Types of Software Needed for Data Management

Table of Contents

INTRODUCTION

SGIGs may need several software and procedural solutions (i.e., policies, processes, or changes to their workflow) to implement complete data systems across the data management lifecycle. Many of these systems are considered "hybrid" systems that can be used for more than one component of the lifecycle. For example, a hybrid system may allow you to collect and store data, or collect and analyze data in a single system.

This document discusses tools for each component of the data lifecycle apart from storage and deletion/destruction, which are covered in the downloadable documents "Data Storage," "Data Backups," and "Data Preservation". This document discusses database systems, data collection tools, tools to use and analyze your data, and data transfer tools.

DATABASE SYSTEMS

Database systems are the systems your data is entered into, stored in, and extracted from. Database systems ensure data is entered and catalogued systematically. These systems define the properties of how data is stored and how it can be queried, searched, or accessed. These systems also play a role in defining the quality, integrity, and reliability of your data. Database systems vary considerably in their level of integration with other systems and processes, the expertise required to use them, and their flexibility in meeting organizational requirements. Options include:

- **Specialized, off-the-shelf, hybrid systems** provided by an external service provider. These often take the form of full hybrid systems and include features like analysis, visualization, and data collection tools. They are meant to solve a specific business need or functional area. These may include software for specific types of work processes (e.g., financial management systems, systems for managing membership data, systems for social services case files, etc.) as well as purpose-built solutions for functions like survey design and administration.

  Implementation considerations:
  Typically, these will be the lowest *total cost* solutions, due to thoughtfulness in data structure design and additional built-in functionality such as intake forms, data

  analysis or visualization. Unfortunately, these products are typically designed for a wide user base, and it is unlikely that these solutions will perfectly meet your government's needs. In particular, they may not meet your government's privacy or auditability requirements.

  These solutions are the most likely to incur significant **vendor lock-in** challenges (see the downloadable document "Procurement – Total Cost of Ownership"), where the system provider owns or controls the data and changing to a new system is challenging.

- **User-managed hybrid database/intelligence systems** such as FileMaker Pro, AirTable or Microsoft Access. Typically, these tools require governments to design the shape or structure of the data being collected (i.e., the fields, their data types, what order and constraints are to be applied), and then provide data intake forms, querying, and export tools specific to that design.

  Implementation considerations
  User-managed hybrid systems have the lowest barrier to entry amongst "custom" systems available, and will often reflect the lowest cost solution.

  These tools can be deployed with minimal training, but it is recommended that you involve experts in both the chosen system and

2

database design broadly, in order to assure good privacy, security, and data structure. Inappropriate decisions made at this stage will impact the data quality, usability and full data lifecycle and can often be very costly to fix after-the-fact.

- **Fully custom or relational database management systems (DBMS)**, used for complex organizational workflows, including solutions like Oracle, Microsoft SQL Server or PostgreSQL, as well as tools like column stores, NoSQL databases, and high-density processing tools.

  Implementation considerations
  Applying a full DBMS grants you the largest degree of flexibility and integrity assurance over your data but will require substantial specific expertise and typically an ongoing procurement cycle with a service provider. If your organization does not have database management expertise in-house, consider this solution only for high-value data with complex needs which may fit into multiple systems' workflows. Often, you may find that custom solutions built on a DBMS product are feasible in partnership with other governments and become **specialized hybrid systems** custom designed for many users.

  The costs, benefits, and risks of applying a full DBMS can vary considerably and depend on things like the need to maintain and support software implementation and the expertise required to make modifications. Vendor lock-in can be less of a concern, so long as the source code and accompanying documentation is owned by the government and available to a future expert.

DATA COLLECTION TOOLS

Data collection tools are often a simple frontend system for your database system. For example, in hybrid systems, it is nearly always the case that data is input through a user interface that is part of the system itself. Some examples of data collection tools include:

- Paper intake forms that need to be entered into the data system.

- Survey tools, such as Qualtrics or SurveyMonkey, used to rapidly create and distribute structured data collection tools.

- Forms management tools, such as Google Forms, Airtable, or TypeForm.

Implementation considerations
You may need to make use of other data collection tools, depending on your government's needs and existing systems, which means that data collection may take place in multiple systems or separately from an authoritative database. When considering data collection tools, it is best to try to design systems where there is a single authoritative source for each piece of data (see the downloadable document "Data Quality" for details).

TOOLS TO USE AND ANALYZE YOUR DATA

Data analysis and visualization applications: The software used to analyse and visualize your data. If your analysis is simple, you may simply require a spreadsheet program like Microsoft Excel. If your analysis is more complex, you may require statistical software or an analytics program to help visualize your data. These applications can be used to create reports and graphic representations of your data, such as charts and graphs, that help to communicate what your data is telling you about your community. Examples include:

- Worksheet tools such as Microsoft Excel or Google Sheets.

3

- Business intelligence suites such as Tableau and PowerBI.

- GIS ("geographic information systems") and their presentation platforms such as ArcGIS.

- Programming environments such as R and Python

- Statistical analysis software such as R, SPSS, SAS, and STATA.

When analyzing data, it is strongly recommended that you acquire a statistical analysis software (R, SPSS, SAS, STATA, etc.). An advanced statistical program will allow you to analyze large amounts of data, produce reproducible scripts (i.e., documented records of your analyses and how you conducted them), and provide you with greater data visualization tools.

Worksheet tools such as Microsoft Excel should ideally not be used for data analysis. Working with and analyzing data in a worksheet tool changes the data in the worksheet, whereas statistical programs leave the original data untouched. Should your government wish to use a worksheet tool for data analysis, it should develop tools/strategies/policies to ensure data is not lost or corrupted (i.e., locked original data sets, file structure, etc.).

For some types of visualization, interactive dashboards such as those constructed by Tableau, PowerBI, or ArcGIS may be a forward-looking tool for your government. For more discussion of this sort of presentation, please see the Acquiring and Working with Data section of the Toolkit.

DATA TRANSFER TOOLS

Encryption, anonymization, and sharing tools: When you share data outside of your government or have to store or transmit data over the public internet, tools should be used to control access and risks related to the data. Tools include:

- Internal networks, data storage and VPNs (these solutions may also facilitate data usage).

- File encryption and sharing, secure file transfer.

- Anonymization workflows and data sharing agreements.

For a more complete discussion of considerations for data control and access, see the Data Privacy and Security section of the Toolkit.