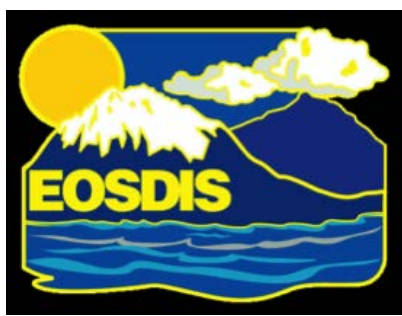
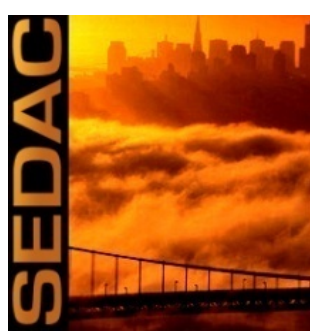


Interdisciplinary Data Quality Information

Robert R. Downs ¹



NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN), Columbia University,
¹ rdowns@ciesin.columbia.edu



Prepared for Presentation to the 2022 ESIP January Meeting
January 18-21, 2022, Virtual Event
Research Showcase: January 19, 2022 4:30 p.m. - 6:00 p.m.

Information about the quality of data and information resources offers potential users of data with capabilities to determine whether the data can meet their needs. Likewise, quality information enables users to understand how data might be used to achieve their goals. Providing data quality information that can be understood by representatives from various disciplines offers opportunities to enable interdisciplinary use of data products and services. An analysis of interdisciplinary data quality information is described.

A pilot study identified 67 terms used to describe the quality of 21 socioeconomic datasets that began being disseminated by SEDAC, the NASA Socioeconomic Data and Applications Center, during 2016 and 2017 for interdisciplinary use. An initial aggregation of the identified data quality terms identified 8 categories for organizing socioeconomic data quality terminology. The categorization of the identified data quality terms is presented to inform discussions about terminology that is used to describe data quality for interdisciplinary data use. Increasing knowledge about the terminology used to describe data quality for interdisciplinary use may contribute to the development of a vocabulary and a classification of terminology for describing data quality to facilitate interdisciplinary understanding and use of data products and services.

Categories of Terminology Used to Describe Data Quality of Interdisciplinary Socioeconomic Data

Caveats

Accuracy, assumptions, appropriate, bias, caveats, comparison, confounding factors, data challenges, deviations, difficulties, evaluation, exceptions, exclusions, inconsistencies, known issues, inflation, limitations, other factors, problems, quality issues, quality problems, rationale, uncertainty, undercounts

Corrections

Adjustments, corrections, estimation, improvements, substitutions

Errors

Commission errors, errors, possible errors, small errors

Missing Data

Gaps, incompleteness, missing values, no data, omissions, unavailable data

Modification

Aggregation, backcast, coarsening, currency, disaggregation, filters, matching, projections, proportions, thresholds

Use

Applicable use, appropriate use, constraints on use, implications for use, inappropriate use, recommended use, suitability for use, usage issues

Quality Control

Data quality indicators, log of changes by version, quality assurance, quality checking, quality control, references on methods, validation

Sources

Alternative sources, data sources, sources’ quality