



OFFICIAL DATA COORDINATOR GUIDEBOOK

City Manager's Budget Office – Data Analytics Team

Visit the DAT intranet page for more information: www.sjcity.net/index.aspx?NID=1096

Contents

1. Background	1
Purpose of this Guidebook.....	1
Why Conduct an Inventory of our Data?	1
What is Data Analytics?	1
What is Open Data?	2
Our Open Data Portal.....	2
2. Roles and Responsibilities.....	3
3. How to Create Your Data Inventory	3
Step 1: Complete Data Source Inventory	4
Step 2: Complete Dataset Inventory	4
Step 3: Review Process	5
4. Next Steps.....	5
5. Appendices	6
Appendix A. Data Source Template and Example.....	6
Appendix B. Dataset Template and Example.....	7

1. Background

Purpose of this Guidebook

The purpose of this guidebook is to provide guidance to Data Coordinators in the City of San José. Data Coordinators should use this guidebook to help them in their new role. We'll update this guide as the role and responsibilities of the Data Coordinators evolve and as we learn more. We appreciate your patience and willingness to remain flexible as we embark on this endeavor together.

Why Conduct an Inventory of our Data?

Across the country, jurisdictions large and small are seeking new ways to use their data. In San José, we are pursuing two initiatives that we hope will lead us toward an enhanced culture of data-driven decision making and increased transparency—data analytics and open data. To facilitate these initiatives, we need to gain a comprehensive understanding of the data that exists across City departments. The City-wide data inventory will help the City of San José achieve this objective.

What is Data Analytics?

Every day, the City of San José collects thousands of data points on everything from traffic to graffiti. The data contains compelling insights about how San José delivers services and conduct the business of government. But in order to distill these insights and transform them into meaningful action, we need a systematic approach for sifting through, analyzing and using our data to improve City services. In many cases, this is already occurring in pockets throughout the City. The City Manager's Budget Office – Data Analytics Team (DAT) will work with departments to bolster these efforts, both in numbers and impact.

Recently, state and local governments have become increasingly adept at using their data to improve service delivery and operational efficiency. Examples abound of data analytics initiatives that improved the lives of constituents or saved a government agency significant time and resources. In New Orleans, the [NoAnalytics](#) initiative recently assisted the local fire department in designing a model to prioritize the distribution of free smoke alarms based on the likelihood of fire deaths and the presence of smoke alarms in a given building. In Chicago, predictive data analytics has helped the city [combat rodent infestations before they occur](#), transforming the process from reactive to proactive. Chicago also recently used data analytics to [prioritize food inspections](#) based on the likelihood of finding severe violations in a given food establishment or neighborhood. The above examples have important implications for public policy, public health, and governance.

By engaging departments across the City, DAT will work with staff to use our data to develop innovative solutions to better serve San José's residents. DAT will undertake data analytics projects utilizing a customer service based approach that treats City departments as our valued customers. Data analytics projects will be conducted using the following framework:

1. Identify the operational challenge or opportunity
2. Make formal arrangements on project scope and DAT and departmental commitments
3. Acquire data for the analysis
4. Conduct the analysis and research
5. Review and pilot the deliverable
6. Fine-tune the deliverable and follow-up with the department as needed

There are a number of other ways in which data analytics can help improve City governance:

Break down silos. Much of the benefit derived from data analytics has come from the sharing of data across departments and across jurisdictions. While data sharing is already occurring in our City, more can be done! DAT will seek to leverage its position in the City Manager's Budget Office to break down silos and facilitate the sharing of data among the various agencies in the City. The Open Data Initiative will

also be a key instrument for success in this regard. (See “Increase internal data sharing” in the next section.)

Increase internal analytical capacity. Continuing to build support for data analytics, and building data literacy and analytical skill sets among City staff are critical components for problem solving through a data analytics approach. We believe in this tenet so strongly that it is included in [our mission statement](#). As much as we love data analysis and would be thrilled to keep all the fun to ourselves, we believe that a city full of data analysts will maximize the positive impact of this work. DAT will explore options to offer professional development opportunities directly related to increasing data literacy and analytical skills and will continue to foster the growth of the new Data Champions group in the City. More information about the Data Champions group can be found on the [DAT Intranet page](#), or you can reach out to DAT Executive Analyst [Erica Garaffo](#) at X5-8126.

What is Open Data?

One of the first questions departments often ask when learning about open data is “Why should my department release its data?” There are a number of reasons, both practical and philosophical, why releasing data can benefit your department and the people it serves.

Stimulate new ideas and services. By releasing open data, City departments may help to stimulate new and innovative ideas from our local technology community. There is great potential for open data to act as the fuel for new solutions and even new businesses that can address common problems or challenges facing those that live in, work in, or travel to the City.

Increase internal data sharing. Open data can also help with some of our internal challenges around accessing data between departments, much in the same way that data analytics can help to break down these silos. Right now, analysts may rely on personal relationships to access data from other departments. San José’s open data site, [Data.SanJoseCa.gov](#), can provide a platform to share data internally. Accessing data from different departments can provide valuable insights into how our city works and how departments may better serve those that live and work in San José.

Simplify Sunshine Requests. Open data releases can be an effective way of responding to requests for data through the City’s Sunshine Ordinance. One open data release may address multiple requests for information than can be repetitive and costly to respond to if addressed on an individual basis.

Reduce unwanted web traffic. Publishing open data can also help reduce unwanted web traffic on department websites, which is often the result of “data scraping” by individuals seeking to obtain data in bulk from the City. This puts unnecessary stress on the City’s technology infrastructure and unneeded burden on City IT staff.

Changing how we use data. Open data can serve as a platform to change how we use, share and consume our data externally and internally, transform data into services, and foster continuous improvement in decision-making and the business of government. Ultimately, open data is about enabling use of data to help support a range of positive outcomes.

Our Open Data Portal

The City’s [Open Data Portal](#) is the next generation of civic technology meant to encourage and inspire the innovative spirit of entrepreneurs, civic hackers, residents and businesses. San José currently has over 175 open data resources available for view and download at [Data.SanJoseCa.gov](#). The data is organized by category and is searchable. The City also provides an application programming interface (API) to encourage the process of developing mobile applications and other uses of our data. Don’t worry... If you aren’t familiar with the term API, you are not alone! You can find out more [here](#).

2. Roles and Responsibilities

Below we included an overview of the roles and general responsibilities in support of the data inventory initiative, open data, and City-wide data analytics projects. The open data policy is slated to be finalized in the fall of 2015. Once published, the open data policy will provide greater detail about the use of open data in the City. In addition, as DAT becomes more familiar with the existing pool of data and challenges facing individual departments, the roles and responsibilities of individual Data Coordinators, Data Administrators, and DAT may change. But don't worry! We know your time is valuable, and we will always strive to minimize the impact on your day-to-day operations.

Role	General Responsibilities
Data Analytics Team (DAT)	DAT is located in the City Manager's Budget Office and aims to provide centralized support for data-related initiatives in the City of San José. DAT's responsibilities include: <ul style="list-style-type: none">• Oversee the Dataset Inventory Initiative• Spearhead data analytics projects with departments• Facilitate capacity building for data analytics throughout the City
Data Coordinators	Data Coordinators will serve as permanent liaisons to DAT and the IT Department. General responsibilities include: ¹ <ul style="list-style-type: none">• Meet with DAT and IT during inventory process• Coordinate the completion of your department's dataset inventory• Act as primary point of contact for inquiries related to open data and data analytics projects• Provide feedback to DAT and IT on data initiatives relevant to your department• Be a data champion!
Data Administrator	The individual in charge of databases, datasets, or information systems. This person has business knowledge of the data and can answer questions about the data. General responsibilities may include: <ul style="list-style-type: none">• Managing the dataset or source and authorizing changes to it• Managing access to and use of the data, including documentation• Managing accuracy, quality and completeness of the data

3. How to Create Your Data Inventory

As a Data Coordinator, you will be responsible for creating an inventory of your department's data through coordination with departmental Data Administrators and other staff. We recognize that this will be, in part, a learning process. As a result, we will make changes to this guide based on your feedback.

The inventory process will include the following three steps:

1. Identification of data sources and completion of the Data Source template
2. Identification of data sets within these data sources and completion of the Dataset Template
3. Review process and gap analysis

Appendices A and B provide an example of the cataloging process for your guidance.

¹ Roles and responsibilities may evolve as new initiatives and projects are developed. These are intended to give a general idea of what to expect as a Data Coordinator.

Step 1: Complete Data Source Inventory

Your data may be housed in a variety of places from information systems or databases to shared drives and folders. This can also include 3rd party vendors and data hosted on vendor systems. Step 1 is about identifying the major data sources in your department.

Questions to help identify and discover data sources:

- What information systems does your department use?
- What databases does your department use?
- What applications capture information or are used in your business processes?
- Are some data resources kept in spreadsheets (on shared or individual drives)?
- What information are we already publishing and where did that information come from?
- What services does your department deliver, and how is information related to those services stored?

For each of the data sources, enter the information in your department's Data Inventory Template on the Data Sources sheet. You will be asked to provide the following details for each of your department's Data Sources:

- Name of Data Source (or System): What is this Data Source called (e.g. PeopleSoft)?
- Data Source (or System) Description: Provide a brief description of the contents of the Data Source. What is its purpose? What kinds of information does it contain?
- Data Administrator: Who maintains this Data Source (may be more than one individual)? Who should be contacted with questions related to this Data Source?
- Date Range Included in Data Source (if applicable): Does this Data Source cover a specific time period? If so, what is it?
- Additional Notes: Do you have any other comments about this Data Source?

See Appendix A for a sample Data Source Inventory Template.

Step 2: Complete Dataset Inventory

Each of your data sources should have an associated dataset(s). Some data sources may actually be datasets themselves. The purpose of Step 2 is to allow Data Coordinators to spend time with Data Administrators in their departments brainstorming datasets to be included on the Dataset Inventory Template. Departments may ask, "what should be included as a dataset?" We've defined a dataset as the contents of a single database table, worksheet, or defined view. For example, if your data source is comprised of tables, listing out the tables may be a good place to start.

The details for each dataset are similar to those requested in the Data Source Inventory but include a bit more detail:

- Name of Dataset: What is this dataset called?
- Dataset Description: Provide a brief description of the contents of the dataset. What is its purpose? What kinds of information does it contain?
- Which Data Source Does This Dataset Come From (if applicable): Does this dataset come from a larger database or system? If so, what system or database?
- Available Data Formats: In what formats can this data be exported (e.g. PDF, CSV, Access, KML, Word, etc.)?
- Data Administrator: Who maintains this dataset? Who should be contacted with questions related to this dataset?

- Divisions/Workgroups: Which divisions or workgroups within the department use this dataset?
- Comments Related to Accuracy, Completeness, and Limitations: Are there any concerns with regard to accuracy, completeness, or consistent entry? Any limitations to this dataset?
- HIPPA or Other Privacy Protections: Is this data protected under HIPPA, or are there other privacy concerns? If there are privacy concerns, but it is not HIPPA protected, please specify.
- Comments: Do you have any other comments about this data? Do you have any ideas about how to use this data for improving operations or service delivery in your department?

See Appendix B for a sample Dataset Inventory Template.

Step 3: Review Process

The last step of the inventory process will consist of a detailed review and gap analysis of each department's submissions. DAT will meet with department Data Coordinators and Data Administrators to review the information and ensure that all data sources and datasets are accounted for, especially those that are directly related to major departmental services and/or operations. While we may never ask for the data in a given dataset, it is important for us to have a comprehensive understanding of all available City data. This will ensure that we are adequately prepared to use our data in new and exciting ways.

4. Next Steps

Following the completion of the Data Inventory Initiative, the City Manager's Office and IT department, will proceed with prioritizing datasets for our open data portal. Criteria for prioritization may include:

- Potential value to San José residents
- Potential value to the San José business and non-profit community
- Potential value to City staff and operations
- System or dataset compatibility with the Open Data Portal.

In addition, throughout the Inventory process, DAT will engage departments interested in pursuing data analytics projects or other data-related initiatives utilizing a customer service based approach that treats City departments as our valued customers. This process will be ongoing and, we hope, institutionalized into the management and operation of the City. The role of Data Coordinators in the ongoing data analytics projects will likely vary by department and project, but we hope that you will stay deeply involved and continue to suggest opportunities for partnerships between DAT and your department. In an effort to generate staff suggestions for data analytics projects, we have included a link to our Data Analytics Project Suggestion Form: <http://www.sjcity.net/index.aspx?NID=1096>

We are looking forward to working closely with you and collaborating on all things data! If you have questions, comments, or suggestions on how we can make improvements, please contact Senior Executive Analyst and Data Analytics Team Lead Kevin Miller at X5-7793 or Executive Analyst Erica Garaffo at X5-8126.

-The City Manager's Budget Office – Data Analytics Team

5. Appendices

Appendix A. Data Source Template and Example

Data Source Inventory Template Directions

Please enter one data source per row in the "Data Sources Template" tab, and fill out the requested information related to each source of data. Before you begin entering your sources, please review the "Template Example" below. We have also included help text below the headers in the template worksheet. If you have questions or concerns, please contact Erica.Garaffo@sanjoseca.gov or x58126.

Column Name and Requested Information

Name of Data Source: What is this Data Source called?

Data Source Description: Provide a brief description of the contents of the Data Source. What is its purpose? What kinds of information does it contain?

Data Administrator: Who maintains this Data Source? Who should be contacted with questions related to this Data Source?

Date Range Included in Data Source (if applicable): Does this Data Source cover a specific time period? If so, what is it?

Additional Notes: Do you have any other comments about this Data Source?

TEMPLATE EXAMPLE: City Manager's Budget Office Data Sources

Name of Data Source	Data Source Description	Data Administrator	Date Range Included in Data Source (if applicable)	Additional Notes
Operating Automated Budget System (ABS)	Contains all operating budget information	Kenn Lee	2002 - Present	System will be replaced soon
Capital Automated Budget System (CABS)	Contains all capital budget information	Bonny Riccobono	2002 - Present	System will be replaced soon
Ordinance Database	Contains all the ordinance information, which is then downloaded into the Financial Management system	Sarah Wood		

Appendix B. Dataset Template and Example

Dataset Inventory Template Directions

Please enter one dataset per row in the "Dataset Template" tab and fill out the requested information related to each dataset. Before you begin entering your datasets, please review the "Template Example" below. We have also included help text below the headers in the template worksheet. If you have questions or concerns, please contact Erica.Garaffo@sanjoseca.gov or x58126.

Column Name and Requested Information

Name of Dataset: What is this dataset called?

Dataset Description: Provide a brief description of the contents of the dataset. What is its purpose? What kinds of information does it contain?

Which Data Source Does This Dataset Come From (if applicable): Does this dataset come from a larger database or system? If so, what system or database?

Available Data Formats: In what format(s) can the dataset be exported (e.g. PDF, CSV, Access, KML, Word, etc)?

Data Administrator: Who maintains this dataset? Who should be contacted with questions related to this dataset?

Divisions/Workgroups: Which divisions or workgroups within the department use this dataset?

Comments Related to Accuracy, Completeness, and Limitations: Are there any concerns with regard to accuracy, completeness, or consistent entry? Any limitations to this dataset?

HIPPA or Other Privacy Protections: Is this data protected under HIPPA, or are there other privacy concerns? If there are privacy concerns, but it is not HIPPA protected, please specify.

Comments: Do you have any other comments about this data? Do you have ideas for how it might be useful?

TEMPLATE EXAMPLE: City Manager's Budget Office Datasets

Name of Dataset	Dataset Description	Which Data Source Does This Dataset Come From?	Available Data Formats	Data Administrator	Divisions/ Workgroups Using This Data	Comments Related to Accuracy, Completeness and Limitations of Database	HIPPA or Other Privacy Protections?	Comments
Proposals	Proposal Budget System	ABS	Excel, PDF	Kenn Lee	CMO Budget, All departments	Requires manual entry by departments/budget staff, is used as a working database and contains iterations and a/b testing of different proposals for a costing perspective, double checked against internal tracking worksheets (Excel) maintained by Budget staff	Working database where departments submit entries and draft entries, eventually informs the proposed budget when downloaded to the Operating Budget System, should not be public	
Mid-Year/Year-End/Annual Report Actions	Mid-Year/Year-End/Annual Report	ABS	PDF	Kenn Lee	CMO Budget	This database is used for formatting purposes for final document production and is double checked against internal tracking worksheets (Excel) maintained by Budget Staff, some of these are then automatically downloaded to the ordinance (such as Mid-Year), while others require manual entry (such as Year-End)	Feeds into the production of the Annual Report (which is published in October of each year), Mid-Year Report (published in February), and the Year-End Clean-up memo (published in June) and the Final MBA published in (June)	