Follow along at:
https://bit.ly/nasadoi

Registering a data DOI at NASA—a standards story

10 May 2023
IVOA Interop Meeting

Mark A. Parsons
https://orcid.org/0000-0002-7723-0950
mark.parsons@uah.edu
@chutneyboy
University of Alabama in Huntsville

IMPACT
Interagency Implementation
and Advanced Concepts Team

NASA SMD needs standards and guidelines because

- Open data is not enough
- Open science requires a new culture **and** a supporting infrastructure

> **"The basic irony of standards is the simple fact that there is no standard way to create a standard, nor is there even a standard definition of 'standard'."**
>
> - Andy Russel and Lee Vinsel, NYT, 2019-02-16

A Policy-Driven Vision

NASA SMD Science Information Policy:

III.C.i. SMD-funded data should follow the FAIR Guiding Principles for scientific data management and stewardship.

III.D.ii. When released, SMD-funded software should follow best practices in the relevant open source and research communities.

A Pragmatic Mission (to start)

Help make NASA's science discovery systems work better.

Help SMD repositories meet the new demands of open science.

Who

- Alan Smale (Astrophysics)
- David Ciardi (Astrophysics)
- Bruce Berriman (Astrophysics)
- Robert Candey (Heliophysics)
- Brian Thomas (Heliophysics)
- Dan Berrios (Biological & Physical Sciences)
- Thomas Morgan (Planetary Science)
- Steve Hughes (Planetary Science)
- Bob Downes (Earth Science)
- Sara Lubkin (Earth Science)
- Steve Crawford (HQ)
- Mark Parsons (IMPACT)

All y'all!

Wisdom of the Crowd

What — Objectives

Establish an SMD-wide 'standards' *guidelines* process to help implement the NASA Information Policy:

- Review and agree on which 'standards' are needed to achieve the policy objectives, including the FAIR Principles etc.
- Foster broad collaboration around and common usage of conventions, agreements, leading practices, specifications, as well as formal standards to create a culture of interoperability.
- Identify mutually satisfactory ways to align all divisional standards goals with the broader SMD goals.
- Determine which standards shall be adopted and how (e.g., which profiles, vocabularies, versions, related protocols, data formats, etc).
- Identify where standards are missing and how that gap should be addressed
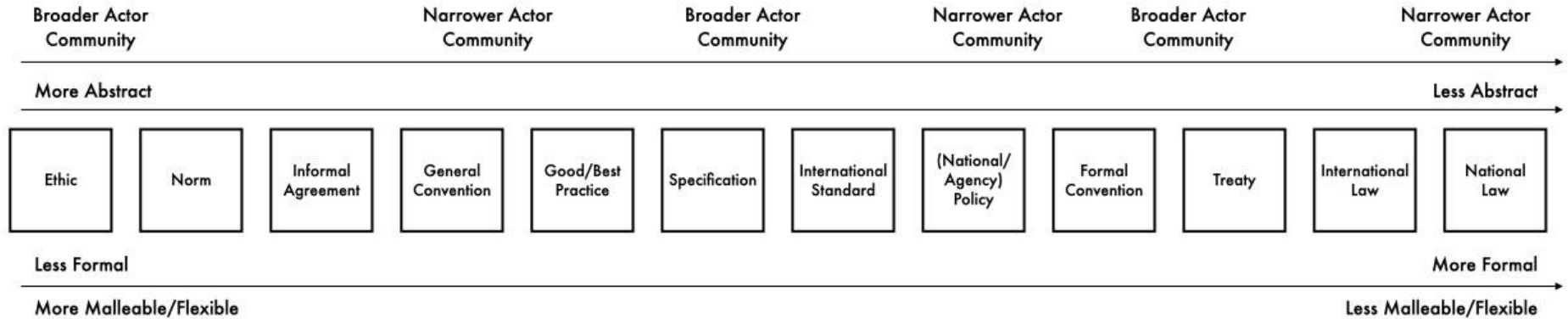
How — Principles

Very Important Slide!
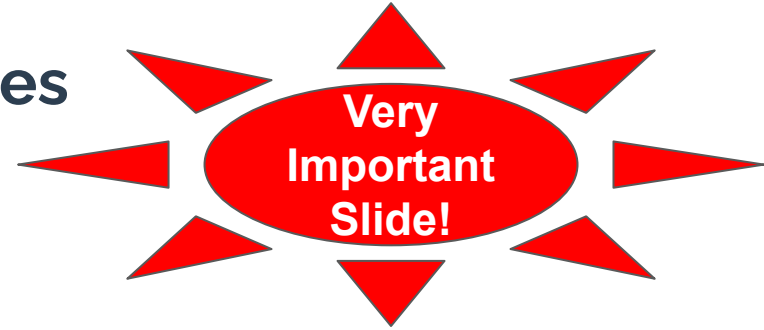
1. Use or adopt existing standards where possible.
2. There is no one (format) standard to rule them all. Disciplinary standards should be respected, but there will be some level of required commonality or crosswalking.
3. Any standard must solve a problem and be actively adopted.
4. Bottom up standards are preferred to top down mandates where possible.
5. The details of exactly how to actually implement standards are as important as the standard itself.
6. Reduce total effort. Make it easy for data providers.
7. The concerns of data providers must be addressed.
8. Emphasize adding value over meeting requirements. Carrots are better than sticks.
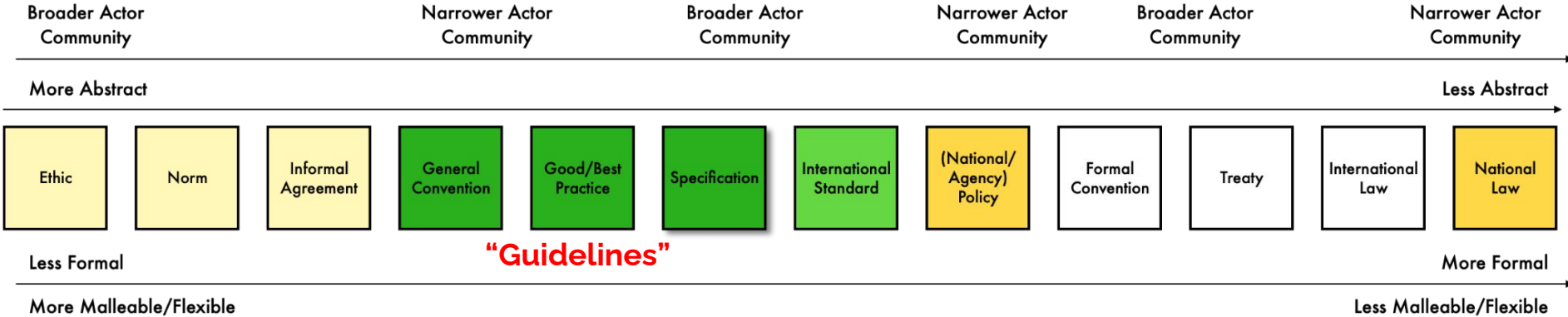
Explicit "community standards" referenced

- FAIR (Should NASA be a FAIR Implementation Network?)
- Open Source Initiative (should address most of the licences and IP issues)
- *Science Journal* open source policy
- *Nature* open source policy
- AGU Data policy
- American Astronomical Society
  - Data Guidelines
  - AAS Software policy

Scope: Where on the spectrum below?

| Broader Actor Community | | Narrower Actor Community | | Broader Actor Community | | Narrower Actor Community | Broader Actor Community | | Narrower Actor Community | |

More Abstract → Less Abstract

| Ethic | Norm | Informal Agreement | General Convention | Good/Best Practice | Specification | International Standard | (National/ Agency) Policy | Formal Convention | Treaty | International Law | National Law |

Less Formal → More Formal

More Malleable/Flexible → Less Malleable/Flexible

Figure based on the work of Peter Pulsifer, Carleton University, Ottawa, Canada

Scope: Where on the spectrum below?

| Broader Actor Community | | Narrower Actor Community | | Broader Actor Community | | Narrower Actor Community | | Broader Actor Community | | Narrower Actor Community |

More Abstract → Less Abstract

| Ethic | Norm | Informal Agreement | General Convention | Good/Best Practice | Specification | International Standard | (National/Agency) Policy | Formal Convention | Treaty | International Law | National Law |

"Guidelines"

Less Formal → More Formal

More Malleable/Flexible → Less Malleable/Flexible

- **Active Involvement**
- **Involvement/Awareness**
- **Keen Awareness. May drive new requirements.**
- **General Awareness. Represent some of the cultural considerations necessary to standardization.**

Reviewed Existing Processes, including

- ESDIS Standards Coordination Office
- International Virtual Observatory Alliance
- International Planetary Data Alliance
- Space Physics Archive Search and Extract Consortium Data Model (SPASE)
- Internet Engineering Task Force
- Research Data Alliance

Seeking a balance of control and flexibility
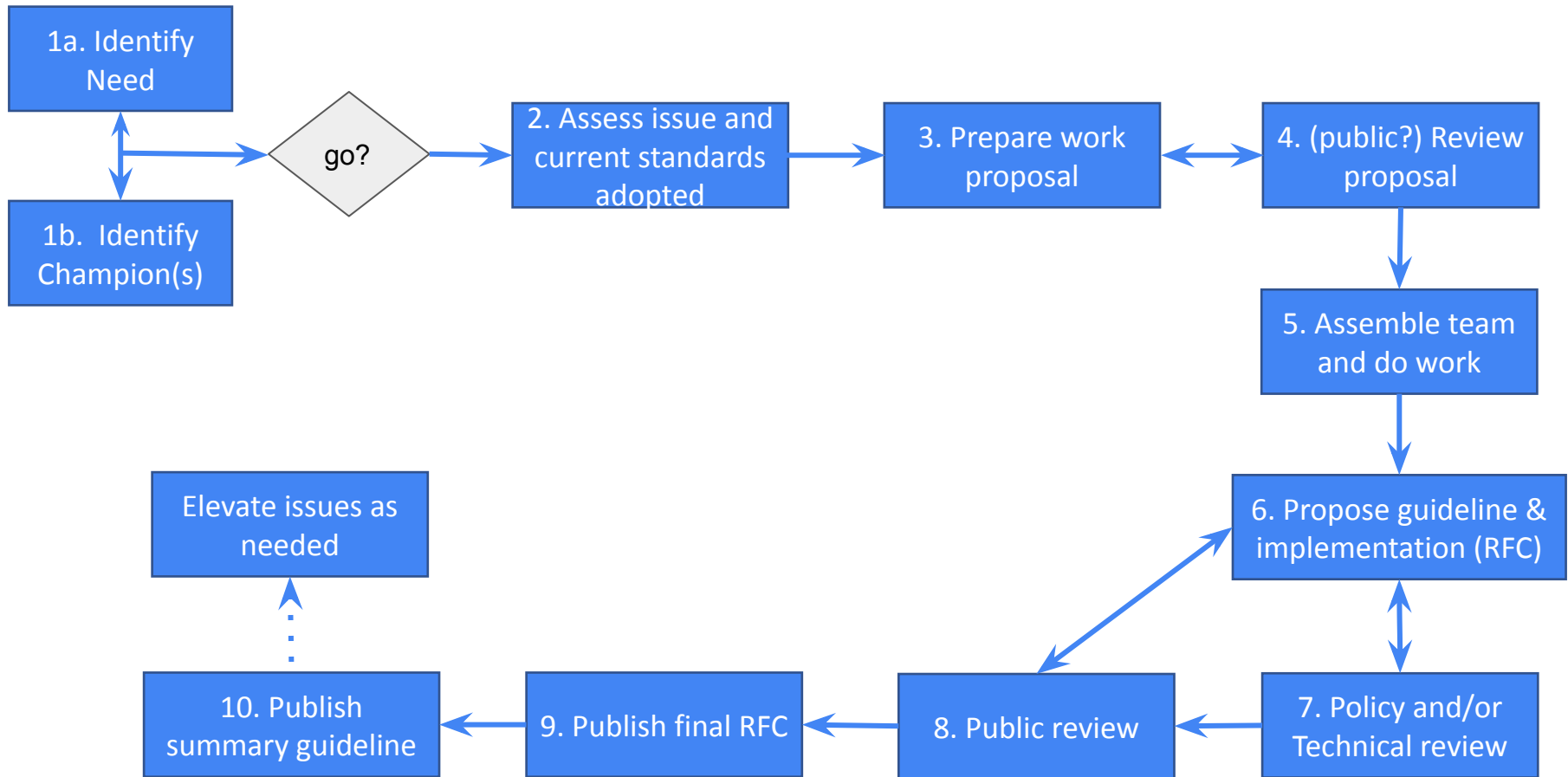
DOIs for data citation as a test case

SPD-41a requires NASA data to be citable "using a persistent identifier". But which identifier and how? So we

- Assembled expert team
- Developed <u>work proposal</u> including existing community guidance
- Surveyed existing practice – 28 of 32 archives responded. Most using DOIs registered through DataCite (in addition to other identifiers). The issue often seems to be mapping a set of local IDs to something considered citable, i.e. "DOI-worthy". Most have published guidance for researchers, archives or both.
- Developed detailed request for comment
- WG and Policy Officer reviewed
- Community reviewed and commented.
- <u>Full RFC</u> (DOI pending 😊)
- <u>Short version</u>

Who

- Dan Berrios, BPS
- Robert M. Candey, HPD
- Mitch Gordon, PSD
- Nathan James, ESD
- Steve Joy, PSD
- Mark Parsons, NASA IMPACT, University of Alabama in Huntsville
- Josh Peek, ASD
- Anne C. Raugh, University of Maryland, College Park (PSD)
- Aaron Roberts, HPD
- Gerald Steeman, STI

Guideline Background and Context

- Policy requires SMD-funded data to be citable with a PID.
- DOIs are commonly used for this in NASA. DataCite is currently the only DOI Registration Authority tailored to data and has three membership types:
  - Member only (don't create DOIs),
  - Direct member,
  - Consortium member.
- Some archives are direct members. STI has a consortium membership for NASA and is developing services.
- Three scenarios for registering DOIs to enable data citation in the literature:
  - Planned
  - Provider request
  - User request

Guideline description (section H)

- Data intended or used for citation in the scientific literature should have a DOI registered through DataCite.
- DOI requests for SMD-funded data should be processed through the responsible repository.
- Metadata requirements should be met by the repository in collaboration with providers.
- The repository should be responsible for maintaining the metadata for the digital object as well as the landing page and resolvability of the DOI.
- Repositories establish their own guidelines on what is required to register a DOI.
- Three options for archives to register DOIs:
  - Request through STI (few)
  - Work through the NASA Consortium Membership (majority)
  - Become a Direct Member (few activists).
- Repositories and STI coordinate to implement an agency registry of DOIs.

Some lessons learned

- PIDs (e.g. DOIs) are essential but the devil is in the implementation details.
- Scoping is essential.
- Context is critical.
- Champions and facilitators are critical.
- Revealed underlying friction of (any) data policy – user needs vs. NASA needs with repositories in the middle.

Marcel DuChamp's *Bicycle Wheel*
photo © nuzz—www.flickr.com/
photos/nuzz/

Friction is inevitable and necessary

"A wheel turns because of its encounter with the surface of the road; spinning in the air it goes nowhere."
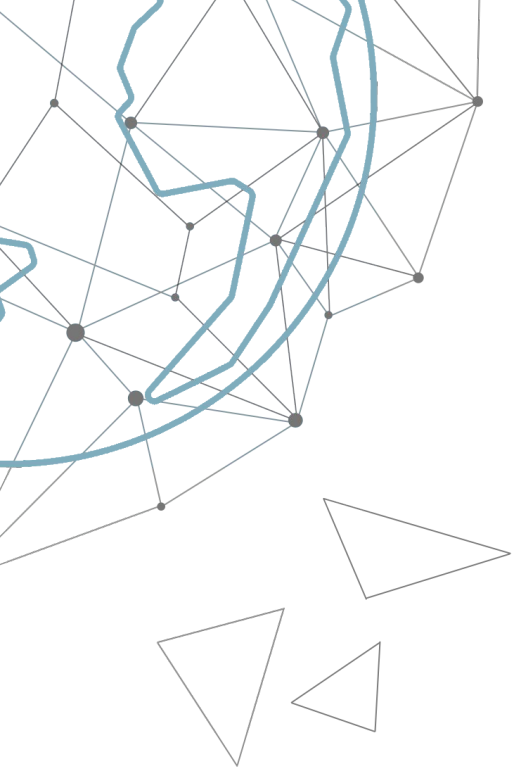cover notes for
Friction—An Ethnography of Global Connection by Anna Lowenhaupt Tsing[3]

"Standardization is dynamic, not static; it means not to stand still, but to move forward together."

1920's motto for the Engineering Standards Committee (precursor to ANSI)
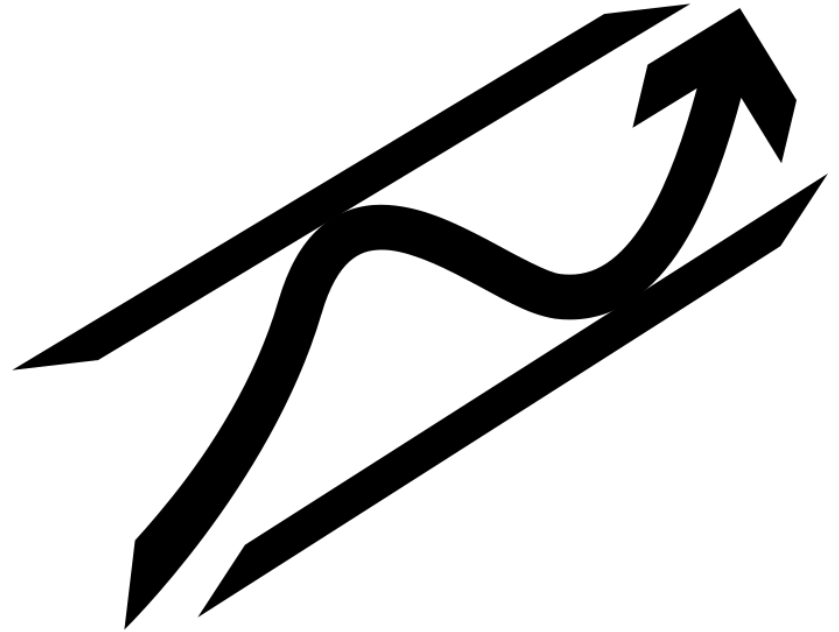
Created by Manuel Waelder, Noun Project

Thank You

Contact me at mark.parsons@uah.com @chutneyboy