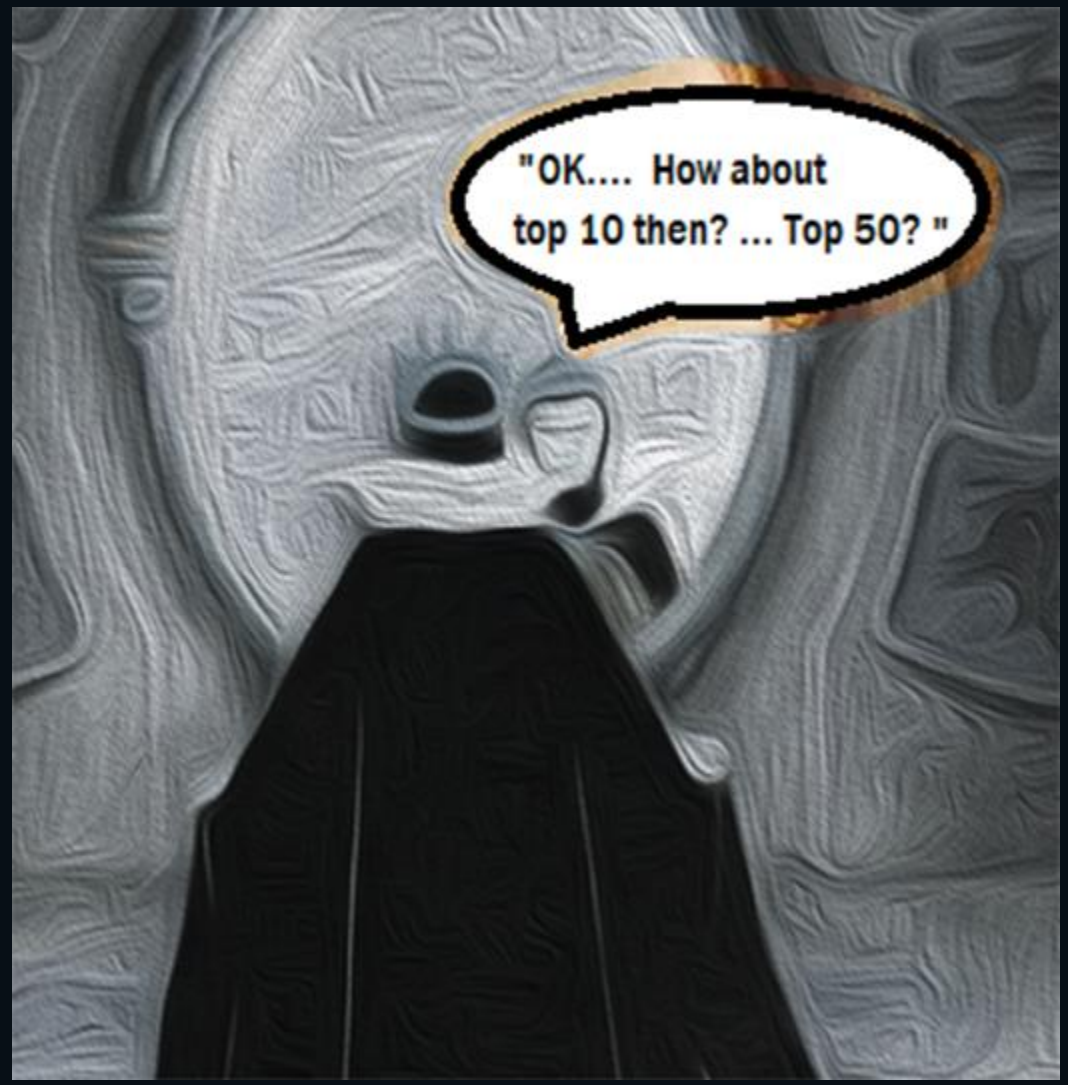
# Governing Data FAIR Compliance

Daniel C. Berrios[1], Gregory D. Eley[2,3], Kathryn C. Richards[2,3], Ruth A. Reitzel[3], John H. Dunn[4], Olga V. Kieschnick[3], Truong Le[3], Sandeep D. Shetye[1], Jessica A. Keune[3]

[1]NASA Ames Research Center; [2]KBR Wyle Services, LLC; [3]NASA Johnson Space Center; [4]Universities Space Research Association

# Agenda

- FAIR Compliance?
- FAIR Governance: Challenges
  - Numerous Tools and technologies
  - Numerous Interested Organizations
- Governance Strategies
  - Centralized
  - Distributed
  - Multi-level/Federated
- Summary



# FAIR Compliance?

- FAIR: *Data* Findability, Accessiblity, Interoperability, Reusability*
- Measurement(s) of the extent to which specific data and systems conform to the required designs stated in the principle
- FAIR compliance is a product of both system and data engineering
  - For example, system designs specify communication protocols referenced in FAIR Accessibility principle A1 ((Meta)data are retrievable by their identifier using a standardized communications protocol)
  - Data designs (formats, standards, terminologies, etc.) constrain compliance with FAIR Interoperability principle I1 (Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation)
- Thus, FAIR compliance necessarily involves many groups, from investigators and data curators, to system engineers and program/project managers

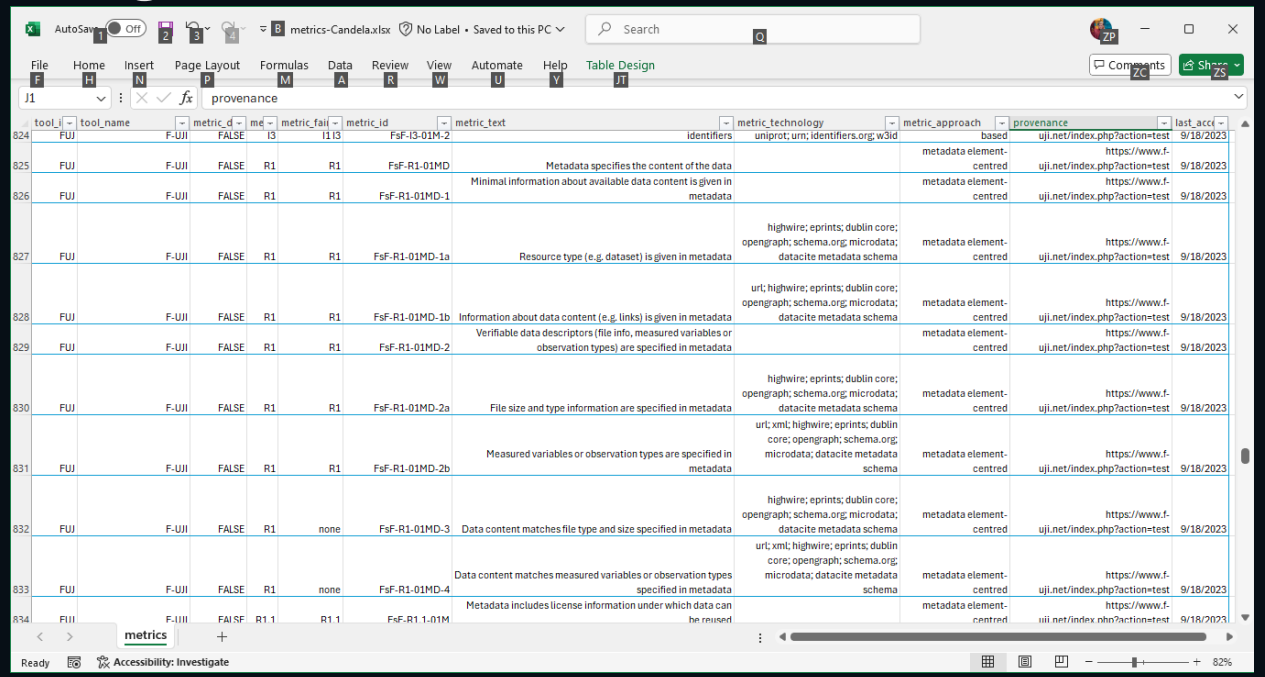
*Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles …. (2016). https://doi.org/10.1038/sdata.2016.18

# FAIR Governance Challenges

- Many interested parties at NASA
  - Boards/Offices: NASA Data Governance Board (see NPD 2800.1), OCIO, CDO, OCSDO, HRP-CSO, HRP-SIO, DMIO, Sci Divisions Data Offices, Org Data Management Boards)
  - Programs (Human Research Program, Space Biology & other SMD Programs)
  - Science/Technology Divisions/Directorates (APD, BPSD, ESD, HPD, PSD, HHPD)
  - Scientific investigator community
- Incomplete/immature set of FAIR Compliance metrics standards and best practices
  - What are best practices for compiling metrics across data sets? Across data collections? Across systems?
  - How should FAIR metrics be weighted according to domain or organization priorities?

# FAIR Governance Challenges (Cont'd)

- Incomplete/immature set of FAIR Compliance tools/technologies for system and data engineers to leverage
- 1180 Metrics in 20 different, publicly available FAIR compliance frameworks
- Not yet widely tested




Candela L, Mangione D, Pavone G 2024
The FAIR Assessment Conundrum:
Reflections on Tools and Metrics.

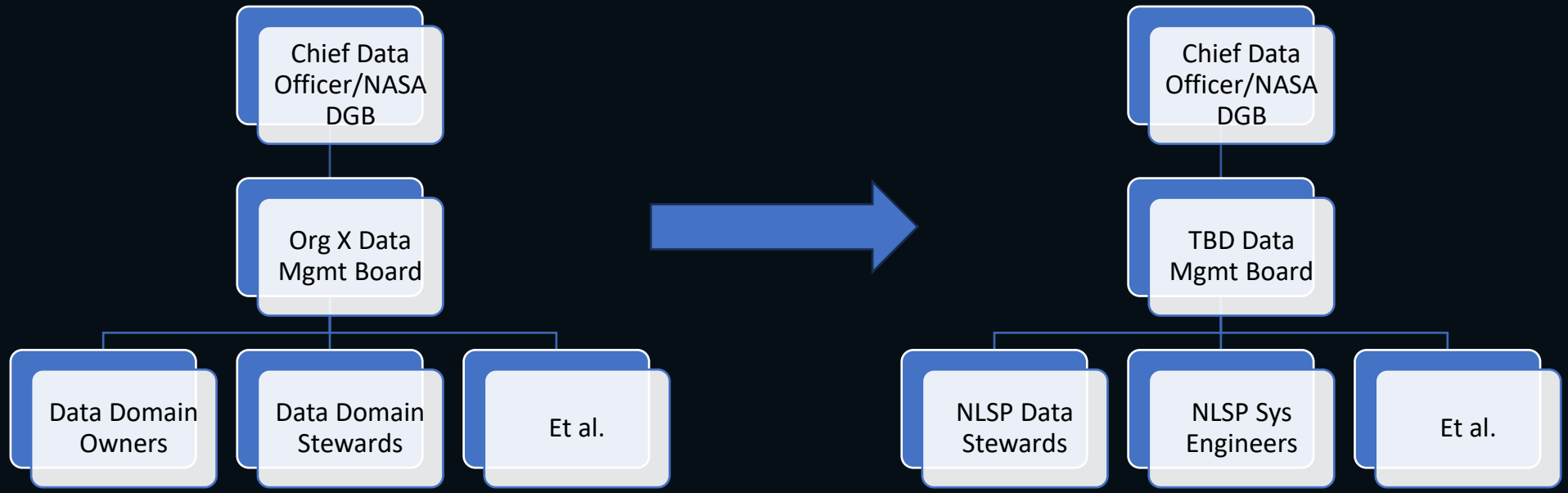
DOI: https://doi.org/10.5334/dsj-2024-033

# FAIR Governance Strategies

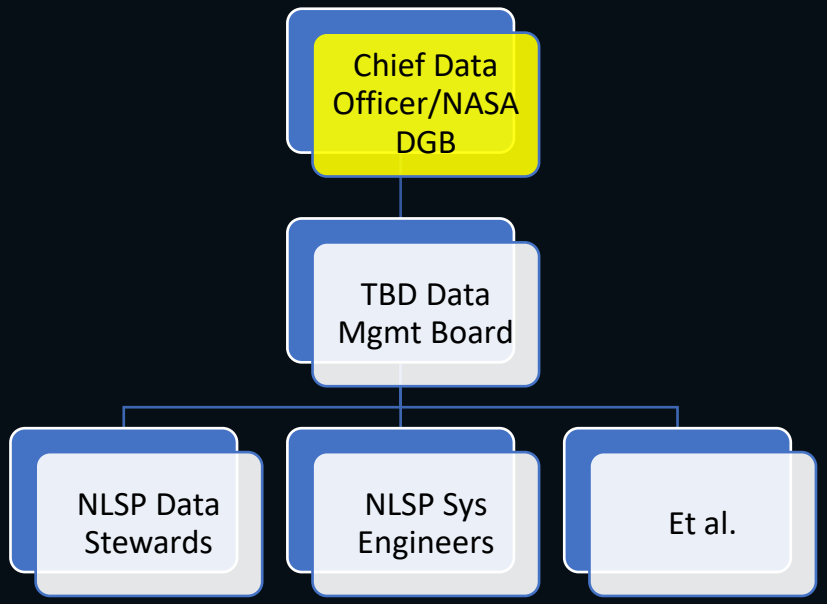
- Sep 2024 NASA Science Data Repositories Workshop
  - Discussion of 3 Possible Strategies: Centralized, Distributed and Multi-level
  - Centralized (4 PRO, 6 CON)
    - "Not the best [strategy]…those not working directly with the project datamay not have a good understanding of nuances that lead to FAIR Compliance…"
  - Distributed (4 PRO, 2 CON)
    - "Good idea but will end up with inconsistent ways even with in a single division"
  - Multi-level (9 PRO, 0 CON)
    - "A layered approach is required… having domain experts 'layer" pertinent requirements on top of basic requirements, but then have external review to safety check"
  - See https://doi.org/10.5281/zenodo.14285970 for entire set of comments

# NPD 2800.1 (Dec 2024): NASA Data Governance, Roles, and Responsibilities

- “The NASA CDO shall review the development of data infrastructure to reduce barriers that inhibit data findability, access, interoperability, reuse, understandability, security, and trust (FAIRUST).”
- Federated Data Governance Framework

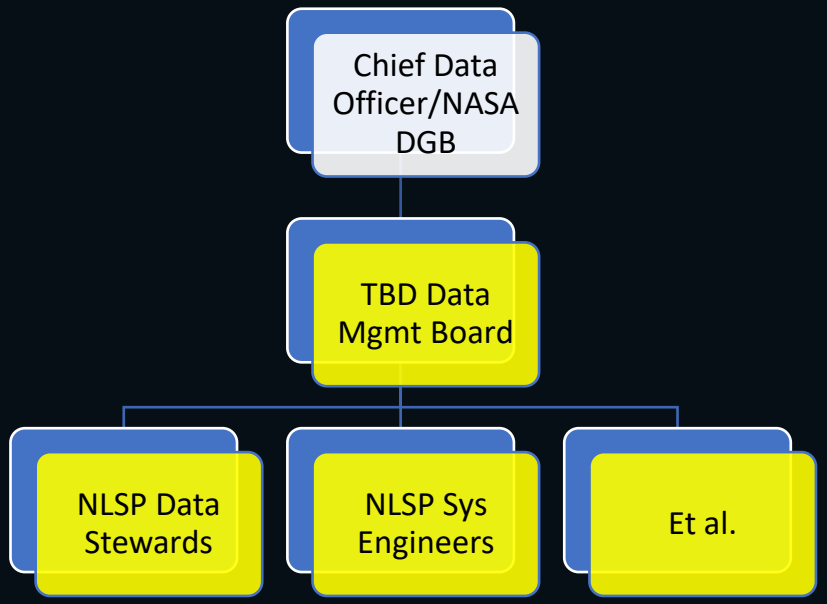


# Federated FAIR Compliance Governance




- As discussed at least partially during the 2024 SMD Data Repositories Workshop, this “multi-layered” strategy:
  - Supports inclusion of domain expertise in processes and designs
  - Allows organizations to review and synergize FAIR efforts by their own projects and systems
  - Provides an enterprise-wide point to gather (or inject) knowledge regarding FAIR Compliance efforts, or coordinate inter-agency efforts
  - Supports persistence of compliance knowledge at higher levels, yet permits evolution over time of compliance activities at lower levels
- Roles of the CDO and EDWG
  - Provide guidance on FAIR compliance goals for organizations (targets)
  - Guide organization towards standards or best practices for measuring system-wide FAIR compliance
  - Monitor / Review enterprise FAIR efforts and overall compliance

# FAIR Compliance Governance: Nuts and Bolts




- Organizational Data Management Boards
  - How much are each of the principles already implemented? How can/should the organizations' projects assess this?
  - Which of the FAIR principles should be highest priority for (further) system and data designs? Which are least costly and quickest to (further) implement?
- Data Stewards and Owners, Sys Engineers
  - Which system/data designs for a given FAIR principle/metric are most effective, sustainable (adaptable at low cost)?
  - Which designs are other systems using (in order to reduce burden on data generators and consumers of adopting and adapting to these designs)

# Summary

- Discussions by NASA science data repository leaders of best strategy for governing FAIR compliance mirror those issued recently by the NASA OCIO:
  - a layered, federated approach in which
  - Multiple data management teams (DMBs) work prioritizing and synergizing compliance efforts within their organizations and
  - Report up to Enterprise-level DMB and/or Enterprise Data Working Group
- While the proposed framework provides great flexibility and should be sustainable, the risk of re-work of organizational DMB wrt FAIR compliance would  be minimized if guidance or at least approval of FAIR compliance targets tools, and processes were made available soon through the framework

# Acknowledgements

Thank you

- Rebecca Ringuette, Lorella Angelini
- Kaylin Bugbee
- Mark Parsons
- David Salvignini
- Carlos De Los Santos
- Byron Smith
- Sam Gebre and Sylvain Costes
- Kevin Murphy, Steve Crawford, Brian Thomas
- The NLSP Development Team: John Dunn, Aaron Lott, Kimberly Heintschel, Patrick Moore, Stuart Williscroft, Ed Jobe, Qi Lin, Sara Glickman, Brian Yu, Sebastian Gutierrez Nolasco

## In Memoriam John H. Dunn

Loved by all of us who worked with him over the last 8+ years, John was the manager and senior engineer (USRA) who led development of the Insight Data Science Platform underlying the NASA Life Science Portal. John's dedication to his team and to his NASA projects resulted in good friendships and great results. John received many awards during his time with USRA and NASA, including the SMA Crystal Ball Team award, the NASA Group Achievement Award, and the USRA President's Innovation Award.