# Best Practices for Data Publication in the Astronomical Literature
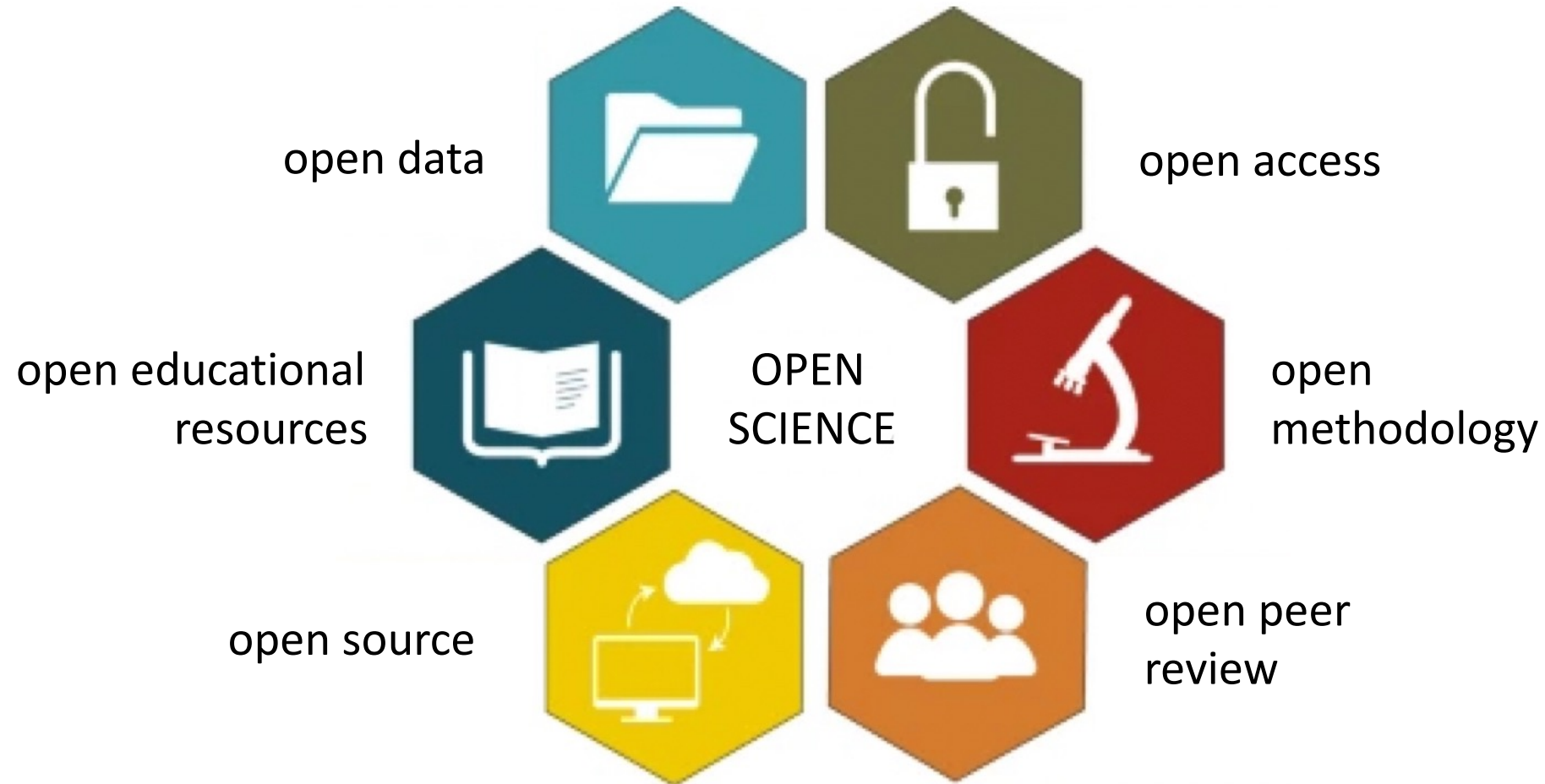
Xiuqin Wu

Caltech/IPAC-NED

# Best Practices for Data Publication in the Astronomical Literature (Chen et al. 2022ApJS..260....5C)

- A broad-based effort led by NED, in response to growth and complexity of data published in the literature

- A call to authors to publish cleaner, unambiguous, and better documented data

- Addresses important FAIR principles

- Strong connection to NASA's Open Science Program: Following BPs will improve the scientific record, enhance reusability, and help streamline ingest into archives

# Core principles of open science



open data

open access

open educational resources

OPEN SCIENCE

open methodology

open source

open peer review

Adapted from Gallagher et al., Nat Ecol Evol 4, 294–303 (2020)
https://doi.org/10.1038/s41559-020-1109-6

# FAIR Principles Addressed

## Findable

- F1: (Meta)data are assigned globally unique and persistent identifiers
- F2: Data are described with rich metadata

## Accessible

- A1: (Meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1: The protocol is open, free, and universally implementable
- A1.2: The protocol allows for an authentication and authorization procedure, where necessary

## Interoperable

- I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2: (Meta)data use vocabularies that follow FAIR principles
- I3: (Meta)data include qualified references to other (meta)data

## Reusable

- R1.2: (Meta)data are associated with detailed provenance
- R1.3: (Meta)data meet domain-relevant community standards

# Open data

- Provide complete names and coordinates for objects studied in an article (§2.1, §2.2)

  E.g., SDSS J1441+0948 --> SDSS J144157.24+094859.1 or

  SDSS J144156.97+094856.5 or

  SDSS J144157.26+094853.7?


- Present the appropriate number of significant figures for numerical measurements and uncertainties that match the precision of the measurements. (§2)
  - Example: (131.32134587, 1.01243229) would imply an accuracy of $10^{-8}$ degrees
  - More than 70% paper NED processed do not give uncertainty of measurements
- Include information for "data behind the plots" (§3.2)
- Provide a complete list of metadata (§4)

  E.g., creation, standards, quality of the data, …

# Open source

- Make source code available (§5.3)

  E.g., use Astrophysics Source Code Library (ASCL) to submit the software used in the research

- Provide full provenance of the data (§5.2)

  E.g., "2MASS (Skrutskie et al. 2006) as downloaded with TOPCAT (version 4.8–3, Taylor 2005) via VizieR (II/246, Cutri et al. 2003)"

- List the software and version used in the production of the article (§5.3)

  E.g., "Stanford Classifier v3.9.2, The Stanford Natural Language Processing Group, https://nlp.stanford.edu/software/classifier.shtml"

# Open methodology

- Describe the method used to estimate photometry (§2.3)

  E.g., PSF fitting, aperture photometry, or isophotal measurements? Size of the aperture and background annulus for aperture photometry? Corrections/assumptions made in the calculation?

- Describe the method used to obtain redshift (§2.5)

  E.g., the particular method (spectroscopic, photometric, etc.) and base assumptions used in the models (template fitting, machine learning, etc.)

- Provide uncertainty and confidence level when reporting a new measurement (§2)

# Open access

- Create the graphics with accessibility in mind (§3.2)

- Append small data sets as part of the publication and deposit large or complex data at an archive with open access (§4)
  - Do not publish data sets at URLs lacking long-term support.

- Include a Data Availability Statement (§4)

  E.g., "All data are incorporated into the article and its online supplementary material" or "Data are available in a repository and can be accessed via a DOI link".
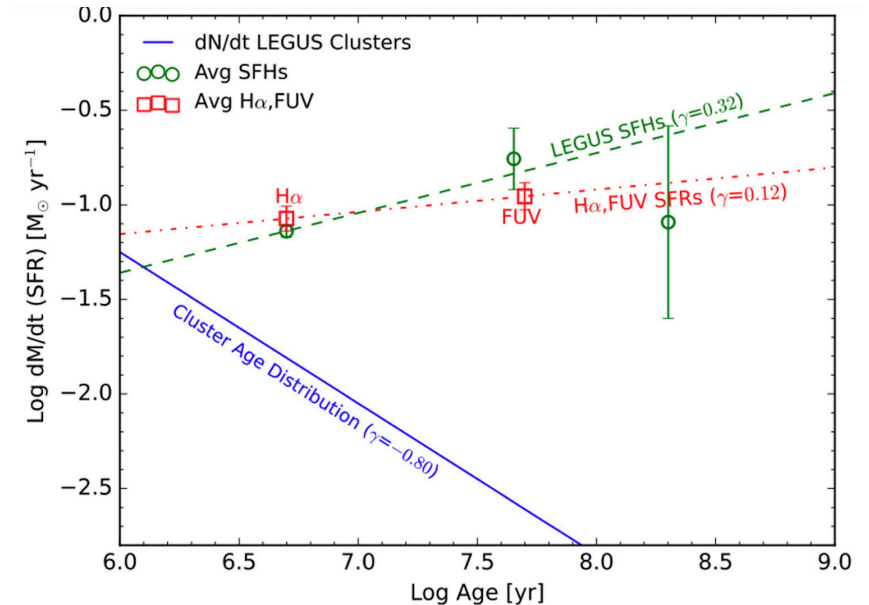


Cook et al. 2019MNRAS.484.4897C

# Open peer review

- "… intended to be used not only by authors during the preparation and submission stages of a publication, but also by referees, editors, and publishers during the refereeing and editing stages before final publication. " (§7)
  - Checklist of BP can be found at http://ned.ipac.caltech.edu/Documents/Guides/BestPractices

- Applying Best Practices requires increased attention to data and methodology during peer review.
  - "…improve the quality of the published research record, expedite the integration of data into the databases with more efficiency and accuracy, and result in long-term preservation and reuse of valuable data."
  - "… enable more scientific discoveries that would otherwise not be possible or practical"

# Integration into the Community

- Links to Best Practices article now in Instructions for Authors
  - AAS journals: AJ, ApJ, ApJL, ApJS, Planetary Science Journal
  - Oxford University Press (MNRAS)
  - Institute of Physics (PASP)

- Recently cited in two Nature Astronomy articles
  - "How to plan your astronomy research paper in ten steps" (Chamba, Knapen & Black 2022NatAs...6.1015C)
  - "How to write and develop your astronomy research paper" (Knapen, Chamba & Black 2022NatAs...6.1021K)

- Chen was interviewed for cross-disciplinary article
  - "How to make your scientific data accessible, discoverable and useful" (Perkel, 2023Natur.618.1098P)

# AAS workshop on Best Practices

- 2pm, Saturday, January 6,  2024, at New Orleans.

- Spread the words among your coworkers, friends, students, etc.

- Please register

    [https://aas.org/meetings/aas243/workshops](https://aas.org/meetings/aas243/workshops)