



OPPORTUNITIES WITH OPEN DATA

Dr. Cynthia V. Marcello

WHAT IS OPEN DATA

- According to the [OpenDataHandbook.org](https://opendatahandbook.org) website,
Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

MOST COMMON TYPES OF OPEN DATA

- Datasets (CSV, EXCEL, JSON, XML, TEXT)
 - Map data (SHAPEFILES (POINTS, POLYLINE, POLYGON, MULTIPOINT))
 - Application Programming Interface (API)
-

USES FOR OPEN DATA

- **Transparency and democratic control**
- **Participation**
- **Self-empowerment**
- **Improved or new private products and services**
- **Innovation**
- **Improved efficiency of government services**
- **Improved effectiveness of government services**
- **Impact measurement of policies**
- **New knowledge from combined data sources and patterns in large data volumes**

CONSIDERATIONS WITH OPEN DATA

- Data cleaning preprocessing, and feature engineering is often needed before using open data.
 - Data anomalies, incompleteness, general quality should be verified before using open data.
 - Open data is all about closing the “data divide” to permit all to access and innovate to create solutions to problems.
 - Many open data sources have no copyright protection and are in the public domain.
 - Transparency requires responsibility – use a risk-based approach to open data.
 - Evaluate the dataset by asking –
 - Who is the dataset provider?
 - When was it published and where?
 - What’s in it and how is it shared?
-

LINKS AND RESOURCES

- THE WORLD BANK - <http://opendatatoolkit.worldbank.org/en/essentials.html>
 - OPEN DATA HANDBOOK - <https://opendatahandbook.org/>
 - 50 BEST OPEN DATA SOURCES - <https://learn.g2.com/open-data-sources>
 - OPEN DATA NEW YORK STATE - <https://data.ny.gov/>
 - OPEN DATA NEW YORK CITY - <https://opendata.cityofnewyork.us/>
-