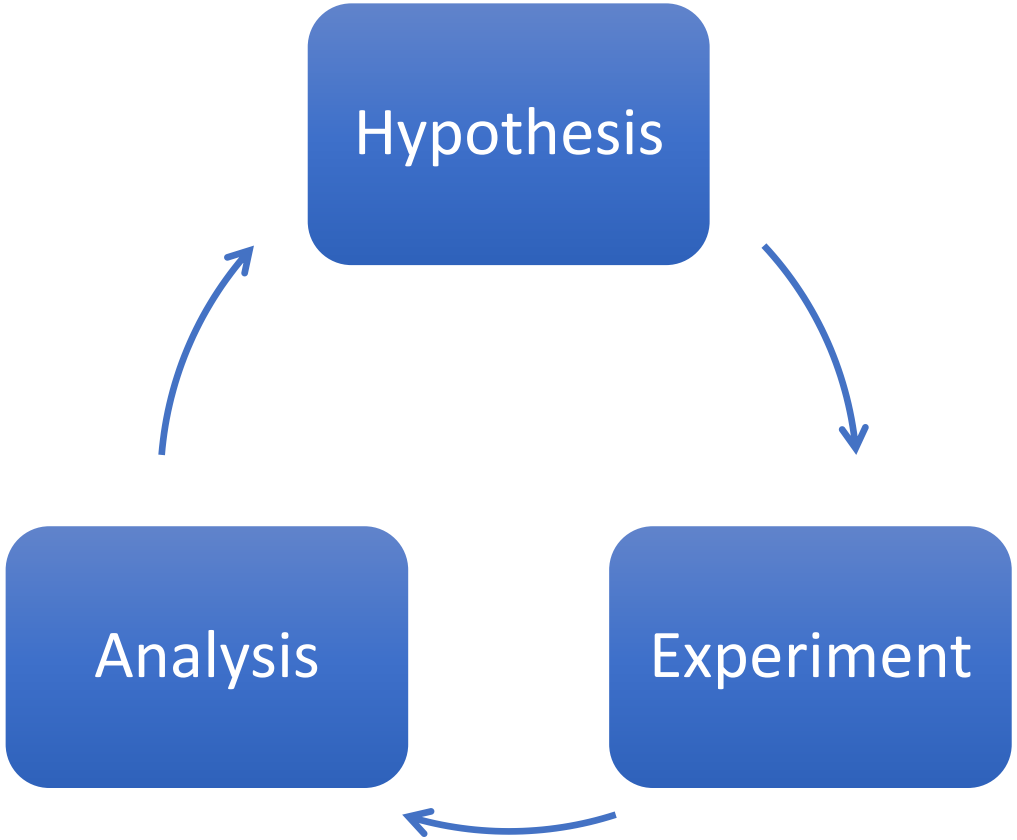
Making Atmospheric Observations FAIR: Challenges and Opportunities

Gao Chen, Crystal Gummo, and Michael Shook

Observations propel advances in research – Inception of Tropospheric Chemistry

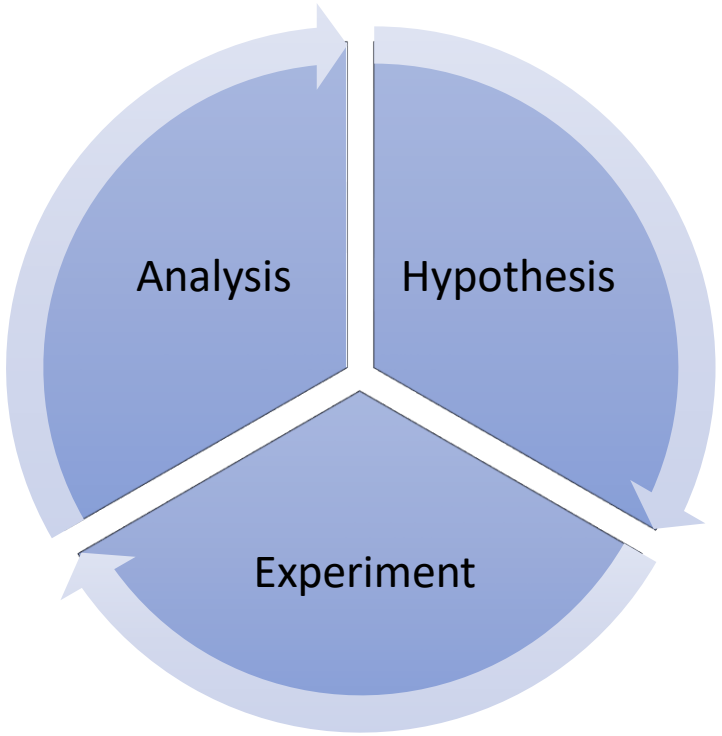


- Observations: Carbon Monoxide lower than expected
- Analysis: transformed some established stratospheric chemistry to troposphere
- Hypothesis: tropospheric radical chemistry: OH level sustained by OH-peroxy radical cycles (Levy, 1971)
- New experiment: NASA GTE program

H Levy II (1971), NORMAL ATMOSPHERE: LARGE RADICAL AND FORMALDEHYDE CONCENTRATIONS PREDICTED, VOL. 173, NO. 3992 , Science, DOI: 10.1126/science.173.3992.141

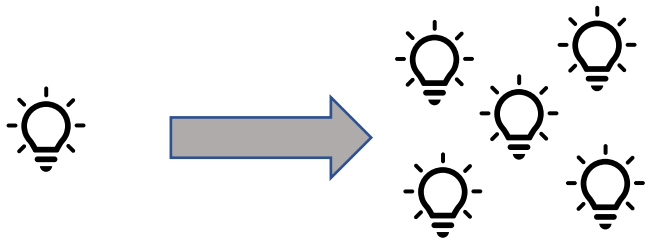
Observations propel advances in research – ARCTAS improved Satellite BrO Observations

- Background: BrO is an important tropospheric free radical
- Hypothesis: BrO column enhancement is linked to surface release and ozone depletion
- Experiment: ARCTAS Observations showed a disconnect
- Analysis: ARCTAS and satellite observations suggest larger spatial variability in stratospheric BrO than previously thought



Salawitch, R. J., et al. (2010), A new interpretation of total column BrO during Arctic spring, Geophys. Res. Lett., 37,L21805, doi:10.1029/2010GL043798

Open Science



- Aiming to accelerate advances in scientific research
- Fostering collaboration, reproducibility, and equity
- Promoting policy and practice and leveraging technology to lower barrier to access research results and process
- Ample examples demonstrated collaboration led to advances in scientific research
- Openness applies to both making and using open science

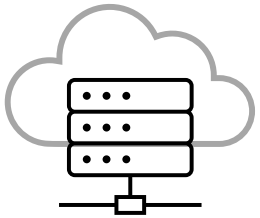
Implementing Open Science



Open Data, Open Results, Open Software, and Open Methodologies enable broader collaboration in all phases of research

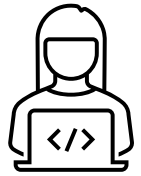
Open Data

Findable

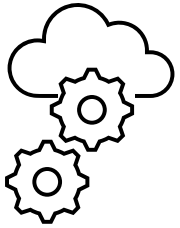


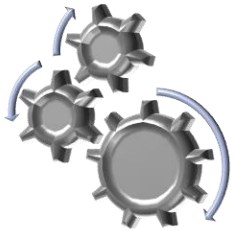






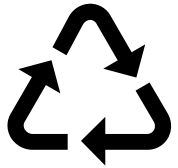
Accessible





Interoperable





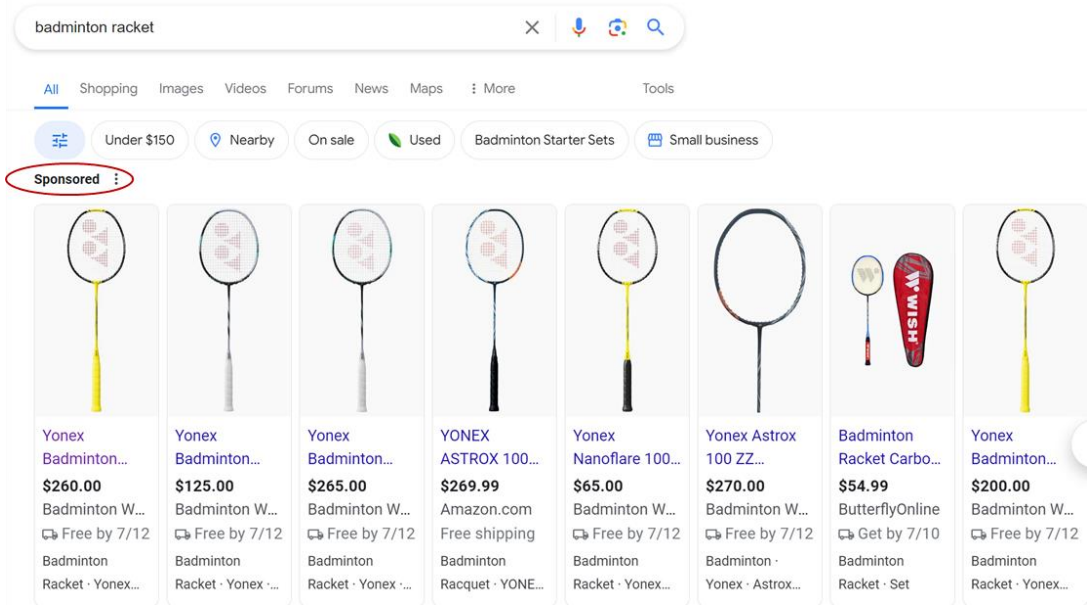
(Re)usable

FAIR Principles (https://www.go-fair.org/fair-principles/)

- FAIR = Findable, Accessible, Interoperable, and (Re)usable
- first introduced in 2016 by Wilkinson et al.
- Domain agnostic, but implementation requires domain knowledge
- Ongoing to effort on FAIR convergence to make the interpretation of the principles more consistent
- Practical implementation guidelines are under development, e.g., A PRACTICAL GUIDE FOR OPEN, FREE & FAIR NASA EARTH SCIENCE DATA PRODUCTS
- This effort requires close collaboration between data scientists and research scientists

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

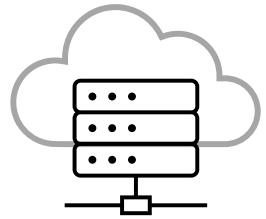
Being findable by both human and machines

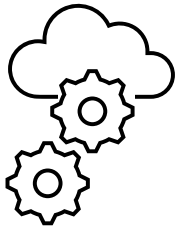


- Assigning data with globally unique, persistent, and resolvable identifiers
- “Rich metadata”
- Atmospheric observational data need to be more findable, especially with better machine-to-machine findability, partially reflecting the complexity of scientific data

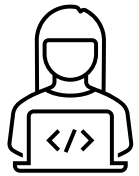
Being accessible by both human and machines

- Both data and metadata
- Free and open
- Standardized authentication process
- Globally implementable: e.g., http, ftp, APIs
- NASA data are open and freely accessible
- Advanced tools can provide visualization and analysis-ready products

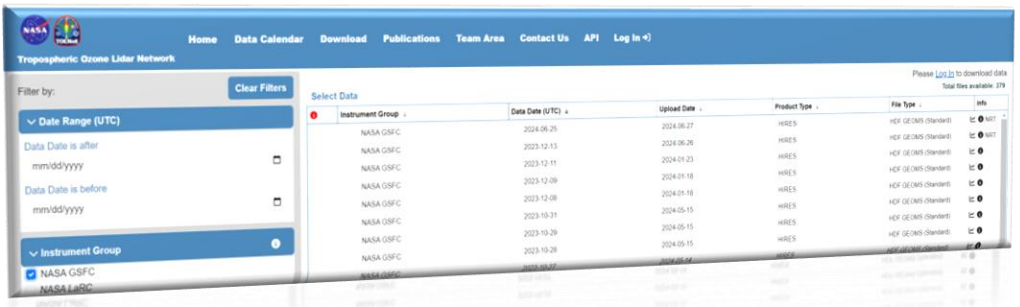




Machines can interact directly with the API



Humans can interact with the website or the API





Being interoperable: usable data across systems, tools, services, and applications

Semantic interoperability: data and metadata can be understood/actioned upon by different systems – ontology and vocabular

Syntactic interoperability: data and metadata structure and syntax, e.g., format, can be understood/actioned upon by different systems

Technical interoperability: exchange protocols compatible with different systems, e.g., HTTPS



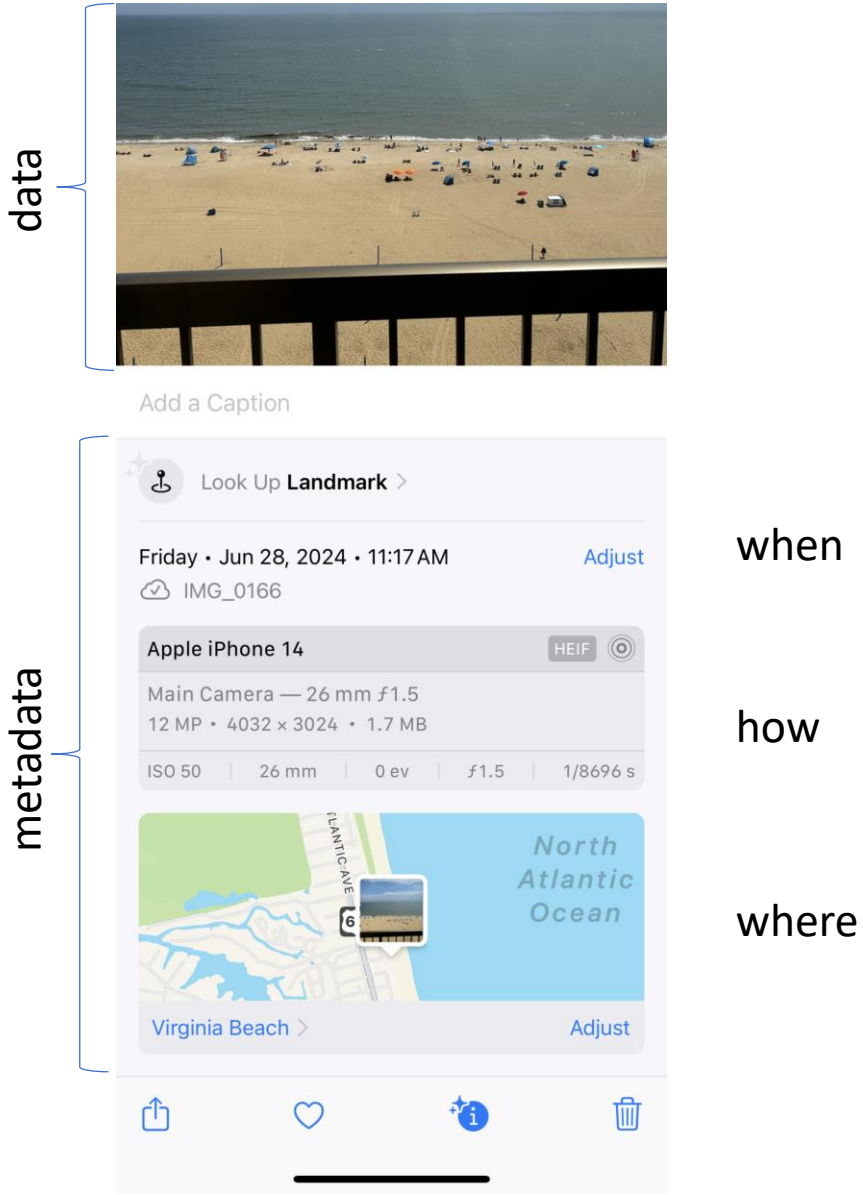
Being (re)Usable: readily interpretable for (re)use

Data and metadata are “richly” described and:

- Following domain-relevant community standards, e.g., metadata requirements and format standards
  - Domain: a research area or a subject area
  - Community standard example: ICARTT format
- Describing provenance – origin and history, including:
  - Collection, e.g., sampling and instrument detection
  - Processing and revision
  - POCs
- Providing usage license

Metadata: data about data

- iPhone photo can be found in different ways, e.g., by datetime, location, and content, which is supported by “persistent” identifiers
- iPhone photo metadata is interoperable, i.e., it can be used by other applications, like Google Photos



Metadata and Data for Atmospheric Composition Measurements

ICARTT format is a community standard developed by NASA

ICARTT format contains both data and metadata file header

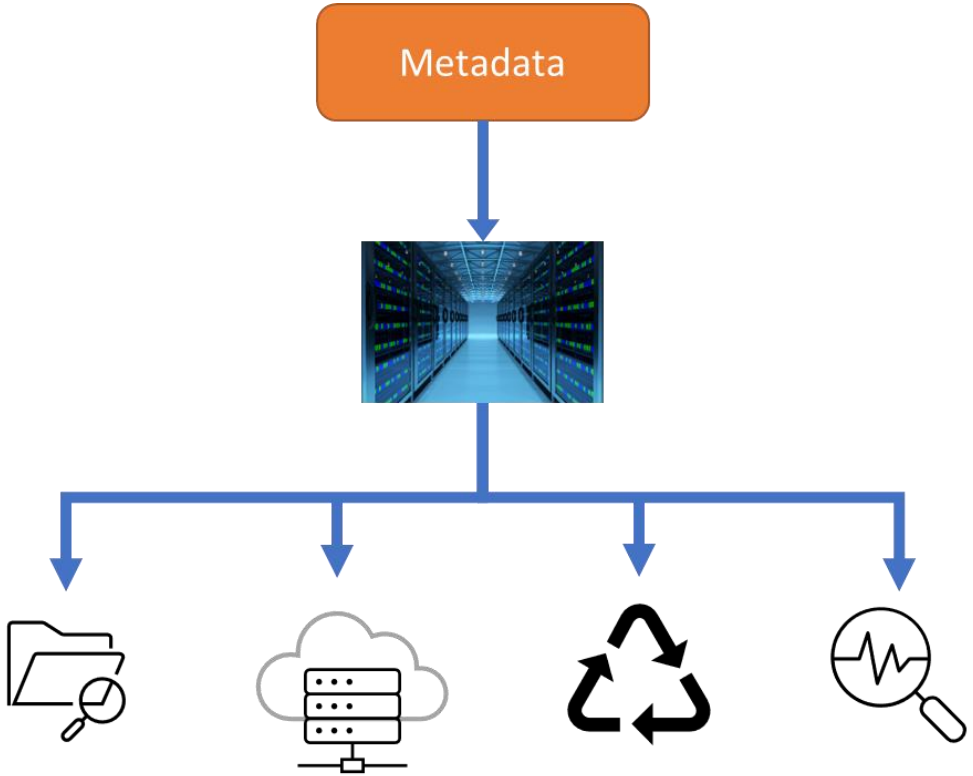
ICARTT Metadata includes

- Structural metadata, e.g., number of header lines, number of variables, data flags
- Descriptive metadata: data source, e.g., instrument, platform
- Definitional metadata: variable name, units, standard name, description
- Provenance metadata: data revision, point of contact
- Data quality metadata: uncertainty and limit of detection

"Persistent" Identifier: Atmospheric Composition Variable Standard Name

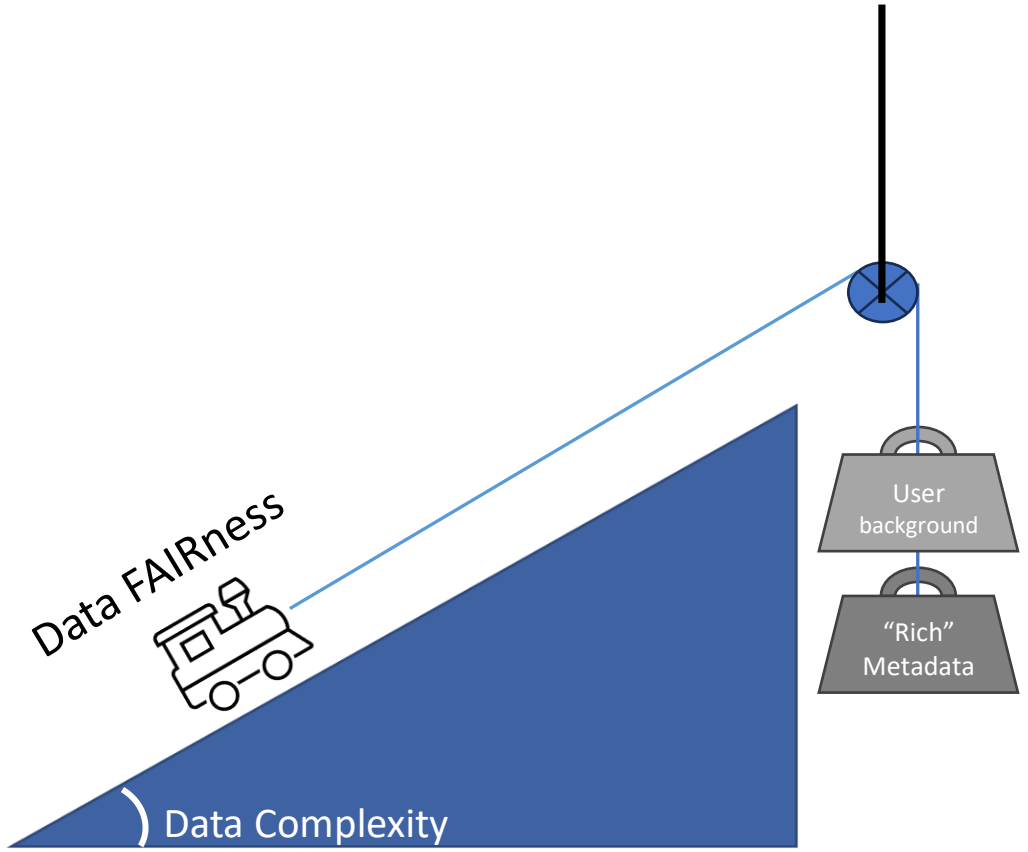
“Rich” Metadata

- Sufficient Metadata through data system is needed to support findability and (re)usability
- “Rich” is purposely vague to make FAIR principles subject/domain agnostic
- Findable metadata need:
  - Faceted search example: Looking for a t-shirt, a polo shirt or, a blue polo shirt
  - A polo shirt would not need to have the same description as a dress shirt
- Usable metadata need:
  - To make food would require more description than to order food from a menu
- For support Earth science research, the data system and metadata standards need to support diverse metadata



FAIRness is not absolute

For a given application:



Challenges from research perspective

Metadata standards and best practices

- Most were developed by data scientists without much subject area knowledge, which can lead to serious applicability issues. The ones developed by research scientists tend to be narrow and may not be machine friendly
  - Bounding box vs. flight tracks for spatially representing aircraft sampling
  - ICARTT format
- Lack of domain-relevant persistent identifiers
- Standard development and maintenance are resource intensive

Vocabulary and ontology

- Ontology implementation can be subjective when dealing with complex variables
- Progress requires close collaboration between the data scientists and research scientists

Machine friendly data formats tend to be human-hostile, especially to those with limit coding experience

- HDF and netCDF files require special tools or coding skills to read
- Keep in mind not only the type of data you are collecting, but how the data might be used in the future. And the future means machine readable, which means self describing data schemas.

How to make your observation data FAIR (I)

- Study and understand the intent of the current data reporting standards and best practices
- Review community standards to find the one best fit the data and data distribution system
- Use persistent identifiers, when possible, e.g., data DOI, ORCID
- Raise questions when run into issues with standards and best practices
- Provide feedback and contribute to the improvement

How to make your observation data FAIR (II)

- Specific actionable items for reporting data:
  - ✓ Time stamp with definition
  - ✓ Persistent identifiers: data DOI, ORCID, and Variable Standard Name, e.g., ACVSNC
  - ✓ Variable name not starting with a number and not containing “-” or special letters
  - ✓ Accurate and understandable variable description
  - ✓ Units and standardized notation
  - ✓ Uncertainty information
  - ✓ Revision



How to use FAIR data

- Learn background information before search or access the data
- Use literature terms or textbook terms as well as persistent identifiers (e.g., data DOI) to search data.
- Do review the metadata included in the data file to understand the data, especially the definition of the variables and limitations of the data
- Even though not required, it has been often proven that collaboration with the instrument scientists brings in more insights and new perspectives
- Properly cite the data, e.g., data DOI and citation statement, and acknowledge collaborators

Opportunities

- Standard and vocabulary development – need expertise not only in data science
  - ✓Franziska Stürzl of MeteoSwiss, math major, invited to work with WMO Expert Team on Metadata
- Data Management – working with instrument scientists to make data FAIR
  - ✓Michael Shook of NASA, Meteorology major, data manager for multiple NASA field studies
- Data Ingest – Making data findable and accessible, assigning persistent identifier, e.g., data DOI
  - ✓Megan Buzanowicz of NASA, MS in Atmospheric Sciences, team lead at ASDC