

Data Infrastructure Starter Guide

Most organizations don't lack data – they lack structure. Even small improvements in organization can significantly reduce reporting time and improve data quality.

PURPOSE

- ▶ Clarify how to organize program data
- ▶ Identify gaps in your current system
- ▶ Take 1–3 immediate steps to improve data usability
- ▶ Data infrastructure tools overview

Use This If You...

- ▶ Are setting up a new program database
- ▶ Have messy or inconsistent data
- ▶ Struggle to pull reports
- ▶ Use disconnected spreadsheets



ABOUT BSRI

BSRI was created to support informed and effective decision-making for our partners with tools, resources, and data for action. As we wind down operations, we are creating these resources for the community to continue the work. We use our guiding principles in engaging with partners and communities, putting people first through accountability and reciprocity, communicating transparently, and sharing our knowledge widely.

ABOUT THIS DOCUMENT

This guide was developed by BSRI based on our experience supporting partners in managing complex data and grant reporting requirements. It is designed to provide a practical, adaptable framework that helps organizations strengthen their data structure, reduce reporting burden, and improve overall data usability. Rather than prescribing a single approach, this document offers foundational principles and actionable steps that can be applied across a range of systems and program contexts. The goal is to help organizations move from fragmented or inconsistent data practices toward more streamlined, reliable, and efficient processes.

By implementing even a few of the strategies outlined here, organizations can improve their ability to track services, generate reports, demonstrate impact, and use data to inform decision-making and continuous improvement.

USE OF ARTIFICIAL INTELLIGENCE (AI)

Our organization utilizes artificial intelligence (AI) tools to enhance efficiency, support data analysis, and strengthen evaluation practices. We also encourage our partners and clients to thoughtfully explore AI as a means to build future-ready evaluation capacity. All AI use is governed by strict data privacy and ethical standards. We do not input any personally identifiable information (PII), protected health information (PHI), or other sensitive data into AI systems in a manner that would violate HIPAA or other applicable privacy regulations.

Additionally, we exclusively use licensed AI tools and platforms that explicitly prohibit the use of submitted data for model training or retention beyond the scope of the intended task. This ensures that all information remains confidential, secure, and protected.

Quick Assessment + Starter Actions

Quick System Check

- ▶ Consistent naming
 - ▶ Use the same names for programs, services, and fields across all datasets
 - ▶ Avoid variations (e.g., “MH”, “Mental Health”, “mental health”)
 - ▶ Be consistent with naming case (e.g., camelCase, snake_case, PascalCase)



Why this matters

Inconsistent naming makes data impossible to combine or analyze

- ▶ One storage location
 - ▶ Store data in one shared, agreed-upon location (not across desktops, emails, and drives)
 - ▶ Everyone knows where the “source of truth” lives



Why this matters

Multiple versions = confusion + errors

- ▶ Standard entry
 - ▶ Staff enter data the same way every time
 - ▶ Use dropdowns or clear instructions when possible



Why this matters

Variation at entry = cleanup later

- ▶ 1 row = 1 record
 - ▶ Each row represents one unit:
 - ▶ One client
 - ▶ One service
 - ▶ One visit



Why this matters

Enables counting, filtering, and reporting

- ▶ 1 column = 1 variable
 - ▶ Each column captures only one type of information
 - ▶ Avoid combining multiple pieces of info in one field

**Why this matters**

Clean columns = usable data

- ▶ Standardized fields
 - ▶ Key fields follow a consistent format:
 - ▶ Dates (MM/DD/YYYY)
 - ▶ Program names
 - ▶ Service types

**Why this matters**

Standardization reduces errors and manual cleaning

Starter Actions

- ▶ Organization → centralize data
 - ▶ Choose one shared location (e.g., Google Drive, SharePoint)
 - ▶ Create a simple structure:
 - ▶ Raw Data
 - ▶ Clean Data
 - ▶ Reports
 - ▶ Use clear file names and store them in a folder you know they won't get lost in (i.e.; clearly named folders that you can easily search for)

**Start here**

If your data is scattered

- ▶ Structure → clean rows/columns
 - ▶ Ensure:
 - ▶ 1 row = 1 record
 - ▶ 1 column = 1 variable

- ▶ Remove:
 - ▶ Merged cells
 - ▶ Multiple values in one cell

**Start here**

If your data is messy or hard to use

- ▶ Consistency → standardize formats
 - ▶ Pick one format for:
 - ▶ Dates
 - ▶ Program names
 - ▶ Categories
 - ▶ Create a simple reference list

**Start here**

If data doesn't match across sources

- ▶ Usability → build one dataset
 - ▶ Identify your most common report
 - ▶ Create one clean dataset that supports it
 - ▶ Reuse it instead of rebuilding each time

**Start here**

If reporting takes too long

Good Data Structure

- ▶ One row = one record
- ▶ Each row represents a single event:
 - ▶ One service delivered
 - ▶ One interaction
- ▶ If multiple services occur → use multiple rows
- ▶ One column = one variable



- ▶ Each column stores one type of information:
 - ▶ Date
 - ▶ Service type
 - ▶ Program
- ▶ Avoid combining values (e.g., “Counseling, Housing”)

✗ Before (Hard to Use)

Client Name	Services Provided	Dates
John Smith	Counseling, Housing	Jan, Feb
A. Brown	MH / Case Mgmt	1/5, 2/10

Issues:

- Multiple services in one cell
- Inconsistent labels (“MH”, “Counseling”)
- Dates unclear and inconsistent
- Cannot easily count services

✓ After (Analysis-Ready)

Client_ID	Service_Type	Service_Date	Program
1001	Counseling	01/15/2024	Outpatient
1001	Housing	02/10/2024	Housing Support
1002	Case Management	01/05/2024	Outpatient

- ▶ What Improved:
 - ▶ One service per row
 - ▶ Standardized service names
 - ▶ Clean, consistent date format
 - ▶ Easy to filter, group, and report

Quick Rules

- ✓ Each row = one person, service, or event
- ✓ Each column = one variable
- ✓ Use consistent naming and formats

✗ Avoid:

- ▶ Multiple values in one cell
- ▶ Merged cells
- ▶ Free text where categories are needed

Data Infrastructure Tools

You don't need fancy tools to start. A well-structured spreadsheet is often enough.

Basic Tools

- ▶ Excel, Google Sheets
 - ▶ Best for: small teams, simple reporting
 - ▶ Tips: one master dataset, avoid duplicates

Intermediate Tools

- ▶ Airtable, Smartsheet, basic CRM
- ▶ Best for: multiple users, structured entry
- ▶ Benefits: forms, standardization, less cleanup

Advanced Tools

- ▶ Salesforce, Apricot, case management systems
- ▶ Best for: large programs, complex reporting
- ▶ Watch out: over-customization, maintenance burden

Choosing the Right Tool

- ▶ How many users?
- ▶ How complex is reporting?
- ▶ Need real-time access?
- ▶ Can you maintain it?

Key Principle

- ▶ The tool matters less than the structure.
- ▶ A messy Salesforce system is worse than a clean Excel file.