IEEE Data Descriptions

An Introduction

IEEE-DATA An Introduction: Outline

1. Why?
2. Case Study
3. Motivation
4. Journal Scope
5. Article Types

6. Article Sections
7. Reviewer Evaluation
8. Website Walk-Through
9. Article Walk-Through
10. Questions

IEEE

Why publish an article about your dataset?

For the data:

▸ Foster innovation and reproducibility in research

▸ Demonstrate dataset value and encourage collaboration

▸ Increase dataset visibility, citations, and academic impact

▸ Highlight unique dataset contributions

For you:

▸ Get the academic credit of having a journal paper with citations

▸ Increase visibility of open science presence and academic impact

◆IEEE

Case Study: The AMPds2 and Scientific Data (1/2)

▸ The Almanac of Minutely Power dataset Version 2 (**AMPds2**)

▸ **Multi-modal data** (electricity + gas + water) unlocks richer insights into household behavior

▸ **11 electrical measurements**: voltage, current, frequency, real/reactive/apparent power and energy, power factors

▸ **File formats**: CSV with timestamps, clear directory structure, and accompanying ISA-Tab metadata

▸ **Data cleaning**: sensor calibration and minor gap-filling to ensure continuity

▸ **Metadata** includes meter location, units, and sampling methods for reproducibility

◈ **IEEE**

Case Study: The AMPds2 and Scientific Data (2/2)

▶ Inspired **AMPds2 (Sci Data 2016)**: 2 years of data with environmental and billing records

▶ **Over 380 citations** across IEEE, Elsevier, and MDPI publications

▶ Adopted in **42 NILM datasets surveys** highlighting its methodological importance

▶ Was referenced in **2 fundamental time-series analysis** papers

▶ Detailed **quality control** with error-rate statistics and algorithm benchmarking

DOI: 10.1038/sdata.2016.37

Cited by 384

2016 2017 2018 2019 2020 2021 2022 2023 2024 2025

5
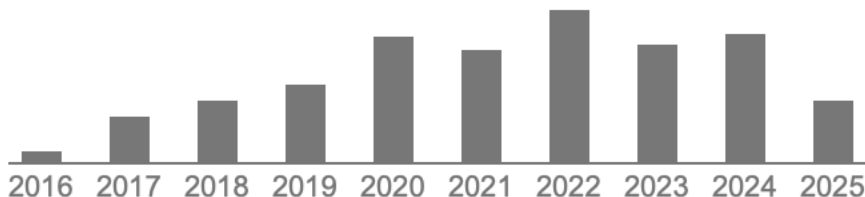
Motivation from Case Study

With the data descriptor journal article:

▸ **Detailed metadata** is essential for reproducibility—incl. units, device specs, and sampling details.

▸ **Open licensing** and clear directory structures drive widespread adoption and citation.

▸ **Fosters cross-disciplinary collaboration** by allowing researchers from diverse fields to integrate (e.g., AMPds) into novel analyses.

▸ **Drives unexpected insights**, as open datasets enable fresh perspectives and innovative reuse beyond the original scope.

▸ **Includes a detailed user manual** covering directory structure, metadata schema, processing scripts, and usage examples—ensuring researchers can quickly onboard and apply the data.

▸ Fostered significant **innovation and reproducibility** in power and energy research.

◈IEEE

IEEE-DATA Journal Scope

▸ Publishes peer-reviewed short articles on datasets: Descriptors, Collections, and Meta

▸ Promotes reproducible science through public data dissemination

▸ Covers data collection methods, quality, metadata, and insights

▸ Supports FAIR principles and future data discovery tools

▸ Accepts data from IEEE DataPort and other persistent/permanent repositories.

Overview of Article Types

There are 3 article types:

▸ **Descriptor** Articles: Detailed description of a dataset

▸ **Collections** Articles: Detailed description multiple related datasets

▸ **Meta** Articles: Discussion of data topics and standards

◈IEEE

Descriptor Articles (Detailed)

▸ Document dataset methods and sources

▸ Validate data quality and accuracy

▸ Describe data structures and storage

▸ Provide insightful usage notes

▸ Example: AMPds Descriptor for previous case study

IEEE

Collections Articles (Detailed)

▸ Structured similarly to Descriptor articles

▸ Include collective statistics and metadata

▸ Ideal for datasets from competitions, hackathons, or multiple-source datasets.

◈IEEE

Meta Articles (Detailed)

▸ Meta articles are articles that **allow for discussion** about data.

Examples:

▸ **standardized formats** for data in a given area
  – (e.g., collecting power data from sub-stations)

▸ **surveys** of a number of datasets of a given type
  – (e.g., GIS, time-series, etc.)

▸ data that are a **library of models** for a given AI or machine learning algorithm or system

▸ data or dataset that is **metadata** of existing published datasets

▸ **comparing** a number of similar datasets in an area to create meta-statistics data

▸ any other data **topics that may not have traditional** scientific experimental results

◈IEEE

Required Sections (Descriptor & Collections)

▶ Abstract

▶ Background

▶ Collection Methods & Design

▶ Validation & Quality

▶ Records & Storage

▶ Insights & Notes

▶ Source Code & Scripts

▶ Acknowledgements & Interests

▶ References

IEEE

Recommended Sections (Meta Articles)

▸ Astract

▸ Background

▸ Proposed Methodology

▸ Applied Analysis

▸ Discussion of Results

▸ Conclusions

▸ Source Code and Scripts

▸ Acknowledgements & Interests

▸ References

◈IEEE

Article Title Strict Format

Descriptor: *The Name of My Dataset (TNM)*

Collection: *Datasets from IEEEXtreme 2023*

Meta: *A Critic of Medical Image Resolutions in Publicly Available Data*

Abstract - Best Practices

▸ 150—250 words, and references/citations are not allowed

▸ They cannot claim new scientific findings or experimental results

Abstract for **Descriptor** and **Collections** articles:

▸ Considered the *Introduction* section in traditional articles.

▸ Briefly introducing or describing it.

Abstract for **Meta** articles:

▸ Considered the *Problem* and *Hypothesis* section

▸ Scientific findings or experimental results are the for the *Conclusions* section

◈IEEE

Background - Best Practices

▶ Must provide an overview of the data collected

▶ Discuss how it fits with other comparable, published datasets

▶ Must make clear the data's value and how it can be reused

▶ Must also summarize any previous publications made using this data with a brief summary and citation for each time used.

▶ **Do not** include a paragraph on how your article is organized; all articles of this type are organized the same way.

IEEE

Collection Methods & Validation - Best Practices

▸ Must provide details on how that data was collected

▸ Includes details on any hardware/system designs used to collect the data
  – (i.e., data acquisition)

▸ Steps and procedures used to collect and process the data in its final form
  – (i.e., computational processing)

▸ Recommended: provide diagrams that show the overall system/procedure used

IEEE

Validation and Quality - Best Practices

- ▶ Detail the accuracy of data and equipment used

- ▶ Provide error metrics and tables

- ▶ Clearly demonstrate technical soundness and robustness of data

IEEE

Records and Storage - Best Practices

▸ Clearly describe file structures and storage

▸ Summarize files and relationships

Example:

▸ CSV file columns, rows, hierarchical relationships

◈IEEE

Insights, Notes, and Source Code - Best Practices

Provide details on:

▸ Caveats of data use

▸ Special cases

▸ Potential alternate dataset uses

▸ List publicly available source code repositories clearly

▸ Include repository links/DOIs where appropriate

▸ Your dataset DOI/PIDs/links are listed on the first page.

◈IEEE

Acknowledgements and Interests Section

▸ Clearly state individual author contributions

▸ Disclose funding sources, or declare if none exist

▸ Declare conflicts of interest clearly and transparently

▸ This section has a strict format.

◆IEEE

Final Checks Before Submission

▸ Revisit title formatting

▸ Ensure completeness of sections

▸ Verify adherence to formatting guidelines (use the templates at IEEE or Overleaf)

▸ Address previous reviewer feedback thoroughly

▸ Review the Author Guide: https://www.ieee-data.org/author-guide/

▸ Review the Best Practices: https://www.ieee-data.org/best-practices-for-authors/

Submit through IEEE Author Portal

Review Criteria Overview

Reviewers assess whether:

▸ Data methods are robust

▸ Technical quality is supported

▸ Data depth and coverage is sufficient

▸ Processing details enable replication

▸ Information supports data reuse

▸ Reporting standards compliance

▸ Repository suitability

▸ Data files completeness

IEEE

Key Reviewer Evaluation Questions

Yes/No, *if no* **Describe:**

1. Is the data generated using robust methods?

2. Is the technical quality convincingly supported?

3. Is the data's depth and coverage adequate?

4. Are data-processing details sufficient for replication?

5. Is all necessary information provided for data reuse?

6. Does it adhere to journal reporting standards?

7. Is the chosen repository suitable?

8. Are the deposited data files complete?

◆IEEE

1. Is the data generated using robust methods?

▸ Document data acquisition hardware and procedures

▸ Include detailed system diagrams

2. Is the technical quality convincingly supported?

▸ Provide validation metrics, error rates, and QC results

▸ Include supporting figures and tables

3. Is the data's depth and coverage adequate?

▸ Detail temporal/spatial coverage

▸ Compare with similar published datasets

◆IEEE

4. Are data-processing details sufficient for replication?

▸ Describe data cleaning and processing steps

▸ List software versions and scripts

5. Is all necessary information provided for data reuse?

▸ Provide comprehensive metadata and documentation

▸ Define variables, units, and formats

6. Does it adhere to journal reporting standards?

▸ Follow IEEE Data Descriptions templates and formatting

▸ Ensure section order matches guidelines

<IEEE>

7. Is the chosen repository suitable?

▸ Choose a persistent, accessible repository

▸ Include DOI and access instructions

▸ We recommend the dataset published on IEEE DataPort

▸ Other persistent/permanent repositories are acceptable
  – Does the data repository assign your dataset a DOI?

8. Are the deposited data files complete?

▸ Verify all files are present and correctly named

▸ Provide a summary table of files in Records & Storage

▸ Provide access to the editors and reviewers so they can check

◈IEEE

More on adequate data depth and coverage…

Reviewers often get this question wrong:

▶ Pertains to the internal comprehensiveness of the dataset

▶ Asks whether all relevant aspects/elements within the scope of the dataset are fully represented

Example:

▶ A dataset on written language recognition

▶ Adequate coverage means that all letters/syllables of a particular language are included

▶ It does not imply that the dataset must cover every known language

Rather:

▶ it should provide a complete and representative sample of the subject matter it intends to capture.

◈IEEE

Journal Website Walk-Through

https://www.ieee-data.org

IEEE Xplore Page

Descriptor Article Walk-Through

https://ieeexplore.ieee.org/document/10704603

Descriptor: *Simon Fraser University Electric Vehicle Parking Dataset (SFU-EVP)*

STEPHEN MAKONIN [1] (SENIOR MEMBER, IEEE), ISLA ZIYAT [2] (MEMBER, IEEE), RYLEN SAMPSON[1], SIYU AN[1], FRED POPOWICH [1], PATRICK PALMER [2] (SENIOR MEMBER, IEEE), AND DAVID AGOSTI[3]

[1] SFU's Big Data Hub, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[2] School of Sustainable Energy Engineering, Simon Fraser University, Surrey, BC V3T 0N1, Canada
[3] Parking & Sustainable Mobility Services, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

CORRESPONDING AUTHOR: Stephen Makonin (e-mail: smakonin@sfu.ca).

ABSTRACT Simon Fraser University (SFU) aims to make a significant contribution to the study of electric

QUESTIONS?