GUIDE-D: Guidelines for Unifying and Improving Data Exposition and Description

Dr. Alaina Pearce

Section	Description		Page Number	Table/Figure Number
Introduction	☐	Describe purpose of the dataset		
	☐	Describe relevant societal and/or scientific value of dataset • Why is data needed? • What value does data provide beyond your primary research paper?		
Related Resources	☐	Associated publications disclosed (e.g., research papers using the data)		
	☐	Companion project repository linked (if applicable)		
	☐	Companion code repository linked (if applicable)		
Methods: General	☐	Study design described		
	☐	Ethical approvals and protocols described		
Methods: Sample (for animal and human subjects)	☐	Description of participants/subjects		
	☐	(Human subjects) Participant characteristics table including descriptive statistics (if possible) overall and/or by group for: • Age • Sex • Gender • Race • Ethnicity • Indicators of socioeconomic status and/or social determinants of health		
	☐	Inclusion and exclusion criteria		
	☐	(Human subjects) Consent/assent procedures (human subjects only) including: • Consent procedures for data reuse and sharing		
	☐	(Human subjects) Describe any data access controls/limitations		
	☐	(Human subjects) Procedures/steps taken to ensure confidentiality is maintained		
Methods: Protocol	☐	Clear description of data collection protocols including: • Dates • Locations • Equipment used • Software used		

	☐	Variables measured and associated protocols including: • Data acquisition protocols (e.g., task designs, sampling methods, etc.) • Rational for each variable • Relevant citations for methods/protocols		
	☐	Describe metadata location/convention. Metadata should include: • Variable names and definitions – if a naming convention was used, describe • Variable units and/or allowed values • Levels and definitions for categorical data • Whether the variable was derived (e.g., computed value based on other variables)		
Methods: Data Processing	☐	Steps taken to get data in current (shared) state including software/code used to process data		
	☐	Missing data conventions • Note: it is not recommended to include imputed data because each user should decide the best approach for dealing with missing data		
	☐	Describe how low-quality data points were assessed for each variable (e.g., chance task performance, failed assay, etc.)		
Data Description: Data Access	☐	Repository name(s) and persistent identifier/DOI(s) • Note: if multiple data repositories were needed, include how the data can be linked between repositories		
	☐	License information		
	☐	Access conditions/controls for sensitive data		
Data Description: Overview	☐	General description of data deposited in repository including: • Number of files • Size of files • File formats – open formats preferred • Metadata standard used		
	☐	Structure of data in repository including: • Directory structure		

		• Links/relationships between data files • Naming conventions and/or data standard used • Important metadata files		
Data Description: Key Variables	☐	Descriptive statistics about key variables such as: • Observed value ranges • Central tendency and variability/variation • Number of missing values • Number of low quality/flagged values		
Technical Validation/Quality Control	☐	Quality control metrics and/or protocols for each variable or set of variables (e.g., blinding, randomization, etc.)		
	☐	Validation of data (e.g., cross-checks, error analysis, signal quality)		
	☐	(if applicable) Phenotypic or genotypic assessments of biological samples (e.g., to confirm cell line or disease status)		
	☐	Known systemic issues or limitations to data		
	☐	Describe how sample-specific recommendations for inclusion/exclusion are documented (e.g., in a data file or in metadata)		
Discussion	☐	Describe the data's unique value (e.g., size, sample, resolution, diversity of measures), novel and/or expensive methods)		
	☐	Describe how the data fills a resource gap if one exists		
	☐	Describe how the data can be applied and/or reused: • Possible research questions the data could help answer • If relevant, describe interdisciplinary uses for the data • If relevant, describe data standardization and note ability to integrate with other datasets		
	☐	Describe limitations or potential pitfalls for users • Any guidance or recommendations to address these when data is used		

		• Warnings about out-of-scope analyses or potential misuse		
Funding	☐	All funding sources with grant numbers, if applicable		