For**

<b>Qualitative Data</b>	<b>Qualitative Data &amp; Quantitative Data Only</b>	<b>Quantitative Data Only</b>	<b>Quantitative Data Only</b>
-------------------------	--	-------------------------------	-------------------------------

below are definitions of the four measurement scales and examples of data points that align with each scale:

### Why do data types and levels of measurement matter?

Levels of measurement impact which types of analysis are feasible and how an analysis can best be visualized. Certain methods of analysis are not possible for all levels of measurement; this point will be reviewed in greater detail in the next section.

Practically speaking, levels of measurement (i.e., Nominal, Ordinal, Interval, and Ratio) influence the data selection process. This is because data needs to be collected in ways that align with multiple factors including the needs of the audience and the use cases for the analysis.

Let's look at two examples of why levels of measurement matter for data analysis projects and why it is important to be thoughtful around designing data collection efforts.

**Example 1:** A Grants Manager at a legal aid organization needs to know the average age of clients served in the past calendar year, to include this in a new grant application. If the organization collects the client's age in years (26, 35, etc.), the Grants Manager could determine the average client age because this uses a ratio scale. However, the Grants Manager would not be able to calculate this average if client age is collected using an age range (0-17, 18-29, etc.) because this uses an ordinal scale.

**Example 2:** A legal aid ED wants to know whether caseloads are distributed evenly across attorneys at their organization. The ED could examine the total number of cases closed by each attorney in the past year. However, on further reflection, the ED will realize that the total number of cases (i.e., one data point) alone does not paint the whole picture because attorney involvement is impacted by case attributes, such as the legal issue, level of service provided, and case complexity. Knowing this, the ED may want to break the analysis down further, perhaps by examining the following data:

- **Legal Issue:** This is measured on a nominal scale; these values cannot be ordered in a meaningful way. The ED could track total cases by legal issue (frequency) and most common case type (mode) for each attorney. However, the ED could not determine the mean legal issue for each attorney.
- **Level of Service Provided:** This is measured on an ordinal scale; these values are not numeric but could be ordered by intensity (from advice to full representation). The ED could track total cases by service level (frequency), most common service level (mode), and what the middle service level is (median). However, the ED still could not determine an attorney's mean level of service.
- **Duration Between Open and Close Date (in Days):** This is measured on a ratio scale; these values are numeric and there is true zero on the scale. Like Legal Issue and Level of Service, the ED could track the frequency, mode, and median for each attorney relating to case duration. The ED could also track the average case duration (mean) and the distribution of cases by duration (standard deviation), for each attorney.

To make quick comparisons, the ED may want to review only a couple metrics, such as each attorney's total closed case count and average case duration from open to close. However, if the ED wants to dive deeper into the data, they may look for analysis grouped by legal issue, level of service provided, or other case and client attributes.

To summarize, when establishing what data points will help in answering the key question, keep in mind equitable and ethical ways to collect the data (see Data Equity section above for related resources), how the data can be analyzed, and whether the structure of data collection aligns with analysis needs.

Once all data collection activities have been completed, the project can move forward into the preparation and analysis phase, which is discussed in the next section of this toolkit.

**Project Example:** Legal Aid of ABC (LAA) is looking to use data analysis to inform their end-of-year performance review process. Managers from each of their three internal teams (housing, family law, and immigration) have requested data from LAA's case management system to help them review staff performance in the past calendar year.

**Data Collection Plan:** LAA will utilize one existing dataset for this effort - data from LAA's case management system from the current calendar year, for cases open at any point during this period.



## Preparing and Analyzing the Data



**How to Use This Section:** Explore concepts, vocabulary, and approaches relating to preparing and analyzing data.

### What is data preparation and analysis?

Before diving into the analysis step, datasets first need to be prepared. This process, referred to as data preparation or data cleaning, involves identifying and addressing inaccurate, irrelevant, duplicative, and incomplete data. It may also involve reformatting data in ways more convenient for analysis. Data is often messy and can be riddled with data entry errors such as typos, misspellings, and missing responses. This data preparation step is critical to make sure the analysis is based on data that is as accurate and complete as possible.

Once data is prepared, it is now ready for analysis. Data analysis involves processing prepared data to uncover trends, patterns, and other useful information. These analysis efforts aim to provide insights which can be used to address the project's key question.

A number of different tools are commonly used for data preparation and analysis, including: [Excel](#), [Google Sheets](#), [Python](#), [R](#), [SAS](#), [SPSS](#), and [Stata](#). Check out the guides and resources from [New York University](#), [Princeton University](#), and [University of California Irvine](#) to learn more about each of these tools.

The remainder of this section introduces key vocabulary and concepts relating to data preparation and analysis. This discussion is geared towards those unfamiliar with this work or those who need a refresher – it will not teach all the technical information necessary to carry out full analysis efforts. For those interested in learning about specific data analysis techniques, check out [Coursera](#) and [Harvard University](#) which both offer a range of free, online classes on this subject.

### How is data prepared and why does it matter?

As a best practice, data preparation starts with duplicating and saving a complete copy of the original dataset. This ensures that the raw, unedited data is always available for reference as needed.

The following example dataset from a case management system will be used to review common data preparation steps. Please note that this example, with only 9 rows of data, is very simplified; data preparation will often involve hundreds to millions of data points.

To prepare data, look out for errors, inconsistencies, and missing or irrelevant data. Examples of steps to take to clean the above dataset include:

Case ID	Intake Date	Status	Client's City of Residence	Client Age	Gender
1363	3/21/2022	Open	Brooklyn	28	Male
6242	2/1/1022	Closed	Manhattan	113	Female
8335	8/11/2022	Open	Bklyn	75	Unknown
7345	6/17/2022	Closed	Queens	28	Female
3992	5/21/2022	Closed	Queens	41	Female
9234	10/18/2019	Open	Bronx	62	Female
2355	1/9/2022	Open	Manhattan	67	
7345	6/17/2022	Closed	Queens	28	Female
4121	9/2/2022	Open	Bronx	43	Male

- Correcting Data Entry Errors:** Case #6242 is listed with a 1022 intake date, a clear typo that can be changed to 2022. This is a reasonable assumption to make, because the because the case management system was not available until the 2000s (so the intake date would not be 1922). When it is unclear how to correct an apparent typo, do not change the data because the adjusted information may be incorrect.
- Resolving Inconsistent Entries:** Respondents manually type in the client's city of residence, resulting in two cases (#1363 and #8335) with different spellings for the same location, "Brooklyn" and "Bklyn". Decide on a standard way to list this city name and record it for both cases.
- Addressing Missing Data:** The client's gender is unknown for cases #8335 and #2355. If possible, reach out to the data entry source to complete the missing data point. If this is not possible, decide on a standard way to mark missing gender values and adhere to that rule for all cases where client gender is not known.
- Reviewing Outliers:** The client's age is listed as 113 for case #6242, which is almost certainly a typo. If possible, go back to the source to confirm if this information is correct and adjust if necessary. If this is not possible, one approach would be to remove client age as a data point for case #6264 and treat this as missing data. In this example, the attorney was able to confirm the client's correct age was 13, so 113 can be replaced with 13.

- **Removing Irrelevant Data:** The analysis will focus on cases with 2022 intake dates, so case #9234 and its entire row of data can be removed because the intake took place in 2019.
- **Removing Duplicate Data:** Case #7345 appears twice in the dataset, with the exact same information in all cells. Only one row per unique case is needed for the analysis, so one of the two rows for case #7345 can be removed.

Data preparation is critical because it ensures the analysis is as accurate as possible. If the above dataset is not properly cleaned, there are clear, negative impacts on the analysis as shown below:

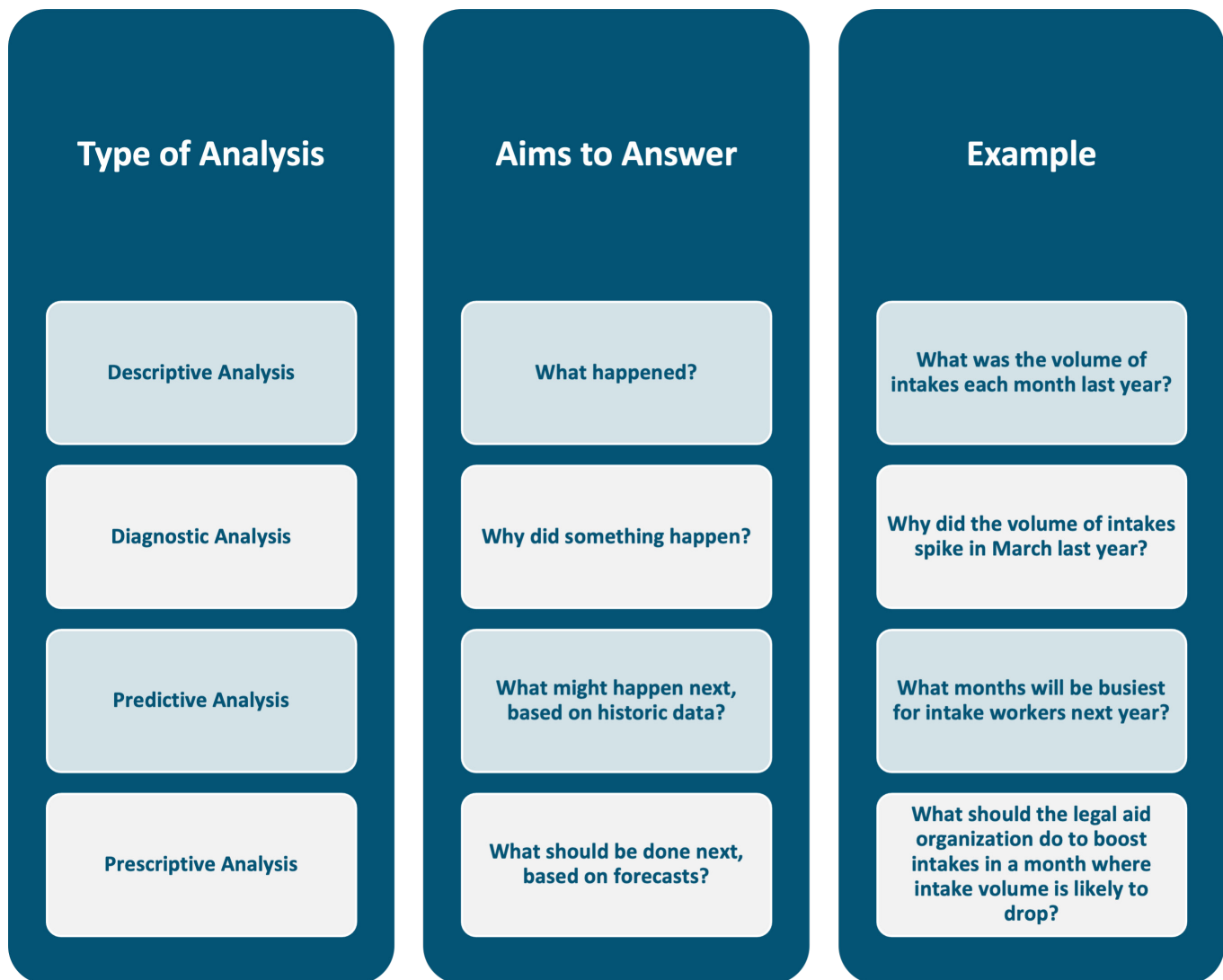
- Failing to address data entry errors and inconsistencies can result in incorrect analysis. For example, if “Brooklyn” and “Bklyn” are left as is, this would appear as two distinct locations in a summary of total cases by city.
- Results can be skewed if outliers are not addressed. For example, if the client age of 113 is left in, the mean client age in the data set would be 53. If it is replaced with the correct value (13), the mean client age is 42.
- If irrelevant or duplicative data is not removed, incorrect analysis may result. For example, if the 2019 case is left in the dataset, this would inflate the total intake in 2022 by one.

While the exact steps needed to properly prepare data will vary depending on the source data, the examples above represent common approaches taken to produce a clean dataset. The following is an example of the dataset listed above, now prepared and ready for analysis:

Case ID	Intake Date	Status	Client's City of Residence	Client Age	Gender
1363	3/21/2022	Open	Brooklyn	28	Male
6242	2/1/2022	Closed	Manhattan	13	Female
8335	8/11/2022	Open	Brooklyn	75	Unknown
7345	6/17/2022	Closed	Queens	28	Female
3992	5/21/2022	Closed	Queens	41	Female
2355	1/9/2022	Open	Manhattan	67	Unknown
4121	9/2/2022	Open	Bronx	43	Male

## How is data analyzed?

Data analysis refers to the processing of prepared data to uncover trends, patterns, and other useful information. There are four categories of analysis that aim to address slightly different questions: descriptive, diagnostic, predictive, and prescriptive. The following chart highlights the different aims of these types of analysis:



See the [Harvard Business School's](#) guide on data analysis for improved decision making for more information on these four types of data analysis.

The remainder of this section focuses on the common components of descriptive analysis, which plays a role in the simplest to the most complex data analysis undertakings. Advanced data analysis techniques are beyond the scope of this toolkit, which are typically needed in diagnostic, predictive, and prescriptive analyzes.

Descriptive analysis uses data to highlight trends and phenomena that are relevant to addressing the key question of the project. Descriptive analysis includes measures of:

- **Frequency:** Describes how often values occur within a dataset (e.g., totals, ratios).
- **Central Tendency:** Describes the central or most typical value of a particular data point (e.g., mean, median, mode).
- **Dispersion:** Describes the spread of the data from the center (e.g., range, standard deviation).
- **Position:** Describes where a value falls in the distribution of all values in the dataset (e.g., quartiles, percentiles).

These measures, in combination, shed light on different aspects and trends of a dataset. To demonstrate this, let's return to the prepared (or cleaned) dataset from the previous section, which will now be used to demonstrate various descriptive statistic calculations:

Case ID	Intake Date	Status	Client's City of Residence	Client Age	Gender	Unique Client ID
1363	3/21/2022	Open	Brooklyn	28	Male	#12
6242	2/1/2022	Closed	Manhattan	13	Female	#56
8335	8/11/2022	Open	Brooklyn	75	Unknown	#74
7345	6/17/2022	Closed	Queens	28	Female	#18
3992	5/21/2022	Closed	Queens	41	Female	#92
2355	1/9/2022	Open	Manhattan	67	Unknown	#77
4121	9/2/2022	Open	Bronx	43	Male	#41

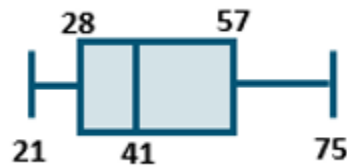
First, let's calculate various descriptive statistics for client age, using the above dataset of 7 unique clients:

Name of Measure	Definition	Example Measure from the Dataset	Calculation
Total	A count of values	4 clients are 40+	Counted the number of clients with an age of 40 or above
Ratio	A comparison of two or more values	57% of clients are 40+	Counted the number of clients with an age of 40 or above and divided by the total number of clients: $4 / 7$
Mean	The average value	42 years	Added all client ages and divided by the total number of clients: $[13 + 28 + 28 + 41 + 43 + 67 + 75] / 7$
Median	The center value	41 years	Arranged all client ages from lowest to highest and located the middle number: 13, 28, 28, 41, 43, 67, 75
Mode	The most common value	28 years	Reviewed the client ages to identify which age was the most frequently recurring
Range	The difference between the maximum and minimum value	62 years	Subtracted the maximum and minimum ages: $75 - 13$
Standard Deviation	An indication of how spread out the data is relative to the mean	22 years; each age deviates from the mean by 22 years, on average.	Coursera provides a free instruction video on calculating standard deviation.*

\*[Coursera Training on Standard Deviation](#)



Data can also be summarized by quartiles, which involves sorting the data in numeric order and dividing it into four equal parts. For the above dataset: 21 is the minimum value, 28 years is the 1st quartile, 41 years is median (2nd quartile), 57 is the 3rd quartile, and 75 is the maximum value. Quartiles are another helpful measure to showcase the spread of the data, and can be showcased with a box plot, as exemplified below:



These different measurements - frequency, central tendency, dispersion, and position – identify spread, outliers, common values, and potential relationships of data points. Descriptive analysis summaries often include measures across these four categories, to help showcase different aspects of the data.

Descriptive analysis can involve looking at one data point on its own (univariate analysis), as with the above example on client age. It can also reference a combination of two data points (bivariate analysis) or several data points (multivariate analysis). Using the example dataset from above, a simple example of bivariate descriptive analysis is the following table of clients by gender and county of residence:

	Female	Male	Unknown	Total
Bronx	0	1	0	1
Brooklyn	0	1	1	2
Manhattan	1	0	1	2
Queens	2	0	0	2
Total	3	2	2	7

Keep in mind that descriptive analysis measures (median, mean, etc.) are not always feasible, depending on the data point’s level of measurement (nominal, ordinal, interval, or ratio).

For example, take a sample of 10 people, of which 3 are female and 7 are male. Gender uses a nominal scale. Nominal scales are categories, which have no numeric significance and have no meaningful order. For this sample, it would be impossible to calculate the “mean” gender. However, it would be possible to calculate the mode (the most common value), which would be male (since 7 of the 10 people are male).

The chart below depicts this relationship between level of measurement and analysis options in greater detail, highlighting whether it is possible to calculate a given metric for each measurement level:

	Total	Mode	Median	Mean	Standard Deviation
<i>Definition of Measure</i>	<i>A count of values</i>	<i>The most common value</i>	<i>The center value</i>	<i>The average value</i>	<i>An indication of how spread out the data is relative to the mean</i>
<b>Nominal</b>	Yes	Yes	No	No	No
<b>Ordinal</b>	Yes	Yes	Yes	Sometimes	No
<b>Interval</b>	Yes	Yes	Yes	Yes	Yes
<b>Ratio</b>	Yes	Yes	Yes	Yes	Yes

Moving forward in the data analysis project, the next section reviews strategies to present on and learn from the analysis.

**Project Example:** Legal Aid of ABC (LAA) is looking to use data analysis to inform their end-of-year performance review process. Managers from each of their three internal teams (housing, family law, and immigration) have requested data from LAA's case management system (CMS) to help them review staff performance in the past calendar year.

**Data Analysis Examples:** Using data from LAA's case management system from the current calendar year, team managers have requested the following metrics be included in the analysis:

- Counts of cases closed during 2022 and open as of the end of 2022, by primary advocate
- Counts of cases closed by month and primary advocate
- Counts of closed cases by level of service provided and primary advocate
- Average duration of closed cases from intake to close (in days), by level of service provided and primary advocate
- Counts of cases by legal issue and primary advocate

**See the attached spreadsheet, Performance Review Project Example Data, for a deeper dive into these data analysis examples. This spreadsheet contains several tabs which showcase descriptive analysis and visualization examples for LAA's CMS data from the current calendar year.**

# Presenting and Learning from the Data

**How to Use This Section:** Consider strategies for presenting data and processing analysis results.

## What are different ways to present the analysis?

When choosing what to include in a presentation, consider questions, such as:





- What is important for the audience to know to answer the key question?
- What analysis findings challenge or bolster existing expectations?
- What trends or outliers are most relevant for the end audience?
- What information does the audience need to be able to fully contextualize results?



Make sure to consider the audience when presenting analysis findings. Presentations do not always need to be visual – it may be possible to succinctly convey a finding in writing, which could be an efficient way to communicate to the intended audience. Tables, which display information across rows and columns, are another popular format that can be used to communicate findings.

Aside from text and tabular formats, there are a range of visualizations that provide convenient, useful ways to convey important findings and examine potential relationships between multiple data points. Review a few of the most common types of visualizations below and take a deeper dive with Tableau’s Guide to Charts:

Good presentations adhere to the following principles:

Bar Chart	Line Graph	Pie Chart	Maps
 <ul style="list-style-type: none"> <li>• Used to compare values across different categories</li> <li>• Avoid using with continuous data (data on an infinite scale, like time)</li> </ul>	 <ul style="list-style-type: none"> <li>• Used to highlight trends over time (or some other continuous progression)</li> <li>• Avoid if there are many categories that would result in lots of lines, which can be hard to distinguish from one another</li> </ul>	 <ul style="list-style-type: none"> <li>• Used to show values across different categories relative to the whole</li> <li>• Avoid using if there are 5 or more categories</li> <li>• Avoid if categories make up similar shares (hard to distinguish slice sizes)</li> </ul>	 <ul style="list-style-type: none"> <li>• Used to display data across geographic locations</li> <li>• Some maps use color or patterns to indicate data trends across geographic boundaries like states, zip codes, or countries (choropleth maps, heat maps)</li> <li>• Other maps use symbols that vary in size over specific locations to indicate data trends</li> </ul>

- **Keep it simple:** Stick to consistent formatting, use [colors](#) and symbols strategically, and avoid adding unnecessary complexity. In other words, don't include three charts when one sufficiently conveys the intended message. Tools such as [I Want Hue](#) and [Color Brewer](#) can help in selecting clear, aesthetic color palettes for visualizations.
  - **Include context:** Add clear titles, labels, legends, and chart descriptions where appropriate so the audience can follow along easily. Include all the information the audience needs to fully understand the analysis somewhere within the presentation.
  - **Be accessible:** Adhere to color contrast rules for color blind individuals, avoid relying on color as the sole indicator in a chart, include descriptions of visualizations for screen readers, and take other appropriate actions to make presentations accessible for all. To learn more, check out the [Digital Accessibility Guide](#) from the University of North Carolina's School of Medicine IT, which includes guidance on color schemes for color blind individuals.
  - **Use appropriate chart types:** Make sure to select appropriate visualizations for the data being displayed. The [Data Visualization Catalogue](#) and [From-Data-to-Viz](#) both contain primers on dozens of data visualization types and their functions. The [University of California Berkley Library](#) provides guidance on this process of selecting appropriate visualizations or charts.

Dashboards offer a way to display a collection of data summaries and visualizations. Dashboards are often interactive and exploratory - the user can typically apply filters and focus on different segments of the data according to their needs and interests. Dashboards are generally for ongoing use and get refreshed with up-to-date data at regular intervals.

To learn more about dashboards visit the National Council of Nonprofit's [dashboard resource guide](#) . Scroll to the end of this section for an example dashboard, based on the Performance Review project example referenced throughout the toolkit.

### What are additional examples and resources for data visualization?

There are plenty of tools available to create data visualizations and dashboards, including [Datavrapper](#), [Excel](#), [Google Data Studio](#), [Infogram](#), [Plotly](#), [PowerBi](#), [QGIS](#), and [Tableau](#).

For more ideas and inspiration on data visualization, look to the following public resources from members of the legal aid community:

- [U.S. Funding for Legal Aid](#): The American Bar Association created a dashboard using Tableau on the sources and amounts of funding for legal aid in the United States.
- [COVID-19 Eviction Surveillance](#): The Eviction Solidarity Network in Travis County, Texas produced a dashboard using ArcGIS on eviction filings, hearings, and outcomes in their county.
- [King County Eviction Tracker](#): The King County Bar Association in King County, Washington developed a dashboard using Power Bi on eviction filings and outcomes in their county.
- [Client Services Dashboard](#): Community Legal Aid Services publicizes PDF files showcasing case and client data on a quarterly basis.
- Legal Information Website Usability Study Results: Michigan Legal Help conducted a usability analysis for their website; results and interpretations of findings are noted in this slide deck. See the bottom of this page to access the file with these results. <sup>1</sup>

---

1 [Legal Information Website Usability Study Results](#)  
[Performance Review Project Example Data](#)

Check out the following webinars to learn more about data visualization strategies and tools:

- [Data Visualization Tools](#): This LSNTAP webinar presents a series of cost-effective tools for data visualizations and reviews the theory behind good data visualization design.
- [Visualizing Your Data Through Dashboards](#): This LSNTAP webinar offers a step-by-step walk-through for creating a legal aid dashboard.
- [PowerBi Community Integrations Call](#): This LegalServer presentation demonstrates an approach by the Legal Aid Foundation of Los Angeles to connect case management system data with third-party data analysis and visualization tool.

Crafting these presentations is essentially a work of storytelling. Focus the presentation on the most salient and relevant findings, instead of displaying all possible data. Keep the end audience in mind, including their level of data literacy, so they can glean insights from the analysis.

### **How should the audience approach data analysis results?**

Data analysis involves selecting relevant data, conducting appropriate analyzes, and presenting pertinent results in a format accessible to the final audience. By adhering to this approach, the audience will have clear end results they can easily use to inform their intended next steps. A few tips to keep in mind:

- **Expect the Unexpected**: Analysis results may not be what was expected or desired – if there was absolute certainty about the trend or phenomena in question, there wouldn't be a need for the analysis. Whether expected or desired, results present a learning opportunity all the same.
- **Focus on Next Steps**: Ideally, the use case for results will be articulated at the very start of the project. Reflect on this specification, adjust as needed, and determine the best course of action based on the results of the analysis. Keep in mind that a data analysis effort may not be one-time-only undertaking; results may lead to new questions or analysis needs.
- **Create a Data Dictionary**: Data dictionaries provide a concise summary of the meaning, purpose, and context behind the data points used in the project's data set. This can be referenced by the audience, to make sure there is a consistent understanding of the underlying data. See the U.S. Department of the Interior's [Data Dictionary Guide](#) for more information.
- **Disseminate Results to Participants**: Those who participate in the project, such as those responding to surveys or taking part in focus groups, have contributed their time, information, and/or resources for these efforts. Sharing results with them is a way to both acknowledge these efforts and encourage future participation.

**What's next?**

This concludes the portion of the toolkit centered on the lifecycle and activities of a data analysis project. Continue onwards for the remaining sections:

- **Building a Strong Data Culture:** Recommendations to build a robust data culture at legal aid organizations, including strategies to maintain clean data and staff data work
- **Overview of Evaluations:** A review of evaluations, including a subset that makes use of data analysis and specific research techniques to inform and improve programs
- **Data Analysis Snapshots:** Examples of data analysis projects which walk through the various steps of the data analysis lifecycle

**Project Example:** Legal Aid of ABC (LAA) is looking to use data analysis to inform their end-of-year performance review process. Managers from each of their three internal teams (housing, family law, and immigration) have requested data from LAA's case management system (CMS) to help them review staff performance in the past calendar year.

**Example Visuals Included in Presentation:** Based on the analysis conducted for LAA's CMS data from the current calendar year, the follow visuals were included in the final presentation for this project:

- Bar chart of total cases closed in the year and open at the end of the year, by primary advocate
- Line graph of total cases closed each month, by team
- Pie charts of cases handled by primary advocate and legal issue

**Decisions Informed by the Analysis:**

- LAA supervisors were able to identify workload trends and spark deeper conversations with staff around their performances.
- LAA supervisors determined there was an ongoing need for access to this information, so staff built out a dashboard with related workload metrics that automatically refreshes with up-to-date case data.

**A screenshot of this dashboard is included below. When LAA staff access the dashboard file, they have the option of filtering results based on team and/or primary advocate (seen in the top left section of the dashboard).**



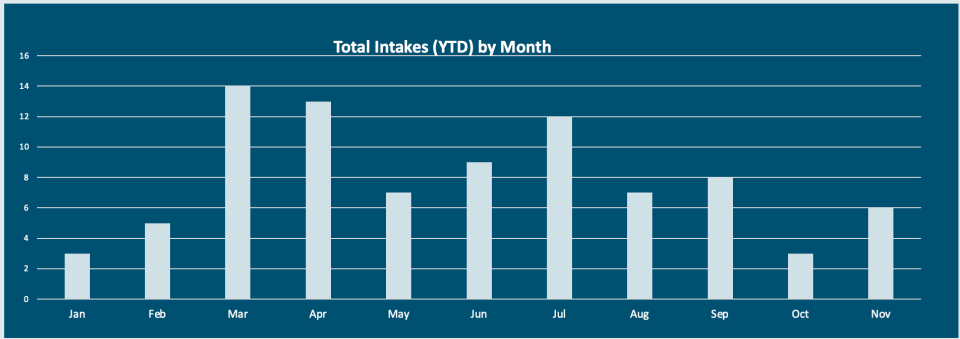
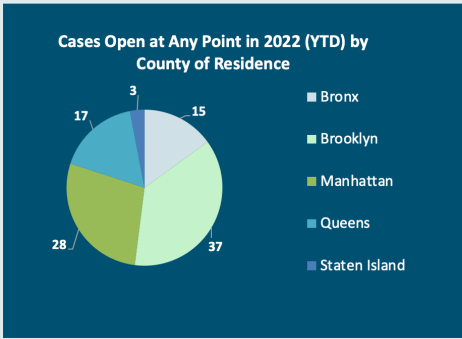
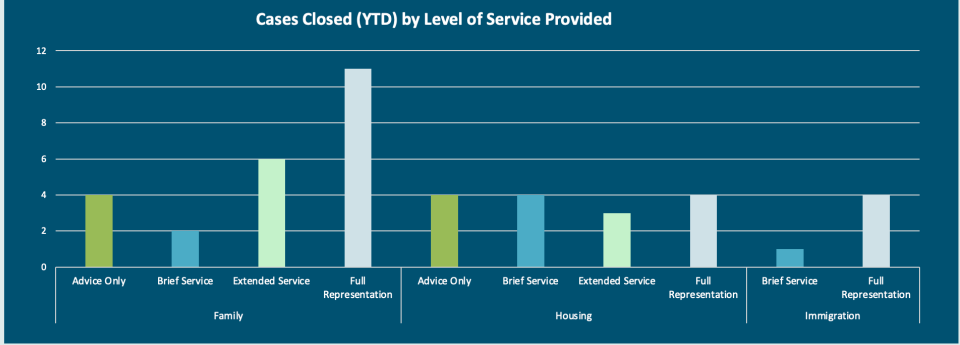
Select teams and/or staff members in the filters below. Filters will be applied to all dashboard results.

**Assigned Program**

- Family
- Housing
- Immigration

**Primary Advocate**

- Cohen, Sammy
- Frances, Sam
- Frye, Nick
- Garcia, Cora
- Jennings, Rachel
- Li, Emily
- Ramirez, Jo
- Smith, Sally
- Wiley, Alex



To explore the dashboard and data visualization example further, access the attached spreadsheet, [Performance Review Project Example Data](#). This spreadsheet provides a deeper dive into the descriptive analysis, visualization examples, and dashboard for this example project.

## Building a Strong Data Culture



**How to Use This Section:** Review tips for developing a robust data culture and staffing for data analysis.

### How can organizations improve their ongoing use of data?

The use of data to routinely inform decisions and operations is referred to as a data culture. There are a few core components to data culture including a commitment to using data, staffing for data analysis work, and practices to maintain and make use of organizational data regularly.

Legal aid providers fall across a wide spectrum in terms of their current use of data. Some organizations are beginning to think about how to use data, while others may have staff dedicated to preparing and analyzing data. No matter where a provider is, there are many actions they can take to strengthen their commitment to making better, more informed use of data.

Some examples include the following suggestions:

- **Encourage Staff Buy-In:** Staff at all levels participate in data efforts, from paralegals collecting intake data to grants managers reporting on grant data, to supervisors reviewing team caseloads. For many organizations, it can be a big shift from under-utilizing data to routinely relying on data. To help encourage staff throughout this transition, make sure to stress the how and why behind the change. How will data be used moving forward? How will processes change as a result? Why is this change needed and how is it expected to help the organization? Answering these questions clearly and repeatedly will demonstrate to staff the benefits of data-driven work and the organization's commitment to this practice. Consider utilizing change management strategies such as those discussed in this [Harvard Business Review article](#).

- Another practice to encourage staff buy-in is to reward those who are adhering well to data practices. Identify which staff are maintaining accurate, timely, and complete data and which staff are using data to inform their work. Rewarding good performance on data efforts can encourage greater participation throughout the organization, as well as acknowledging staff contributions to this change.
- **Make Data More Prominent:** If staff can't access data, they can't use it. Minimize the burden on staff to seek data out by making the data more accessible in the first place. This transparency demonstrates an organizational commitment to utilizing data routinely. Efforts to make data more prominent will take different forms, such as broadly disseminating analysis results or reports, discussing metrics in staff or team meetings, including data summaries in internal newsletters, and building dashboards for ongoing use.
- **Address Data Literacy:** No matter how accessible data is, staff are likely to ignore or misuse it if they do not understand how to properly interpret data. Start by assessing data literacy within the organization. To inform this discovery work, consider asking staff to take a data literacy assessment such as [this test](#) offered by the Data Literacy Project. From there, tailor trainings, webinars, and other educational events for staff to help boost knowledge around data analysis vocabulary, techniques, and meanings.
  - Staff who are more familiar with data analysis can be resources for these educational activities. Providers could also turn to volunteers, local academic institutions, or external partners for training support and resources.
- **Ask More Questions:** Recognize that data analysis work is not a one-time-only activity. Legal aid providers face operational and programmatic decisions all the time that can be informed by data. Don't stop with just one project; consider other ways that data can be used or what new analysis questions could be asked based on existing results.

The above actions all help to foster a culture of data use and encourage data-informed, continuous improvement. Providers can also strengthen a data-driven culture by maintaining clean, quality data for routine use; strategies relating to data quality maintenance are discussed in the next section.

### How can organizations maintain high quality data?

Legal aid providers typically have a large quantity of data on hand from case management systems ("CMS") and other data repositories. A strong data culture is bolstered by frequent review and upkeep of this data, to help ensure that the data being collected routinely aligns with organization's needs. This data can be used for ongoing analysis but is only as useful as it is clean. Consider asking the following questions when reviewing the quality of case management system data, one of the most common data sources for legal aid providers:

- Is there missing or outdated data?
  - **Example:** Case handlers are behind on closing their old cases, so it is not possible to get an accurate count of active, open cases by staff member.
- Are there structural issues with the way any data is collected?
  - **Example:** A mandatory intake field "Do you have a disability" only has response options of "yes" and "no". Intake workers select "no" to continue with the intake when the answer is unknown to them, because a response is required. This clouds the meaning of "no", to include those who do not have a disability AND those who did not answer the question.

- Are respondents able to select values that are accurate and true to them?
  - **Example:** A client identifies as “non-binary” but the only options in the gender field are “Male”, “Female”, and “Unknown”. There is not a value that accurately captures their gender identity.
- Is there a shared understanding of the definitions behind fields and values by those responsible for data entry?
  - **Example:** A staff member selects “Brief Service” for their case’s Closing Level of Service after providing advice to a client for two hours via phone. A different staff member selects “Advice and Counsel” for their case’s Closing Level of Service after providing the exact same service. The same work is captured differently in the Closing Level of Service field, because there is not a common understanding of the difference between “Advice and Counsel” and “Brief Service”.
- Addressing the above questions is a great starting place towards improving the quality of data that gets collected on an ongoing basis. To help with this, consider implementing routine data cleaning practices to help maintain quality data. These practices could include quarterly reports to review data completion rates for new intakes or cases with no recent updates that may need to be closed or automated reports pulling cases from the CMS with incomplete data for staff to address. Also schedule regular administrative reviews of data collection processes throughout the year, such as intake and closing forms, to make sure these are set up to facilitate complete and accurate data collection of key data points.

### **How can organizations ensure they have adequate staffing for data analysis work?**

When thinking about staffing for data work, keep in mind that this includes not just those conducting the analysis, but also staff who enter or review data into case management systems and other data repositories. Building a strong data culture can involve staff time and effort from all across the organization, and it is important to acknowledge this work as part of data analysis efforts.

In terms of hiring for data analysis work, it is not always feasible for legal aid providers to hire a full-time staff member or team dedicated solely to data analysis. For organizations that find themselves in this situation, here are a few alternative approaches to staffing for data analysis:

- Combine data analyst responsibilities with another related position, such as a grants coordinator, case management system administrator, or technology manager.
- Bring on part-time support or interns for this work.
- Talk with community partners to see if there is a shared need for data analysis staff and consider splitting a position across multiple organizations.
- Connect with volunteers in the community who may be willing to engage in data analysis work for low or no cost. Ask board members to identify individuals in their network who may be willing to volunteer their data analysis services.
- Connect with local academic institutions or hire external consultants for discrete analysis projects.



## Overview of Evaluations



**How to Use This Section:** Learn about common use cases and examples of evaluation in legal aid.

### How do evaluations fit in to data analysis work?

In the legal aid sector, the term “evaluation” is often used as a stand-in for “assessment”, covering a broad range of use cases and activities such as:

- Monitoring case types and volume in a period
- Collecting feedback from end users on a new tool
- Conducting staff performance reviews
- Reporting on case outcomes and financial benefits secured
- Tracking demographic trends of client populations
- Identifying the service needs of a given community

Each of the above activities can be informed and bolstered by collecting and analyzing relevant data, including through methods described in earlier sections of this toolkit. In the research community, the term “evaluation” has a narrower meaning used to describe a specific set of activities designed to determine the effectiveness and efficiency of a program, as described in the Centers for Disease Control and Prevention’s [framework for program evaluation](#). The phrase “program evaluations” will be used throughout this section to denote this particular meaning.

Program evaluations are used most often to demonstrate a successful implementation, confirm the positive impact of a program, or learn from negative or unexpected results. These evaluations aim to generate information that can be used to ultimately improve a program. Program evaluators make use of a specific set of social science research and analysis methods, including advanced statistical analysis, to produce rigorous results. Program evaluations make use of data analysis, following the steps outlined in this toolkit of data collection, preparation, analysis, presentation, and learning.

## What is program evaluation?

Some program evaluations take place before or during program implementation, to inform its design and delivery (i.e., formative evaluations). Others take place towards or at the end of the program to assess outcomes and impact (i.e., summative evaluations). Two of the main types of evaluation are:

**Process Evaluations:** This formative evaluation type is used to determine whether the program is being implemented according to plan. It involves a detailed examination of how the program is being carried out, whether the intended program activities are taking place as expected, and what obstacles are occurring that impact program operations. Process evaluations also assess if designated resources are being used, target communities reached, and outputs produced as expected. These evaluations help to identify issues in program service delivery, so that appropriate adjustments can be made while the program is ongoing.

**Impact Evaluations:** This summative evaluation type is used to determine whether the program successfully achieved its intended outcomes. It assesses the causal relationship between the program and any measured outcomes. This is addressed through two questions:

1. What were those measured outcomes for program participants?
2. Did the program cause these outcomes or were they a result of other factors?

To answer these questions, evaluators seek a counterfactual: what outcomes would have occurred if the program did not happen? Crafting an estimated counterfactual (or control) and determining the program's causal effects typically involves experimental or quasi-experimental research methods. It is important for this work to be done by those trained in these methods to avoid producing inaccurate, biased, or incomplete results.

These two types of evaluation go hand-in-hand. For example, it is important to determine whether a program was implemented as intended (i.e., through a process evaluation), so that impact evaluation results can be attributed to the actual program activities that occurred (i.e., through an impact evaluation). If the program was implemented differently than intended, then the program impact stems from however the actual implementation occurred and not as it was originally planned.

## Why would an organization do a program evaluation and how do they start it?

As with any other type of data analysis project, it is important to start with the why: why is a program evaluation needed? Taking this a step further, it is also necessary to consider how results will be used. Common reasons for program evaluations include:

- To improve a program by understanding what is and is not working well
- To secure additional funding or justify program expansion
- To comply with reporting requirements from funders
- To generate knowledge around a program's impact

For an organization interested in program evaluation, a good starting point would be to articulate a theory of change ("TOC") and create a logic model for a particular program. A TOC reflects why a program is expected to achieve its intended outcomes. A logic model captures the how of a project, documenting the inputs and activities that comprise the program, the anticipated outputs, and the intended short and long term outcomes. Both process and impact evaluations are based around a program's underlying TOC. More on TOCs and logic models can be found in LSNTAP's [Project Management toolkit](#).



Starting with a TOC and logic model will help an organization lay the groundwork for what should be evaluated and what measures are most appropriate, either for a process or impact evaluation. The TOC and logic model may also spur questions and ideas for what, if anything, needs to be evaluated. Beyond this, organizations will need to take other factors into consideration including:

- **Timing:** Both process and outcome evaluations could be planned for before a program starts, but in general occur at specific times in the program lifecycle. Process evaluations typically happen at the start or during program implementation, so results can be used to adjust the program before it concludes. Impact evaluations typically occur after the program has ended to assess the short and/or long-term program impact.
- **Cost:** Program evaluations can be time and resource intensive efforts to draw accurate conclusions. Cost will vary significantly depending on the size and scope of the evaluation effort.
- **Staffing:** Because of the complexity involved in this type of research, program evaluations are often administered by individuals trained in this subject. Organizations should consider reaching out to academics, researchers, or evaluation consultants to see if they might engage with your organization for program evaluations. One such resource is the Harvard Law School's [Access to Justice Lab](#). Even if an outside entity is leading the efforts, it is also important to consider what internal staffing resources will be needed to inform and support the evaluation.
- **Feasibility and Relevance:** Whether a program evaluation is feasible or relevant is important to consider at the onset, particularly for impact evaluations. For example, the program participant size might be too small to be able to draw accurate statistical conclusions, there may be ethical concerns with a potential study, or it might not be possible to generate results in time for them to be of use.

This toolkit section has introduced important concepts and considerations for program evaluation. To learn more, check out these evaluation resources from [Innovations for Poverty Action](#), the [International Organization for Migration](#), and the [US Office of the Administration for Children and Families](#).

### What are Technology Innovation Grant (TIG) evaluations?

As part of their central role in funding dozens of legal aid organizations across the country, the Legal Services Corporation (LSC) operates the Technology Initiative Grant (TIG) program to support a wide range of initiatives that leverage technology to improve, expand, and facilitate delivery of legal services by existing LSC grantees. In addition to the traditional larger Technology Innovation Grants, the TIG program also offers smaller grants (up to \$35,000) for Technology Improvement Projects (TIPs) to enable planning for technology needs. Like other grant opportunities, TIGs and TIPs come with reporting requirements, some of which guide the recipient in evaluating the project. These reports also enable LSC to understand how the funds were spent and learn about the methods used and the resulting outcomes.

For the purposes of the TIG program, LSC categorizes data by whether it is obtained by administrative means, collected in surveys, or gathered through direct engagement with participants in the target system. In addition, LSC distinguish between qualitative and quantitative data. This distinction can inform selection of sources according to the needs of the intended evaluation:

- Administrative data is the easiest to obtain, particularly in high volumes, but most limited in scope and depth;
- Direct participant engagement is more costly but is more flexible and can offer a much richer

understanding.

- Organizations should use a mix of data sources to balance complexity and cost while ensuring they will adequately assess their intended evaluation targets.

For organizations preparing for a TIG or TIP evaluations, consider these LSC resources that provide support and ideas for:

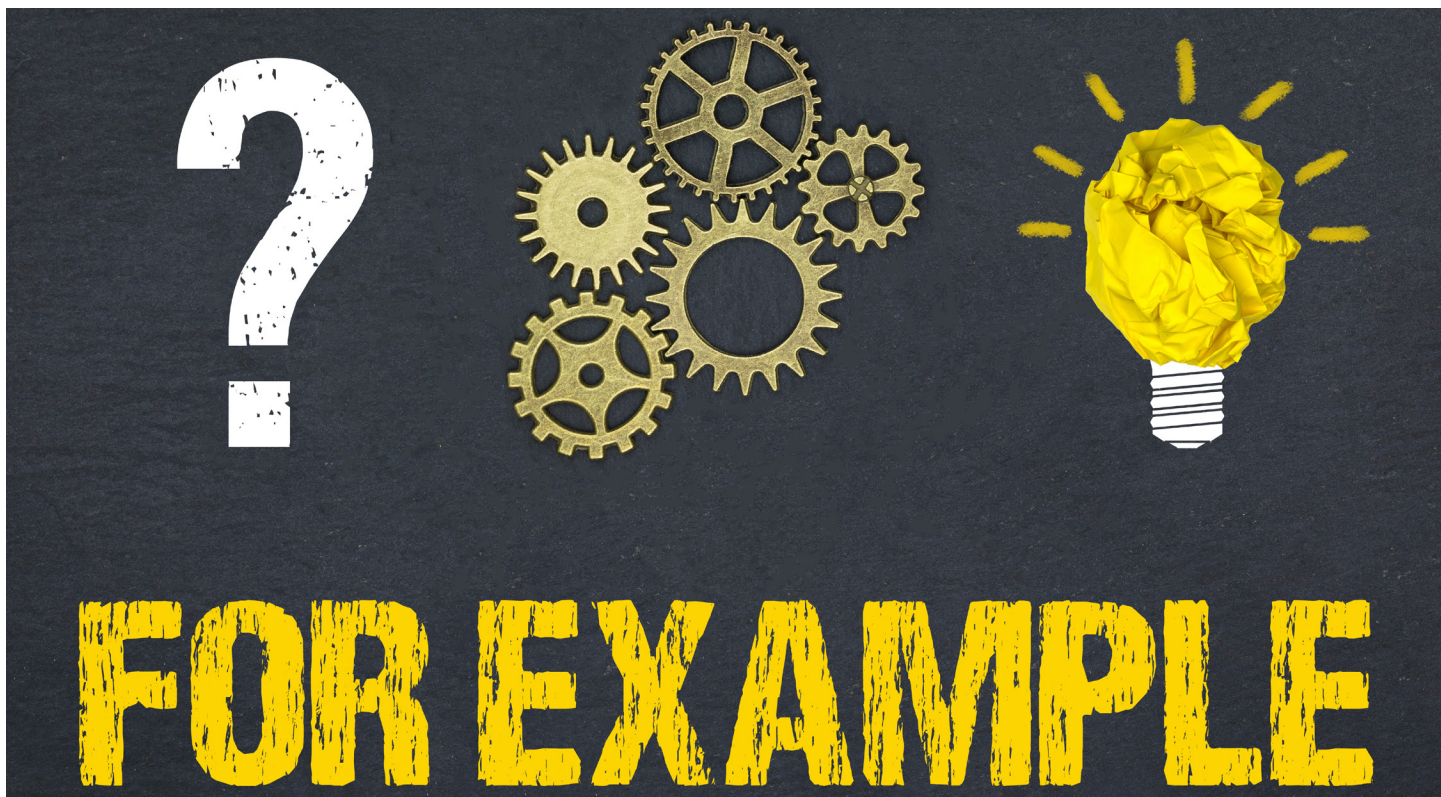
- Specifying project [goals](#) and [objectives](#)
- Identifying project [strategies and activities](#) for each objective
- Selecting the [methods and datasets](#) for the evaluation

TIG and TIP Final Reports follow a standardized format to encourage reflective analysis by the grantee. These reports include components from the Evaluation Plan which is required as part of the application process. Successful reports will often include appendices with additional data, analysis, and/or sub-reports. Many organizations contract with outside consultants to evaluate a TIG or TIP project and include those findings in the final evaluation report.

Given its role in funding a broad range of legal aid organizations, LSC requests multiple components of the analysis that will inform future projects, including impacts on a broader population of those served, if applicable, and suggestions for standards usable by those undertaking similar efforts.

Grantees describe challenges that arose during the project, strategies for addressing them, and overall lessons and recommendations for replication or expansion of the effort undertaken. The diversity of projects undertaken through TIGs makes the specialized knowledge that is developed in this manner all the more important.

## Data Analysis Project Examples



**How to Use This Section:** Explore examples of data analysis projects on online intake systems, language access, and staff workloads.

## Project Example 1 - Website Example

Legal Aid of ABC (LAA) is looking to complete the following three projects that will be informed by data analysis. These example projects follow the lifecycle of a data analysis project as outlined in sections 2-6 of this toolkit. These snapshots highlight types of questions that may be answered through data analysis and examples of data collection and analysis approaches for these questions.

### #1: Website Example

- **Background:** Legal Aid of ABC hosts an online intake application that is publicly accessible via their website. The online intake has been live for a year without any refinements. LAA staff have heard feedback from multiple clients that the application was confusing and time-consuming to navigate. The organization is now looking into this complaint to assess if and how the online intake should be altered to improve usability.
- **Key Question:** What should be done to improve the usability of the online intake application?
- **Intended Use of Analysis:** LAA will use results to inform what changes should be made to their online intake and how to prioritize those changes.
- **Audience:** The Director of Technology will use results to make a list of recommended changes to the online intake. The ED will need to sign off on all changes to the online intake.

### Data Collection Plan:

- LAA will collect data for this effort through three methods:
  1. Adding a user survey at the end of the online intake application to collect feedback on usability
  2. Conducting accessibility testing using an automated software tool
  3. Adding fields to track new metrics on the online application around completion rates and duration
- To incentivize and thank user survey respondents for their participation, all respondents will be entered in a drawing for a \$25 gift card.

### Data Analysis Examples:

- Percentage breakdown of applicants who do not complete the online intake by drop-off page
- Average and range of duration it took applicants to complete each page of the online intake (in minutes)
- Frequency table of common phrases in responses to an open-ended survey question concerning pain points with the online application process
- Average scores for various categories of accessibility metrics; scores produced by the automated software tool on a scale of 1-inaccessible to 10-meets accessibility standards
- Comments made during the usability testing of the existing online intake application form

### Example Metrics Included in Presentation:

- Bar chart of total applications by average time spent completing each step of the online intake process
- Table with the accessibility metrics sorted from worst to best score
- Pie chart of total applicants who did not finish the online intake, broken down by the last page completed
- Summary of comments obtained from usability testing

### Decisions Informed by the Analysis:

- LAA was able to identify specific steps to improve the accessibility and usability of the online application and placed priority on addressing the accessibility items that received the 3 lowest scores.



## Project Example 2 - Language Access Example

### #2 Language Access Example

- **Background:** Legal Aid of ABC has acquired funding to hire three new paralegals to assist with intakes. They are looking to hire individuals who speak another language in addition to English but are unsure what languages to prioritize in the hiring process. LAA staff are interested in learning about the language access needs of their current client population, as well as of the general, low-income community which they serve.
- **Key Question:** What languages should LAA prioritize when hiring bilingual paralegals?
- **Intended Use of Data Analysis:** LAA will use results to inform the job description and hiring criteria, to prioritize candidates with needed language skills.
- **Audience:** The Intake Manager, who will supervise these new positions, will work with the Director of Operations to decide together on which language skills to prioritize.

### Data Collection Plan:

- LAA will utilize three existing datasets for this effort:
  - 1. Data from LAA's case management system for new clients from the past year to date, including their primary language and need for an interpreter
  - 2. Aggregated public data made available by LAA's local government on the demographics of the population in their service area, including poverty status and primary language spoken
  - 3. Data from LAA's HR platform on the languages spoken by current staff

### Data Analysis Examples:

- Counts of clients by primary language spoken and need for an interpreter
- Percentage breakdown of clients whose primary language was not one that any current staff are fluent in
- Counts of clients whose primary language is one for which there is no fluent staff member
- Percentage breakdown of primary languages spoken by the local, low-income population

### Example Metrics Included in Presentation:

- Pie chart of clients who need an interpreter, broken down by primary language spoken
- Lists of the five most common non-English languages which are clients' primary language and for which there are no fluent staff
- Table displaying number of staff who are fluent in each non-English primary language, by level of language expertise
- Zip code map of LAA's service area displaying the most common, non-English language spoken by low-income population in each zip code

### Decisions Informed by the Analysis:

- LAA was able to identify that there was a gap in staff fluency of two primary languages commonly spoken by clients, Arabic and Korean. They included Arabic and Korean language skills as part of their job postings for the new paralegal positions.

## Project Example 3 - Performance Review Example

### #3 Performance Review Example

- **Background:** Legal Aid of ABC has three internal teams: housing, family law, and immigration. The team managers have requested data to compare staff performance between last year and this year. They plan to use this information to inform end-of-year performance reviews for those they supervise.
- **Key Question:** How did staff performance change between the last and current year?
- **Intended Use of Data Analysis:** LAA will use results to assess performance and inform decisions around salary increases, promotions, and bonuses for staff.
- **Audience:** The supervisors for each of the three teams (housing, family law, and immigration) will review results and use them to inform their end-of-year performance reviews with supervisees. The ED will also review the results to ensure consistency across the organization in terms of compensation and promotion.

#### Data Collection Plan:

- LAA will utilize one existing dataset for this effort:
  - 1. Data from LAA's case management system from the last and current year for cases open at any point during this time period.

#### Data Analysis Examples:

- Counts of intakes and cases closed per month
- Average cases open as of the start of each month
- Counts of closed cases by level of service provided
- Average duration for closed cases from open to close, sorted by level of service provided (in days)

#### Example Metrics Included in Presentation:

- Bar chart of total closed cases by level of service with different color bars for each of the two years
- Line chart of total cases closed each month, with one line for each of the two years
- Pie charts of all cases handed each year, broken down by legal issue

#### Decisions Informed by the Analysis:

- LAA supervisors were able to identify workload trends and spark deeper conversations with staff around their performances. LAA supervisors determined there was an ongoing need for this information, so staff built out a dashboard with related workload metrics that automatically refreshes with up-to-date case data.



# Acknowledgments



**LSNTAP produced this toolkit with support from Just-Tech, LLC.**

LSNTAP and Just-Tech would like to thank the following individuals for their feedback and assistance with this endeavor: Holly Stevens (Legal Services Corporation), Najae White (New York Legal Assistance Group), Susan Bales (Michigan Advocacy Program), Whit Froehlich (LSNTAP), and Zach Ballard (New York Legal Assistance Group).



**Legal Services National  
Technology Assistance Project**