Metadata Standard

Committee on Information Technology

DataSF hosts datasets for public use and dissemination. These datasets include metadata (i.e. data about our data), which helps users search, find and understand our published data. Metadata can also help summarize and track our data publications via basic metrics (e.g. datasets published by department). Our current metadata elements are essentially "out of the box" and missing key elements and in some cases, controlled vocabulary.

PURPOSE AND SCOPE

The metadata standard was the result of a comprehensive survey of existing and best practices and the contributions of a working group. The COIT Architecture and Policy Subcommittee approved the standard. The table below summarizes the standard:

REQUIREMENTS

Group	Purpose	Fields
Basic Descriptive Information	Provide the core information to describe the dataset, including the source department. Each of these fields help our users discover and distinguish between datasets.	- Dataset title - Description - Category - Department
Detailed Descriptive Information	Support informed use of the data. They allow users to assess the appropriateness of the dataset for their needs (including data coverage, size and other details), address common questions or misconceptions, and provide a means of conveying additional detail.	- Data dictionary - Row count - Geographic unit - Temporal coverage - Tags - Program link - Data notes - Related documents
Publishing Details	Allow users to understand what to expect in terms of how often the data is updated and its relative "freshness". This informs how the data can be used and helps users assess if it is appropriate for their desired use.	- Last updated - Frequency of data change - Frequency of data publishing - License and rights

---

COIT Policy Dates

Approved: October 16, 2014
Next Review Date: FY 2018-19

Web & Technical Information	Provide web and technical details that support web or application access to the dataset. These fields are heavily used by programmers and administrative users of the data platform.	- Unique identifier - Permanent link - URL - Endpoint (for APIs) - Download URL - Format
Internal Management	Support internal management of datasets for publishing datasets and answering data questions. These metadata fields are private.	- Public access level - Public access comment - Data steward name - Data steward email