MTC Conceptual Data Model Strategy Planning

by

Victoria Giryavets and Michael Ziyambi

Table of Content

Executive Summary

Approach

Conceptual Model Development Framework

Results

Conclusion

Executive Summary

Integrated Planning Section at MTC uses and maintains an extensive variety of data that is used to support growing number of projects and initiatives. Currently the data is spread between several databases, shared drives, box and personal drives in many different formats. “By 2020 IDC” [International Data Corporation] “predicts the amount of digital information will grow by a factor of 30, the number of files, by a factor of 60”. MTS’s N-drive alone is holding more than half a million files and folders containing more the 60,000 datasets. Furthermore, “IDC surveys find that the time spent searching for information averages 8.8 hours per week.” The integrated planning section typically invest significant amount of time searching for data relevant to a particular project, initiative or task, as well as hire outside consultants to provide that data. The goal of this Analytical Services Internship was to work with staff to develop a strategy for data discovery within the planning groups of the Air District, MTC and its transitional employees. The solution design is not to drastically change how things are done, but rather to improve and streamline the process with clear understanding of rules when it comes to data methodology to eliminate time, effort and cost spend on finding relevant data and greatly enhance the efficiency of the unit.

Approach

Interview with the Head of the Department and at least one main Power Users to understand the need.

Schedule 15-30 minutes one-on-one interviews with the main data users to understand how they use data.

Analyze responses from one-on-one interviews to draw conclusions on the initial questions and methods to gather information about datasets.

Investigate option for publishing survey: google forms, survey monkey, in-house webform.

Sent out initial survey as an in-house web form and as excel template providing incentive such as prize to complete with in shorter timeframe.

Schedule short one-on-one interviews with all users that didn’t respond to the survey.

Gather and document feedback on the Initial Data Survey.

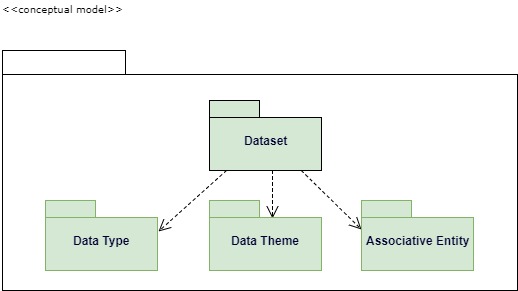
Analyze at least one source of data files that participants frequently referred to as location to evaluate whether or not comprehensive inventory of all data files/datasets is valuable:

Come up with Conceptual Framework Development

Document the Results

Present findings

Conceptual Data Model Development Framework



Dataset with Types and Themes and its Associative Entity (such as project or relationship to other datasets, projects, etc)

Data Types: Spatial, Non Spatial, Containing Spatial-Compatible field

Data Themes:

**Current:** Administrative, Demographic, Policy, Transportation, Environmental, Projects

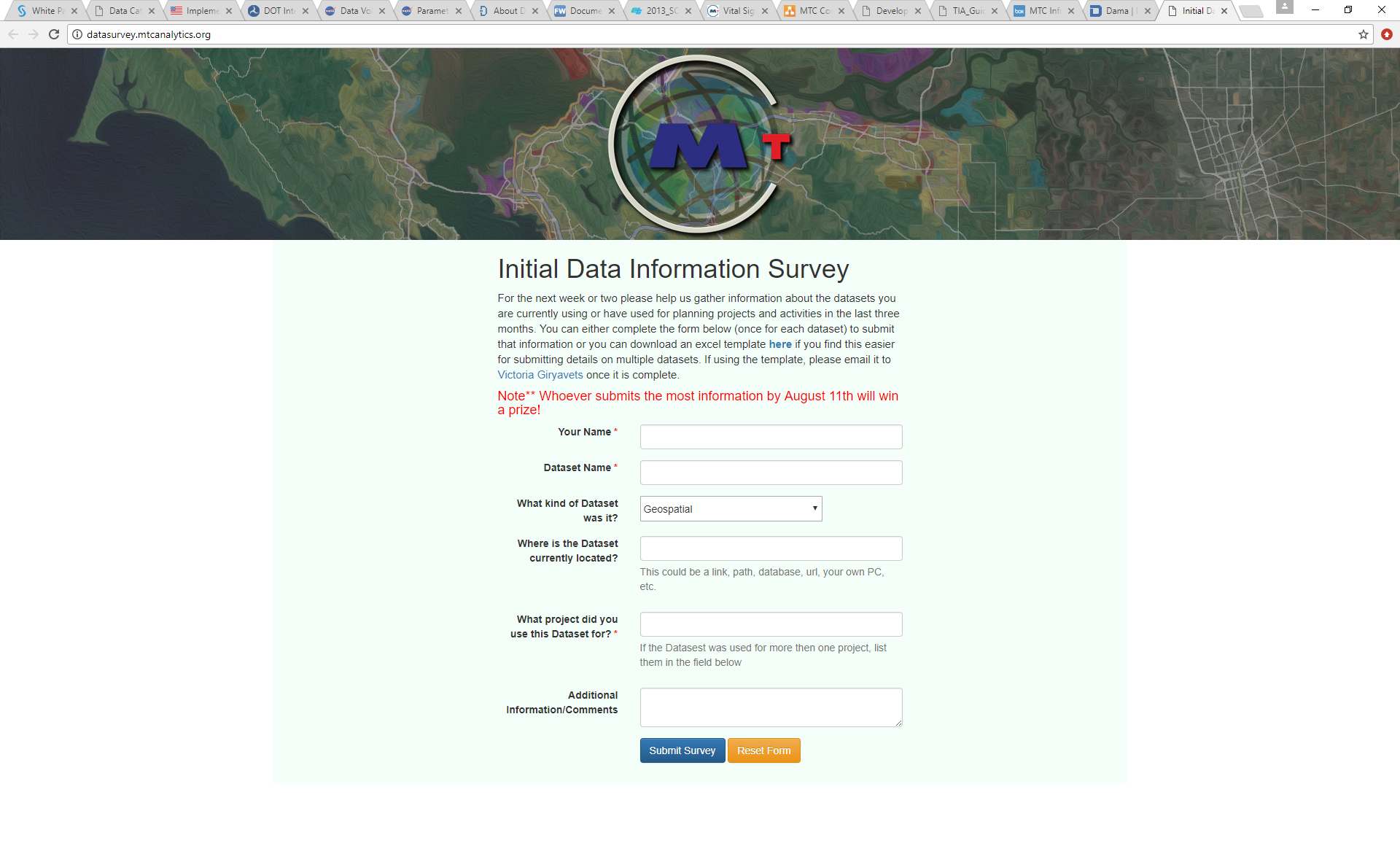
|  |  |  |  |
| --- | --- | --- | --- |
| Bay Trail | Waze RSS Feed | Employment Statistics | Congested Corridor Analysis |
| Bay Trail Grants | RM3 | Occupational Projections | PeMs |
| PCA | Bay Trail Shapefiles | Income/Employment Data | Freeway Reliability Analysis |
| TAZ | Natural Hazard Layers | Residential Permit Data | StreetSaver |
| MAZ | Infrastructure Assets | County Business Patterns | Regional Transit Capital Inventory |
| Bay Trail GIS Layers | Vital Signs external data | TIGER GIS files | CIRB Building Permits Data |
| P-Tap | Death Records (Vital Signs) | Traffic Counts Data | ES-202 Confidential Data |
| Walkscore | Bike Transit Score | California Assembly Districts | Long Term Neighborhood DB |

**Possible:** Emergency Operation, Hazards, Structures, Government Units, Utilities, Addresses and Names, Transportation, Cadastral, Hydrography, Environmental, Land Use/Land Cover, Basemap, Elevation, Imagery, Geodetic Control

Results

**Initial Data Information Survey**

Webform <http://datasurvey.mtcanalytics.org/>



* Survey Participation results: 37% of people completed it or responded.(Michel Smith’s submission was past the Prize Deadline)

Data can be quantified with volume, variety and velocity. Pattern from initial survey suggests that MTS planning department uses a high volume and High Variety of micro sets of data, analysis is performed on variety of data files from multitude of sources.

Approximately 30% of files actually should be inventoried and only 10-15% will be published. This trend might carry through any other source/location of datasets. N-Drive is an unstructured source of information thus makes it difficult to locate and use correct information.

**Comprehensive Data Inventory Cost and Benefit**



Evaluation

Techniques used

* In Person Interviews
* Web Survey
* Excel Template
* Mass email
* Personalized Emails
* Follow up strategy

What worked and what didn’t

Conclusion

It was difficult to accomplish the given task of developing the conceptual data model due to that lack of participation and limited time of the internship compared with the level of desired objective. However, a great beginning for the future of the data driven decision-making was made.

**Suggested Stages for Data Model Strategy:**

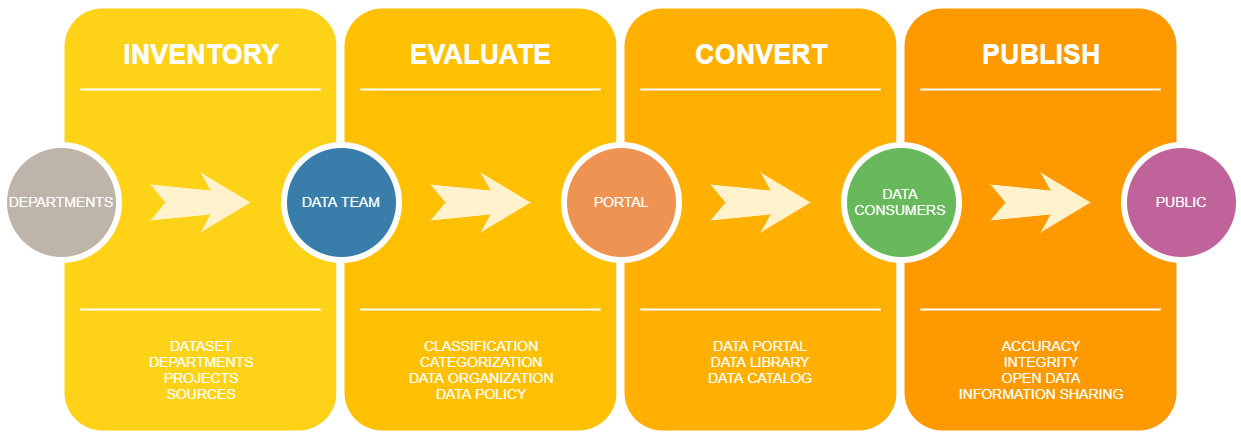


Figure 2: Victoria Giryavets (Analytical Services Intern)

**Smart Data Tracking Mechanism**

**Initial Data Audit**

**Dataset Catalog**

**Data Library**

Case Studies

Massachusetts Department of Transportation

City of San Francisco

United States Department of Transportation

U.S. General Services Administration, Technology Transformation Service

National Aeronautics and Space Administration

Federal Geographic Data Committee

Valley Transportation Authority

State of New York Open Data

State of California Open Data

Chicago Regional Transportation Authority

Rudin Center for Transportation Policy and Management