# Changing the Culture of Data Management and Sharing — A Workshop

## Workshop Highlights from the Co-chairs

Maryann Martone, University of California, San Diego
Richard Nakamura, *Retired*, NIH Center for Scientific Review

A Virtual Workshop
National Academies of Sciences, Engineering, and Medicine
April 28-29, 2021

# Changing the Culture of Data Management and Sharing — A Workshop

## Workshop Highlights from the Co-chairs

Maryann Martone, University of California, San Diego
Richard Nakamura, *Retired*, NIH Center for Scientific Review

Summary Remarks from Day 1: April 28, 2021

# Memorable Quotes Heard from Speakers Today

- “Prioritize adoption over perfection”
- “Engage now; it’s more productive than waiting”
- “Data should be elevated and made “Science Ready””
- “Datafication of biomedical science”
- “Most important issue of data sharing is respect but needs to be earned with an equal amount of generosity”
- Move away from “Data banking” and “Eat your brussel sprouts”
- “Future you will be a major consumer of data sharing”
- “The goal of data management isn’t to archive perfect data”
- “Barriers do not need to be roadblocks”
- “Nudge points”
- 1% of papers have associated data
- “Infiltrate the committees;  it comes down to who is in the room”
- “But what about…” (more could always be done)

As we practice, we will learn, and the infrastructure will get better

# Select Key Points (Maryann Martone)

- A lot of focus on “long tail data” not just large prospectively shared datasets
- Emphasis on impact and not “box ticking”
- Glass half full:  Parts of the infrastructure (e.g., data citation) are still in a state of becoming
    - We will get some things wrong and some things right, but the basics are there
    - Understand better the “edge” cases that will be encountered
- Raw statistics shared today may be slightly discouraging, testify to the magnitude of the challenge before us
- Costs are a major concern to the research community as well as promotion and tenure
- Libraries/librarians will play a big role in implementation but  need to engage with the wider infrastructure
    - 10% of research data are in institutional repositories
    - 90% of data will be published outside of an institution
- Prepare to share FAIR: Data management and sharing requires partnership and collaboration among many entities;  we all need to do our part
- Act local;  think global:  management in the lab and small scale sharing within and across labs as a bridge to open data
- Human infrastructure: Workforce issues:  professionalization of data stewardship, building “guilds” or “pools”;  oversight

# Select Key Points (Richard Nakamura)

- Important to think about data management and sharing early on and throughout the whole research process.  Keep data and metadata (scripts)together in study workflow.
- Entire community of stakeholders are being invited in to participate and contribute to this effort
- Key Issues to resolve:
  - Funding
  - Resources
  - Incentives
  - Tools
  - Training modules
  - Discovery Science
  - Citations standards
  - Tests for Data Reusability
  - Repositories and Metadata Best Practices
  - Visibility of PI sharing practices

# Changing the Culture of Data Management and Sharing — A Workshop

## Workshop Highlights from the Co-chairs

Maryann Martone, University of California, San Diego
Richard Nakamura, *Retired*, NIH Center for Scientific Review

Opening Comments from Day 2: April 29, 2021

# Where We Started on Day 1 (April 28, 2021)

- Many benefits to data sharing: impact during pandemic was clear
- Will involve a shift in scientific culture to “elevate the data,” including data from “failed experiments”
- NIH policy has flexibility built in
- Responsible data sharing is a partnership across the research ecosystem – community is invited to participate in this effort

# Introduction: Perspectives on Sharing Across Different Data Types

- Panel Discussion
  - What will successful sharing look like?
    - Proactive approach included early and throughout the process
    - Emphasis on impact not box-ticking: data reuse, innovative, action-oriented
    - Careers will be started with other people's data
    - Culture and incentive structures will be different
    - Partnership, collaboration

# Session I: Strategies for Managing and Sharing Data

- Big data- don't neglect "variety"
- Build multi-species and -center data repositories
- Involve the community early: How would you make your data FAIR?
- Meet researchers where they live
- Innovative architectures that include "data visiting" not just downloading
- Innovation in tools: examine old data for new insights
- Flexible means for sharing negative data
- Research repositories
    - Sustainability relies on meaningful metrics
    - Researcher-driven repositories offer flexibility
    - Institutional and generalist repositories also make up the ecosystem
    - Professionalization of data stewardship will advance FAIR

# Session II: Monitoring and Evaluating Data Management and Sharing

- Data citation adoption should be the goal, not perfection – act now and the infrastructure will catch up
  - Repositories should support it
  - Researchers should practice it
  - Publishers should implement it
- NIH could push on asking for DOIs for data
- We have data on data sharing: Small number of papers make data available; sharing is associated with international collaborations and higher impact
- Proactive approaches to data sharing/management plans as important planning and research tools
- Provide training opportunities for developing plans – institution, department, PI/lab-level

# Session III: Encouraging Uptake of Data Sharing in the Scientific Community

- Assess quality and reuse of data
- Support practice of adoption over time
- Incremental sharing: start within and across labs first (shows direct gain)
- Establish norms for sharing and meaningful carrots/sticks
- If we *plan our work*, we want to *work our plan* by capturing who/what/where/when/how
- Commit to formal relevant, structured, and mentor-training opportunities for PIs and grad students **and** provide examples for others to follow
- Cultivate subject matter data stewards
- Build in clear rewards for sharing
- Oversight/compliance checking will be needed

# Key Issues Still to Resolve

- Funding/costs
- Resources
- Incentives: career value
- Training modules
- Tools
- Discovery science
- Citation standards
- Privacy/IP
- Tests for data reusability
- Repositories and metadata best practices
- Visibility of PI sharing practices

# Changing the Culture of Data Management and Sharing — A Workshop

## Workshop Highlights from the Co-chairs

Maryann Martone, University of California, San Diego
Richard Nakamura, *Retired*, NIH Center for Scientific Review

Key Takeaways from Both Days: April 29, 2021

# Session IV: Value and Costs of Managing/Sharing Data

- Data management **vs research process/workflow**
- Data management is the **gateway** to developing a culture of data sharing; **Change the culture by providing working solutions**
- **Train for the lab**, not the individual (partner institutions, societies) - will have discipline-specific needs
- Sharing benefit not emphasized yesterday: Development of better data analysis methods (e.g. in genomics, use of public data for adv statistics)
- Models from outside biomedicine may be helpful in navigating our difficult waters
  - **Differential privacy**: "Every query leaks some privacy and validity"
  - Can we view some data as "normal economic goods"?
- "Comparative data reuse vs. integrative reuse"
- "**Data Creators' Advantage:**" Data speed data - collaborations among data users and creators are key
- Costs: Funded by direct or indirect? Cost/benefit considerations for federal budgets

We have data on data sharing and we should get more. Studies of practices and attitudes of researchers are very valuable in navigating this culture change.

# Session V: Shaping a Culture of Sharing

- New regulations impact whether data should be shared and will likely continue to be a moving target in the context of clinical data
- **Let's focus on who has the rights to do what with data** (as opposed to data ownership)
- "Data Governance: **Neither a carrot or a stick but a set of guard rails and some signage"** - community should have input on guard rails
- **Trust** is at the heart of data governance
- "Active stewardship" **Trust promotes and encourages use**
- Who will benefit from the new policy? What **values** are we trying to promote? (fair information practice principles)
- Ethical principles from cultures/communities may not align with those of researchers/policies and these communities deserve our **respect**
- **Don't be blind to people** and their interests behind the research
- How can we **build** a data **trust** that is trustworthy?
- Funders need **tools** too! Need to be able to monitor the policy
- Don't want to penalize anyone for the policy
- Training and educational materials can have a very large impact
- **Reward good data sharing behavior**

# Session VI: Are We Ready for 2023?

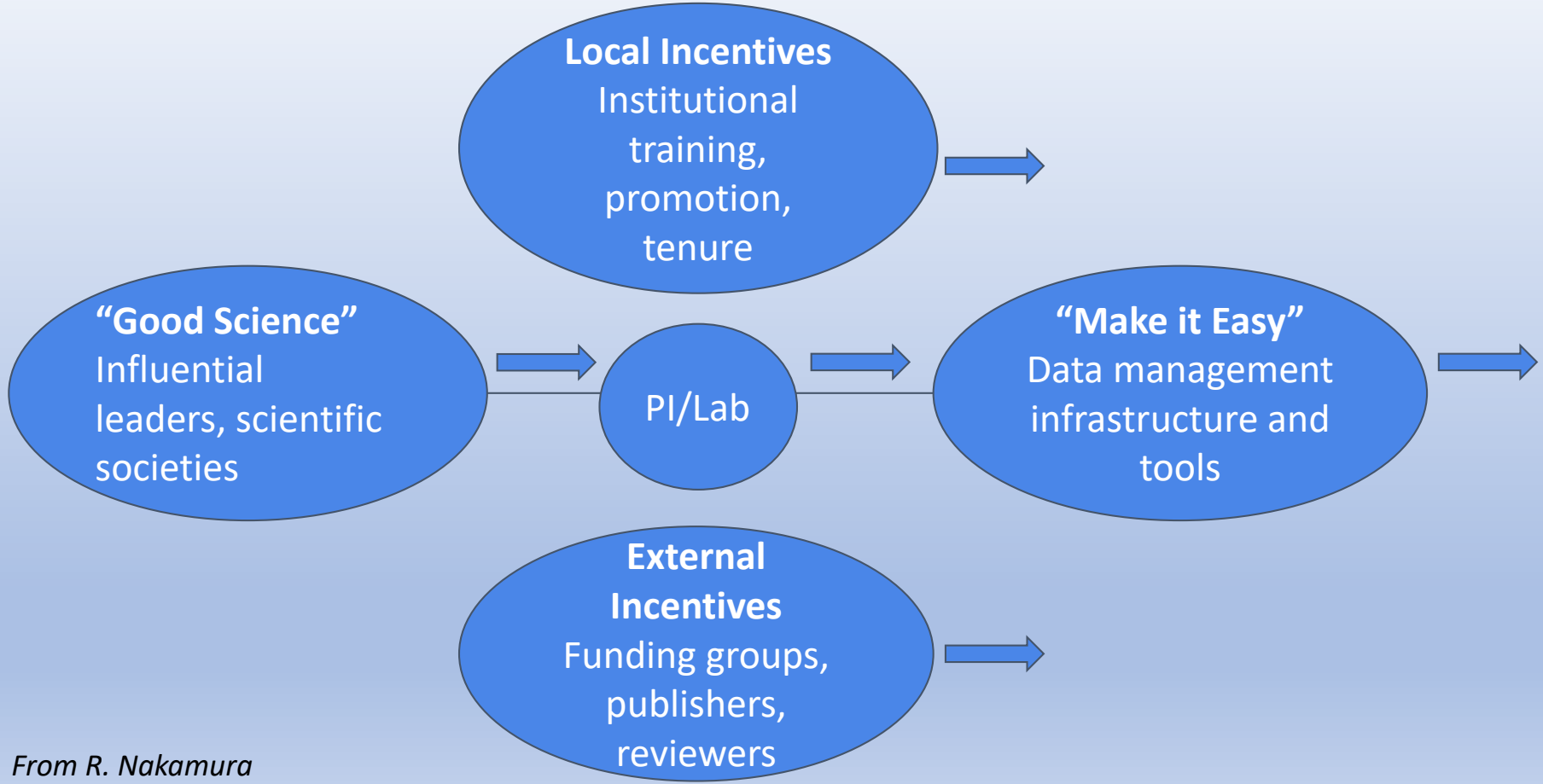
- Managing and sharing takes a village
  - If we create the community, we can create the practice of sharing/managing
- Support for training and putting learning into practice is key
- How can NIH use all of their processes to support the policy? What do the staff need?
- The success of the policy is focused on advancing research to drive health
- Need to integrate and embed equity in all aspects of data management/sharing

# Some Overarching Thoughts

- The tenor of conversation around data sharing has changed: desire that data sharing be impactful
- Data management is the gateway to data sharing
  - Reproducibility requires that data be structured in a way that it can be shared
  - Powerful examples shared of the benefits a well-managed laboratory can offer researchers: Are such tools and systems within reach of most?
  - Think about the unit that should be targeted
- Truly breathtaking body of work, tools and experiences that can be drawn upon as we seek to change the culture, so reinforces we will learn by doing.
  - Training materials are desperately needed “FAIR Data Resource”: Guidelines, examples, expertise need to be shared on a wide scale
  - A lot of that body of work occurs outside of biomedicine: need to bring these into biomedicine (economic models, digital information principles, NIST Research Data Framework)
- Trust and respect: Trust is at the heart of data sharing and data governance but it needs to be earned.
  - Difference in trust will have asymmetric results; will impact data sharing. AI/tools use public data
- Culture change is system-wide: All participants need to have their point of view shifted. Funding agency included.
  - Measurement tools are needed to support funders and institutions

Human expertise is still important in a digital world and that knowledge is distributed

# The System Must Shift to Change Culture

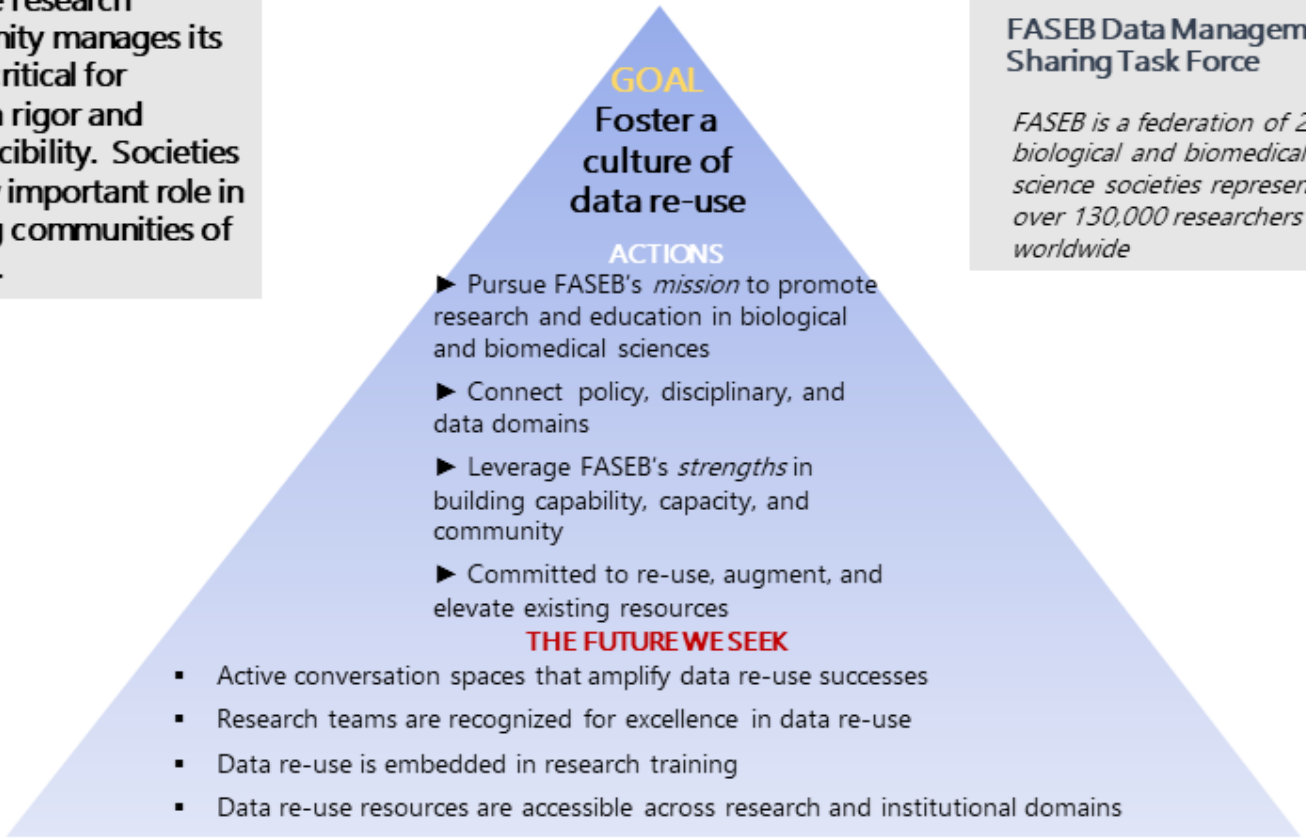



*From R. Nakamura*

How the research community manages its data is critical for research rigor and reproducibility. Societies can play important role in evolving communities of practice.

## FASEB Data Management & Sharing Task Force

*FASEB is a federation of 29 biological and biomedical science societies representing over 130,000 researchers worldwide*




***First program components to launch later in 2021***

FASEB
Federation of American Societies for Experimental Biology

*From Yvette Seger*

# Considerations for NIH: Can we be *more* ready?

- Prioritize research projects that focus on data re-use
- Incentivize the sharing of data from “failed” experiments
- Prioritize adoption of data citation methods
- Encourage the use of digital object identifiers (DOIs)
- Make data management and sharing plans machine readable
- Encourage alignment of incentives for sharing across the science “system”
- Clarify funding sources for data sharing and management
- Assist disciplinary organizations and societies in facilitating FAIR sharing
- Upgrade the data management and sharing infrastructure across biomedicine
- Generate more data on data sharing: study practices, track costs, provide examples
- Create flexible consent guidance but with guide rails
- Request tangible evidence of respect for data sharing: e.g., biosketch should include a contribution to data sharing
- Develop tools to make this policy stick
- Address data sharing broadly at NIH, giving clear guidelines to NIH staff and the tools that they need

# What Does Success Look Like?

- **Growing proportion of publications** indicating availability of background data and calculations needed to check and support conclusions
- Researchers, societies, institutions, libraries, journals, funders take **an active role** in developing a culture of data sharing
- **Data is valued** by researchers, institutions, societies, journals and funders as a first class research output
- **Data are born FAIR**; use of PIDs for key entities is routine
- **Increase in reuse** of shared data; new careers and new industries are launched
- Availability of information about sharing practices of scientists to reviewers, funding program staff, other scientists and public
- Expanding, **sustainable repositories** with FAIR data
- **Data stewards are professionalized** and their expertise increases FAIR
- Shortening of average time to data sharing and reuse; **more creative use of public data**
- Growing proportion of **training programs** with DMS modules
- **Effective data management** accelerates across the biomedical enterprise
- Biomedical data sharing operates with a clear set of **guard rails and sign posts** for ethical and just sharing of information
- Trust and respect are the norms

# What's Next After This Workshop?

- Workshop materials and videos will be posted to the website
- You should have received a **post-workshop survey** by email; please share your thoughts
- Stakeholders at all levels are encouraged to think about the challenges discussed over the last 2 days to help move this field forward:
  - Individual researchers - PIs, trainees
  - Departments
  - Scientific/professional societies
  - Libraries
  - Publishers
  - Repository staff

- Thank you for participating with us!