

Sarang Chernov

LIS545b

Friday 1/27/2023

Word Count: 1000

Term Project:

Data and Metadata Profile

DYCD after-school programs: Fatherhood

The Dataset

Data.Gov - <https://catalog.data.gov/dataset/dycd-after-school-programs-fatherhood>

NYC Landing Page - <https://data.cityofnewyork.us/Social-Services/DYCD-after-school-programs-Fatherhood/esw6-z4id>

Dataset Contents

The dataset titled “DYCD after-school programs: Fatherhood” contains information regarding New York City (NYC) facilities offering Family Support after-school “Fatherhood Initiative Programs” for young adults ages 16-24, listed by agency and site (Data.Gov, 2020).

Data Origin

Department of Youth and Community Development (DYCD)

Key Stakeholders

- NYC fathers ages 16-24 who are eligible to or do attend Fatherhood Initiative Programs
- Families and support people for the young fathers
- DYCD
- City of New York
- Dataset owner: New York City OpenData
- The agencies and sites listed in the dataset

- Other NYC social and health services with a shared client-base
- Other cities interested in similar programs

Data Files

This dataset contains a single data file.

The file contents are in a tabular data format with 19 rows and 18 columns. Columns are titled Program Type, Program, Site Name, Borough/Community, Agency, Contact Number, Grade Level/Age Group, Location 1, Latitude, Longitude, Community Board, Council District, Census Tract, BIN, BBL, NTA, Postcode, and Address. This is true both on the landing page and the downloadable data on the data.gov page. The landing page also offers a “View Data” button. When selected, 5 additional columns appear for Zip Codes, Community Districts, Borough Boundaries, City Council Districts, and Police Precincts. There is no explanation stated for this apparent update to the dataset.

The formats available for download on the data.gov page are CSV, RDF, JSON, or XML. On the landing page, the dataset is also available for export as CSV, CSV for Excel, CSV for Excel (Europe), TSV for Excel, RDF, RSS, or XML formats.

Usage Restrictions

The intention behind sharing this dataset is to make it available for public access and use. While no usage restrictions are clearly listed, there are terms of use. The dataset was found on data.gov, but those terms of use do not apply because the dataset is non-federal. NYC has its own terms of use that are legal formalities regarding representation and usage of the NYC website.

Software

The original format of the dataset is XLSX, so the easiest software to use to open it is Excel. The XLSX format can be used with other software and there are other formats available on data.gov and the landing page.

Metadata: Location, Extent, & Standards

Metadata is provided by two sources: data.gov and the NYC landing page. The metadata offered by data.gov is more comprehensive than that of the landing page and has been updated more recently. The metadata on the landing page addresses usage information (see

Figure 1), while the metadata on data.gov seems to cover more detailed documentation regarding the dataset (see Figure 2).

| | | | | | |
|--------------|------------------|---------------|-------------------|-----------------------|-------------------|
| Date Updated | Data Provided by | Dataset Owner | Data Last Updated | Metadata Last Updated | Number of Rows |
| Date Created | Views | Downloads | Agency | Update Frequency | Number of Columns |
| Automation | Date Made Public | Attachments | Category | Tags | |

Figure 1: Metadata on the NYC landing page

| | | | | | | |
|--------------------|-----------------------|-----------------------|---------------------|----------------------------|----------------|-----------------------|
| Topics | Contact | Social Sites | Federal/Non-Federal | License | Downloads | Resources |
| Resource Type | Metadata Created Date | Metadata Updated Date | Publisher | Maintainer | Identifier | Date First Published |
| Date Last Modified | Category | Public Access Level | Metadata Context | Metadata Catalog ID | Schema Version | Catalog Described by |
| Harvest Object ID | Harvest Source Id | Harvest Source Title | Homepage URL | Source Datajson Identifier | Source Hash | Source Schema Version |

Figure 2: Metadata on data.gov

A single Data.json file containing metadata and data is available for download on data.gov, but not the landing page. The landing page provides a downloadable spreadsheet; however, it is only partially filled out with general dataset information. The spreadsheet's sections for Detailed Description, Column Information, and Revision History are blank.

A metadata standard was likely used to structure the metadata since the dataset is available for download in formats like JSON and XML. The XML file states, "This XML file does not appear to have any style information associated with it." It is unclear whether or how the metadata is structured according to any metadata standard.

Enriching the Data & Metadata

Additional information could enrich both the data and metadata. To improve users' ability to discover the dataset in the repository environment, it would help to create a DOI. The inconsistency of metadata could be remedied by comparing data.gov information with the

landing page and updating the pages accordingly. The data updates also need consistency. If 5 columns are added to the dataset table in one location, it should be matched and/or explained elsewhere.

The metadata spreadsheet available on the landing page needs to be finished to assist those unfamiliar with the data make use of it for new purposes. There is a blank section for Detailed Description that intends to provide context for someone not familiar with the agency's operations. On the same spreadsheet is the opportunity to define terms or acronyms, but that is blank. If the dataset Revision History were filled out, that would provide further context for future use of the data.

To make these changes, the DataOne Best Practices Primer could be used to guide the creator/curator through the management of this dataset. These include best practices for each part of the data life cycle. These parts are the following actions: plan, collect, assure, describe, preserve, discover, integrate, analyze (DataOne, n.d.). These best practices could be an iterative approach to following up with and improving this dataset.

Publications

Since there is not a DOI, it is difficult to research whether publications were written based on this dataset. Search terms such as the title showed no results on databases such as DataCite and ProQuest. A Google search brought up the dataset from different sources and included some publications that referenced the program the dataset is about, however there were no clear uses or citations of the dataset itself. There are views and downloads counters on the landing page that suggest a level of interest. The dataset has been downloaded 2,462 times as of this assignment. This leads me to believe there could be publications relating to this dataset out there, but that the metadata available does not make searching easy.

Sources:

DataOne. (n.d.). Best Practices of Data Management. *Data Management Skillbuilding Hub*.
<https://dataoneorg.github.io/Education/bestpractices/>

DYCD after-school programs: Fatherhood. (2020). *Data.Gov Data Catalog*.
<https://catalog.data.gov/dataset/dycd-after-school-programs-fatherhood>

NYC OpenData. (2011). DYCD after-school programs: Fatherhood. *NYC OpenData*.
<https://data.cityofnewyork.us/Social-Services/DYCD-after-school-programs-Fatherhood/esw6-z4id>

Terms of Use. (n.d.). NYC.Gov. <https://www.nyc.gov/home/terms-of-use.page>