**Data Collection – six steps to success**

If an organization is considering whether to collect data on its own or get help from an external consultant, it will need to have enough information to make an informed decision about how to proceed.

This section outlines some of the key considerations that may arise during various steps in the data collection process. There is no requirement that these steps be followed or pursued in the order that they are written. The model presented is offered as a reference tool. How data is gathered and analyzed depends on many factors, including the context, the issue that needs to be monitored, the purpose of the data collection, and the nature and size of the organization.

The main consideration is to make sure that any information collected is done in a way and for a purpose that is consistent with the *Code*and complies with freedom of information and privacy protection legislation. In the interest of effectiveness and efficiency, it is recommended that efforts be made to collect data that will shed light on issues or opportunities. To protect the credibility and reliability of data, information should be gathered using accepted data collection techniques.

**Step 1: Identify issues and/or opportunities for collecting data**

The first step is to identify issues and/or opportunities for collecting data and to decide what next steps to take. To do this, it may be helpful to conduct an internal and external assessment to understand what is happening inside and outside of your organization.

Some organizations, like FCP and Legislated Employment Equity Plan (LEEP) employers, are given specific direction on what issues should be explored and how data must be collected. Other organizations may have more flexibility to decide when and how to collect information to achieve certain goals. Some of the non-exhaustive questions identified below may apply to a diverse range of organizations and audiences, including employees and service users. Depending on the organization, these questions may be considered at Step 1, or at different stages in a data collection process.

Conduct a review of all policies, practices and procedures applicable to employees, service users or another appropriate audience:

* Does the organization have human resources and human rights policies, practices and procedures that are accessible to all employees or to the people they serve?
* Does the organization have clear, transparent and fair complaint procedures in place to deal with allegations of discrimination, harassment or systemic barriers?
* Have any claims, grievances or allegations been made or received relating to discrimination, harassment or systemic barriers?
  + Do any signal barriers to persons protected under the *Code*and/or other individuals/groups in society based on a non-*Code*ground?
  + Have any been dealt with appropriately and in accordance with existing polices, practices and procedures?

Explore organizational culture from a human rights, diversity and equity-inclusion lens:

* What are the organization's mandate, goals and core values?
* What is the history of the organization?
* Are equity, diversity and inclusiveness supported, reflected and promoted by senior leaders throughout the organization?
* Are performance measures in place to motivate the achievement of an organization’s strategic human resources, human rights, equity and diversity goals?
* Do employees feel that the organization is diverse, inclusive, and provides equal opportunity for learning and advancement?
* How are decisions made?
* How are employment, programming or service delivery opportunities advertised?
* Does the organization have formal, transparent and fair processes in place to recruit, hire, promote, terminate and retire staff?
* Does the organization have a clear system of discipline?
  + Is this system perceived to be applied fairly and consistently?
  + Do service users feel that they are welcome, valued, and able to use the services offered by the organization?

Assess external context:

* Are there best practices in the industry/sector or among similar organizations that can be learned from?
* Are there objective data or research studies showing that discrimination or systemic barriers exist or do not exist in the organization, industry/sector or similar organizations?
* Is there evidence from other organizations or jurisdictions that a policy, program or practice, similar to one in place at the organization, has had a positive or negative impact on Code-protected persons or other marginalized persons in society?
* How is the organization perceived by the community it operates in?
* Have the media or advocacy groups complimented or criticized the organization about human rights, human resource or equity issues?
* What are the demographics of the people the organization serves or the community it operates in?
  + Are the demographics changing or projected to change in the future?
  + Is the organization proactively looking at ways to make sure that it has the skills and knowledge to meet the potential needs and concerns of this changing demographic?

Check representation:

* Compare the organization’s workforce makeup to the availability of labour or the demographics of the service users in the community, city, region, province and/or country it operates in.
* Is the organization representative of and responsive to the needs of the community it serves?
* At this stage, a detailed comparison is not needed. The goal here is to identify key issues and/or opportunities that may need further study by noting obvious gaps, disparities or trends.
* Organizations can:
  + Estimate how people or groups identified by *Code*grounds and other persons/groups are represented and distributed among their employees or service users by levels of responsibility, occupation, branch, department or other appropriate measure.
  + Are there any areas in the organization or in service delivery where the persons or groups seem to be obviously over-represented or under-represented?

Finding the above information can be challenging for smaller organizations, but the internet offers a wealth of resources to choose from. Media reports may offer insights, as well as on-line resources offered by the OHRC, Statistics Canada, the City of Toronto, government agencies, and community organizations that focus on *Code*and non-*Code*ground-related topics. Information may also be gathered from various sources using accepted data collection research methodologies discussed in Step 3.

It is to be expected that an internal and external assessment of the organization, in light of the questions listed above, may result in a number of potential issues and/or opportunities for exploring data collection. Before proceeding to Step 2, organizations may wish to consider whether there are any preliminary actions that can be taken to address these issues and/or opportunities, without collecting data (*e.g.*training, policy development).

**Example:** The review in Step 1 may have identified the following issues and/or opportunities for collecting data:

* Positive public feedback received about a pilot community policing project in high-crime neighbourhoods.
* Unclear and inconsistent human rights policies and procedures in place to address sexual harassment.

The above examples present a potential opportunity or problematic human rights issue, respectively, and could lend themselves to data collection. Decisions need to be made about how best to address the identified opportunities and/or issues and whether it would be appropriate to act, based on the assessments in Step 1 (either instead of or together with further data collection).

If the results of the internal and external assessment seem to show that the organization does not have any pressing problems with discrimination and/or systemic barriers, and is generally in compliance with the *Code*and OHRC policies, consider whether the organization could still benefit from proactively implementing a data collection initiative (for example, to help monitor the ongoing effectiveness and suitability of policies, programs and intervention strategies).

**Step 2: Select issue(s) and/or opportunity (ies) and set goals**

The focus of Step 2 is choosing a priority issue(s) and/or opportunity(ies) for collecting data, and then setting goals and objectives.

The organization reviews the issues and/or opportunities identified from the internal and external assessment done in Step 1, and picks one or more specific issues and/or opportunities for starting a data collection project from among the list of priorities. Some of the questions an organization can consider when deciding to prioritize an issue and/or opportunity for gathering data include:

* Is there a fundamental reason or opportunity to collect data from which other issues and/or opportunities seem to arise?

**Example:** An aging taxpayer base provides a government body with a pressing reason to collect data on this group’s projected size, needs and revenue base. This changing demographic also presents an opportunity for the government body to ensure that it is proactively developing policies, programs and services that are accessible and appropriate to meet the needs and concerns of these taxpayers.

* Did the internal and external assessment of the organization in Step 1 reveal any critical gaps or trends that are apparent in the organization, industry/sector or similar organizations?
* Is there one particular area that has drawn positive/negative media attention or been subject to multiple complaints, internal rumours and concerns?
* Does there seem to be a greater diversity or lack of diversity in one area compared to others?

Goal-setting

While the organization may intend to collect data relating to multiple issues and/or opportunities at the same time, the next steps, including goal-setting, should be individualized for each issue and/or opportunity.

The specific goal(s) defined for each issue and/or opportunity may depend on a hypothesis or guess about what is happening that can be tested using data collection techniques and analysis.

**Example:** A downtown Toronto hotel receives complaints from guests, who self-identify as being gay, about the unwelcome treatment they received from staff. A hypothesis might be that hotel staff lack sufficient awareness and training about how to deal respectfully with guests who are gay, or are perceived to be from the larger LGBT community. The goal is to get enough evidence to test this hypothesis.

Step 2 can also involve an organization brainstorming a smaller set of questions that may be answered by collecting data. Rather than asking a general question like, “Is there any evidence of discrimination on the basis of sexual orientation or gender identity in this hotel?” one might ask, “What percentage of hotel guests self-identify as being part of the LGBT community?” and “What are the perceptions of the service received by self-identified LGBT patrons?” Ultimately, data that is collected should be rationally connected to the goals set and the overall purpose for collecting the data.

**Step 3: Plan an approach and methods**

In Step 3, organizations will make decisions about who will be surveyed, how data will be collected, the sources of data that will be used, and the duration of the data collection project, among other questions. These decisions may be made in consultation with an expert. The methods and approaches will flow from the goals set in Step 2, and will vary significantly depending on a number of factors, including the organization’s context, size, resources, and the purpose and complexity of the issue(s) or opportunity(ies) selected.

Some of the questions to consider at this stage include:

Who will the data be collected about?

The “group of interest” (e.g. youth service users of a local community centre who cannot read and speak English as a second language) will be the focus of the study, and the data collection methods used will refer to this group, or the persons within it, depending on the goals of the project.

Understanding discrimination

* When thinking about who the data will be collected about, it is important to consider who you think will be most affected by, for example, the discrimination or inequities that you wish to measure. Is it a broad category (e.g. all service users who *cannot* read), or a sub-set of that category (e.g. *youth* service users who *cannot*read)? The italicized words refer to a unique characteristic about a broader group that an organization may wish to gather information about.
* Depending on factors like the goals of the data collection project, the organization’s size, resources and time, data may be gathered about many sub-sets within a broader group of interest (e.g. *youth*service users who *cannot read* and who *speak* *English as a second language*).
* Collecting data about a group of interest that shares characteristics, based on several *Code*or non-*Code*grounds, can help an organization understand the behaviour, perceptions, values and demographic makeup of services users and other subjects of interest. Generally speaking, gathering data that reflects more than one *Code*and/or non-*Code*ground can allow for richer, nuanced information and more complex analysis.
* It is important to recognize that based on their unique combination of identities, people may be exposed to particular forms of discrimination. Multiple forms of discrimination can intersect and compound to form a unique experience of discrimination. This perspective is referred to as an “intersectional” analysis of discrimination.

**Example:** A South Asian male youth service user, who cannot read and speaks limited English, may face discrimination on any of the grounds of age, race, colour, ancestry, ethnic origin, place of origin, gender, disability or perceived disability (*e.g.* could be seen as having a learning disability). However, he may also be exposed to discrimination on intersecting grounds based on being identified as a “young, illiterate Indian male from a foreign country,” based on the various assumptions or stereotypes that are uniquely associated with this socially significant interaction of multiple identity factors.

To better understand the potential impact of multiple identity factors, or intersectionality, when collecting and analyzing data about a group of interest, it may be helpful to consult with communities, and review applicable research and other relevant documents that highlight how the dynamic of discrimination and disadvantage can play out in a practical way for persons identified by *Code*and non-*Code*grounds. The OHRC’s recent edition of **Human Rights at Work** is a useful reference for this purpose. The OHRC has also developed policies and guidelines that provide a more detailed outline of how the *Code*applies to the various grounds (see Appendix G for a list of OHRC guides, policies and guidelines).

Who will the group of interest be compared to?

The “comparator group” should be persons who share one or more characteristics with the persons in the group of interest, but differ in the key characteristic(s) being studied (e.g. youth service users who cannot read but *can* speak English fluently). The experiences of youth service users who cannot read and who speak English as a second language can then be compared to youth service users who cannot read but can speak English fluently.

What locations or geographical areas will the data be gathered from?

Some data collection initiatives require gathering data from multiple sizes, groups or communities located in different locations and geographical areas. When determining where to collect information from, key factors to consider include who the data will be collected about and who the data will be compared to.

**Example:** A local community center is interested in making its current youth literacy program more responsive to the needs of an increased number of youth in the surrounding area who cannot read and who speak English as a second language. The community center plans to gather information about the community it serves and the geographical region it is located in. Data is gathered from the community center’s pre-existing records relating to its service users, including people who attend the youth literacy program or have expressed an interest in it. Publicly available information about the characteristics of the surrounding neighbourhood is also explored, among other data sources.

What categories will be used to identify the group of interest and comparator group?

Choosing categories provides a way to organize the information that is collected. This can be done either before collecting data, as discussed in this step, or after data is collected (see Step 5).

In some cases, although it is not required, it is preferable to use pre-determined categories such as those developed by Statistics Canada. There are certain benefits to this approach.

**Example:** Organizations can be confident that the 12 racial groups used by Statistics Canada will represent how the majority of Canadians racially classify themselves. In addition, use of these categories is most likely to produce reliable and valid results and enable researchers to directly compare the results of their studies to Census data collected by Statistics Canada.

The limitations are that if these categories are used, some respondents may not identify with them or may object to them. Another limitation is that Statistics Canada does not produce Census data on all grounds (for example, on sexual orientation).

For a fee, Statistics Canada will customize its data. For example, it can break it down to “disaggregated” data for a local labour market or for a particular occupational category.

Another limitation is that the Statistics Canada categories may be too broad depending on the goals selected in Step 2.

**Example:** Using a broad category such as “racialized” can mask important differences between racialized groups, since racialized groups are not subject to exactly the same experiences, racial stereotypes and types of discrimination. However, when it is necessary to describe people collectively, the term “racialized person” or “racialized group” is preferred over terms like “racial minority,” “visible minority,” “person of colour” or “non-White” as it expresses race as a social construct rather than as a description based on perceived biological traits. Also, these other terms treat “White” as the norm that racialized persons are to be compared to, and have a tendency to group all racialized persons in one category, as if they are all the same.

Consider other categories to describe the groups selected (for example, relating to job or service categories). Organizations may ultimately choose the categories that best reflect where the organization is at in terms of achieving its human rights, equity and diversity goals.

How should data be collected?

In the context of human rights, social-science researchers are commonly asked to lead or help with data collection projects. Two types of data are used in social science research: qualitative and quantitative. A good research effort involves the use of both types. Both approaches, while distinct, can overlap and rely on the other to produce meaningful data, analysis and results.

Qualitative data:

* Typically, data is called “qualitative” if it is in the form of words, but may also include any information that is not numerical in form, such as photographs, videos and sound recordings.
* Qualitative methods are aimed at describing a specific context, event, people or relationship in a broad contextual way, by trying to understand the underlying reasons for behaviour, thoughts and feelings.
* Common qualitative research methods include observation, one-on-one interviews, focus groups and intensive case studies.

**Example:** A restaurant chain wants to improve service and access to customers with disabilities. Management decides to collect qualitative information using focus groups consisting of a range of stakeholders, including customers and representatives of organizations from the disability community.

**Potential strengths**

* qualitative data excels at "telling the story" from the participant's viewpoint (it helps participants feel like they have been heard)
* can help others better understand the issue or problem by providing the rich descriptive detail that explains the human context of numerical results

**Potential weaknesses**

* perceived that the accuracy of qualitative data can be influenced by false, subjective or manipulated testimonies. Good qualitative data, checked by a professional researcher and gathered using accepted data collection research methods, can address the impact of such factors
* depending on the nature and size of the project, as well as the sophistication of the methods and analysis used, can take a significant amount of time, be very labour-intensive, and yield results that may not be general enough for policy-making and decision-making purposes.

Quantitative data:

* Typically, data is called “quantitative” if it is in the form of numbers.
* A quantitative approach can be used to count events or the number of people who represent a particular background.
* Common quantitative tools include surveys, questionnaires and statistical data (such as Statistics Canada census information).
* It is important to note that all quantitative data is based on qualitative judgment. In other words, numbers cannot be interpreted by themselves, without understanding the assumptions that underlie them.

**Example:** A simple 1- 5 rating variable for the survey statement, “My union handles human rights grievances in a sensitive and efficient manner” gives respondents the option of circling: 1 (Strongly Disagree), 2 (Disagree), 3 (Neutral) 4 (Agree) and 5 (Strongly Agree).  
  
A respondent circles “2 = Disagree.” To understand the value of “2” here, a researcher must consider some of the judgments and assumptions that are behind this choice. Did the respondent understand the term "human rights grievance"? Has the respondent had experience filing a grievance with the union? Does the respondent like unions generally?

**Potential strengths**

* perceived to be more credible and reliable than qualitative data because of the use of numbers, which are seen as an objective source of data. This is not necessarily the case. The accuracy of quantitative data can be influenced by manipulation and bias of the researcher, among other factors, unless checked by the researcher’s professionalism and the use of accepted data collection research methods
* quantitative data excels at summarizing, organizing and comparing large amounts of information, and drawing general conclusions about a research topic of interest
* can help measure progress and success
* good at identifying trends and determining the magnitude of a research topic of interest.

**Potential weaknesses**

A focus on numbers and rankings alone can overly simplify or lead to an inaccurate understanding of complex situations and realities, unless a broader context is provided

**Example:** An employment data survey of the Custodial Services Division of a large organization reveals that 80% of the cleaning staff are women and that 6 of 7 Custodial Services supervisors are men. A comparison between these figures and gap data from Human Resources and Skills Development Canada (HRSDC) shows that, while there is an overrepresentation of women in the ranks of cleaners, there is no gap for women in the ranks of supervisors.

The reason for the seeming discrepancy is that HRSDC gap data is based on availability. Nationally, so few women are Custodial Services supervisors that there is a statistically insignificant availability, giving rise to the conclusion that there is no numerical gap with respect to women supervisors. This conclusion, however, does not make sense since the organization knows that the 200:40 women to men cleaning staff ratio is supervised by a 6:1 male to female supervisory staff ratio. The organization decides to ignore the HRSDC data and apply common sense by setting up career advancement mentoring and other policies and programs to increase the number of female supervisors in its workforce.

* subject to multiple interpretations of what the numbers actually mean, which can lead to a distorted understanding of a research topic of interest. This potential weakness can be minimized by using accepted quantitative research methods and identifying appropriate warnings to explain the parameters and assumptions underlying the research
* depending on the nature and size of the project, as well as the sophistication of the methods and analysis used, it can be costly to gather the required information
* in areas of research that are relatively new or where tools, indicators, procedures and sources are far from settled, statistical data can be lacking or of unequal quality, causing problems for comparisons. These difficulties are often compounded by other issues, like definition problems (e.g. the meaning of the word “freedom” – depending on the interpretation of the word chosen, it may produce different issues and results).

What sources of data should be used to collect information?

Qualitative and quantitative data are generally gathered from more than one source. Where possible, two or more of the following sources should be used together to strengthen reliability and consistency in results.

Pre-existing or official data

Pre-existing or official data is information that has already been documented (e.g. newspaper clippings, case law, Statistics Canada census data, photographs) or is created by an organization during its routine business operations (e.g. employee personnel files, student registration forms, annual reports, occurrence reports). This data may contain information that directly relates to specific *Code*grounds like race, but more commonly will relate only indirectly (for example, in the form of names, place of origin or ethnicity). This type of information could be used as proxies or stand-ins for race, but would be less reliable than actually having self-reported racial data.

**Potential strengths**

* is efficient. Avoids the time, energy, expense and disruption involved in collecting data as a separate step from running daily operations

**Example:** Outcomes of workplace recruitment, hiring, promotions and terminations can be recorded, as can events such as interventions by security guards and customer complaints. When recording these events, relevant *Code*ground and non-*Code*classifications could also be included. This data could then be examined for trends over time to show whether discrimination or systemic barriers exist, may exist or do not exist.

**Potential weaknesses**

* to be a useful source of information, organizations need to be willing to collect the data as part of their ordinary record-keeping procedures
* the reliability of this data will depend on the diligence and accuracy of the reporting done by the people collecting it.

Survey data

Survey research is a broad area and generally includes any measurement procedures that involve asking respondents questions. A "survey" can range from a short paper-and-pencil questionnaire to an in-depth one-on-one interview (interviews will be discussed further below).

In designing a survey, it is important to consider the specific characteristics of the respondents, to make sure that the questions are relevant, clear, accessible and easy to understand. Some practical considerations to keep in mind are whether the respondents can read, have language or cultural barriers, have disabilities, and can be easily reached.

**Potential strengths**

* very useful for documenting an individual’s perceptions and perceived experiences of an organization’s work culture, service delivery or other areas of interest

**Example:** The TDSB’s 2006 Student Census, Grades 7-12 System Overview included a component on how senior and secondary school students generally perceived their schooling and out-of-school experiences in 10 areas, including school safety and home support and involvement.

* can contain questions that are quantitative or qualitative in nature, or a combination of both
* can be conducted on a small or large scale.

**Potential weaknesses**

* quality and reliability of survey data depends on factors like the expertise of the people conducting them, the design and appropriateness of the questions asked, and the credibility of the methods used to analyze and interpret the results
* may not provide an accurate measure of how others perceive a person’s background or experience.

**Example:**A transgender employee may self-identify as female but a third party may identify her as male.

Focus groups and interviews

Interviews and focus groups (also referred to as “group interviews”) allow for information to be provided orally, either individually or in a group setting. The data can be recorded in a wide variety of ways including written notes, audio recording and video recording.

Focus groups:

In focus groups, the interviewer facilitates the session. A select group of people are brought together, asked questions, encouraged to listen to each other's comments, and have their answers recorded. The same set of questions may be used for a number of different groups, each of which is constituted slightly differently, and for a range of purposes.

Focus groups may be facilitated by professionals, but this is not always needed. The decision to hire a professional facilitator may depend on the goals of the focus group research, the nature of the questions asked, the skills and experience of staff taking part, and the need for confidentiality or anonymity.

**Example:** To get the unique perspective of each group, an organization may wish to hold separate focus groups for representatives of each of the organization’s internal and external stakeholder groups, such as senior management, front-line employees, service users, union representatives and community groups. Or, it may be of greater value to organize a group that includes people representing all key internal and external stakeholders, to allow for contrasting ideas to be expressed and discussed.

Whatever format is chosen, it is important that the focus group is structured and managed in a way that cultivates a “safe space” for people to share their experiences. In some cases, this may not be possible without setting up separate focus groups or hiring a professional facilitator who is not connected to the organization.

**Potential strengths**

* focus groups allow for multiple narratives to be voiced in one “interview” about a research topic of interest
* act as tools for education because discussion among participants can illuminate the participants’ and the researcher’s views, helping to further refine research about a particular topic of interest.

**Potential weakness**

* does not allow participants to fully express their individual opinions and narratives, or ask questions when they immediately come to mind, because of the need to hear and accommodate other voices.

Interviews:

Typically, interviews involve a set of standard questions being asked of all respondents, on a one-on-one basis, so that accurate trends and gaps can be drawn from the data. Interviews are commonly conducted face-to-face, but for more rapid results, can also be done over the telephone, or, as technology advances, through video-conferencing and other means.

**Potential strengths**

* interviews can provide a rich, detailed perspective, impression or story on a research topic of interest
* the interviewer generally has the opportunity to probe more deeply or ask follow-up questions than when in a focus group setting
* data from both focus groups and interviews can provide valuable context for understanding and informing research, numbers, events, behaviour and other research goals
* depending on the size of the organization, the purpose of the data collection, the internal expertise available and other factors, focus groups and interviews can be done with relatively little expense.

**Potential weaknesses**

* one-on-one interviews allow for just one narrative or perspective on a research topic of interest
* can be very time consuming and resource intensive
* respondents in interviews and focus groups generally want to "look good" in the eyes of others.

Depending on the questions asked, they might “spin” their response to avoid being embarrassed, particularly in a face-to-face setting. Skilled interviewers may be able to address this potential weakness by doing a few things, like designing good questions, being perceptive, asking follow-up questions and cross-checking responses with other credible sources of information

* interviewers, in both individual and focus group settings, may distort an interview by not, for example, asking questions that make them uncomfortable or not listening carefully to respondents on topics that they have strong opinions on. The impact of this potential weakness can be addressed by taking steps like making sure that interviewers are properly trained and using standard interview questions.

Observed data

Trained staff or external experts can gather data by identifying and recording the characteristics and behaviour of research subjects through observation, either within or outside of an organization. Observed data can include information gathered using all of the senses available to the researcher, including sight, hearing, smell, taste and touch.

**Example:** A human rights organization that offers a mediation service hires a mediation expert to observe mediators and service users and provide feedback about any issues of concern related to human rights. To minimize potential stress and anxiety experienced by the people being observed, staff and service users are informed in advance of the purpose and goals of the exercise. Service users’ consent is sought. Staff is advised that the observed data gathered will only be used for research purposes and not shared with their managers. The expert maintains access to the data, and the results are reported on an aggregated and summarized basis to prevent individuals from being identified.  
  
Hiring experts, while potentially expensive, can add validity and credibility to research analysis because they are often perceived as having no vested interest in the research results.  
  
Information gathered using observation techniques differs from interviewing, because the observer does not actively ask the respondent questions. Observed data can include everything from field research, where someone lives in another context or culture for a period of time (participant observation), to photographs that show the interaction between service providers and service users (direct observation). The data can be recorded in many of the same ways as interviews (taking notes, audio, video) and through pictures, photos or drawings.

**Potential strengths**

* an effective and capable observer can provide an objective third viewpoint on what is going on, and draw out implications that are not obvious or that people are unaware of
* can be relatively inexpensive depending on factors like the size of the project, its goals, the organization’s resources and the duration of the project.

**Potential weaknesses**

* an observer, trained or otherwise, can influence the behaviour of the people being observed (for example, people could be motivated to behave better while under observation), which may ultimately affect the accuracy of observed results
* can cause potential stress and anxiety for the people being observed, more than the use of other data collection methods. Efforts can be made to minimize stress and anxiety by using effective communication strategies to inform participants, in advance, of the purpose, goals, confidentiality measures, duration of the project and other key information
* an observer, trained or otherwise, may not always be able to accurately differentiate within or between certain groups of people, particularly when an identity (ies) is/are non-evident (e.g. religion, mental illness, sexual orientation). A survey requesting self-identification information might be more effective in this regard.

Each source of data used to collect information has its strengths and weaknesses. Some of the more common potential strengths and weaknesses identified above have been highlighted. Analyzing data from multiple perspectives and relying on data from different sources can strengthen the conclusions drawn from research. A combination of statistical analysis, observational data, legal analysis, documentary analysis, in-depth interviews and external and/or internal consultation can help maximize understanding of a given situation. Organizations should choose the sources of data that best suit their program goals, context, resources and organizational culture.

How long will the data be collected (the scope of data collection)?

Data can be collected and analyzed on a short-term or project basis in response to situations or needs that arise from time to time. A short-term data collection project would include a start and a finish date, with set deliverables to be carried out over a certain period of time.

The best practice is to collect data on an ongoing, permanent basis, and to analyze this data as often as is needed to identify, address and monitor barriers to *Code*-protected persons or other persons based on non-*Code*grounds.

Data collected in a time-limited study may be less complete than data collected through ongoing monitoring. This is because short-term studies do not allow for the assessment of trends, patterns or changes over time. However, where costs, time and resources are a factor, short-term studies may be the preferred choice to fulfil a need and project goals.

Other factors may also influence the reliability of the data. For example, people may modify behaviour while under scrutiny during the data collection period.

**Step 4: Collect data**

When planning on how best to collect data in Step 4, it is important to be aware of the practical considerations and best practices for addressing logistical challenges organizations often face at this stage of the process. Implementing a data collection plan requires attention to matters such as:

* Getting buy-in from senior leadership and key stakeholders, in or outside of the organization. This group could include boards of directors, management committees, union representatives, employees, community groups, tenants, customers and service users.
* Establishing a steering committee or selecting a person(s) to be consulted and held accountable for all major decisions about the data collection process, such as design, logistics, communication management, coordination and finances.
* Determining who will collect the data (e.g., experts or trained employees).
* Identifying the logistics, resources, technology and people needed to develop and implement a data collection initiative.
* Anticipating and addressing key stakeholder concerns and questions about the project.
* Designing a communication and consultation strategy that will explain the data collection initiative and encourage the highest possible participation rate.
* Protecting privacy and personal information by using carefully controlled procedures for collecting, storing and accessing data that comply with privacy, human rights and other legislation. Dignity and confidentiality must be respected.
* Minimizing the impact and inconvenience for the people affected in the workplace or service environment, which includes choosing the best time to collect the data.
* Aiming for flexibility to allow for changes without great expense or inconvenience.
* Considering a test period or a pilot phase to allow you to improve and modify data collection methods, as may be needed.

**Step 5: Analyze and interpret data**

Step 5 involves analyzing and interpreting the data collected. Whether quantitative and/or qualitative methods of gathering data are used, the analysis can be complex, or less so, depending on the methods used and the amount of data collected.

Explaining the technical steps involved in analyzing and interpreting data is beyond the scope of this guide. An organization will have to determine whether it has the internal capacity and expertise to analyze and interpret data itself, or whether it will need the help of an external consultant.

A smaller organization that has basic data collection needs may be able to rely on internal expertise and existing resources to interpret the meaning of gathered data.

**Example:** An organization with 50 employees wants to find out if it has enough women working in management positions, and if there are barriers to equal opportunity and advancement. The organization counts the number of female employees it has (25), and determines how many of these employees are working in supervisory and management positions (two). A few motivated employees identify some issues of concern, like gender discrimination, that may have broader implications for the organization as a whole.

After deciding to do an internal and external assessment (Step 1), and gather qualitative data using focus groups and interviews with current and past employees, senior leadership decides that barriers exist for women in the organization’s recruitment, hiring, promotion and human resources policies, processes and practices. Efforts are made to work with female employees, human resources and other staff to address these barriers. The organization makes a commitment to foster a more equitable, inclusive work environment for all employees.

**Step 6: Act on results**

Once an organization has analyzed and interpreted the results of the data collected, it may decide to act on the data, collect more of the same type of data or modify its approach.

Quantitative and qualitative information can provide a solid basis for creating an effective action plan designed to achieve strategic organizational human resources, human rights, equity and diversity goals identified through the data collection process. If an organization feels it has enough information to develop an action plan, it should consider including the following elements:

* a summary of the results of the analysis and interpretation of the data
* identification of the barriers, gaps and opportunities that exist or may exist for Code-protected persons and other individuals/groups based on non-*Code*grounds
* steps that will be taken to address these barriers, gaps or opportunities now and in the future
* realistic, attainable goals with short-term and longer-term timelines
* input sought from stakeholders and affected communities
* how progress in meeting these goals will be monitored, evaluated and reported.

In some cases, an organization may decide that it needs to collect more information because there are gaps in the data collected, or areas where the data is unclear or inconclusive. This may prompt them to conduct a more detailed internal and external assessment (go back to Step 1) or try another approach.

In the end, there is no one or “right way” to conduct a data collection initiative. The experiences of Mount Sinai Hospital, KPMG Canada, the Keewatin-Patricia District School Board, TD Bank Financial Group, the University of Guelph and the *DiverseCity Counts* project and featured in the Appendices reflect this statement, yet also show some similarities in terms of the best practices and lessons learned.

**Six steps to success**

Step 1: Identify issues and/or opportunities for collecting data

Step 2: Select issue(s) and/or opportunity(ies) and set goals

Step 3: Plan an approach and methods

* Who will the data be collected about?
* Who will the group of interest be compared to?
* What locations or geographical areas will the data be gathered from?
* What categories will be used to identify the group of interest and comparator group?

How should data be collected?

* Qualitative Data
* Quantitative Data

What sources of data should be used to collect information?

* Pre-existing or official data
* Survey data
* Interviews and focus groups
* Observed data

How long will the data be collected (the scope of data collection)?

Step 4: Collect data

Step 5: Analyze and interpret data

Step 6: Act on result