National Institutes of Health
Office of Data Science Strategy

Concept Clearance

Developing Experts for Better Biomedical and Behavioral Research Data: FAIR and AI/ML-Ready Data

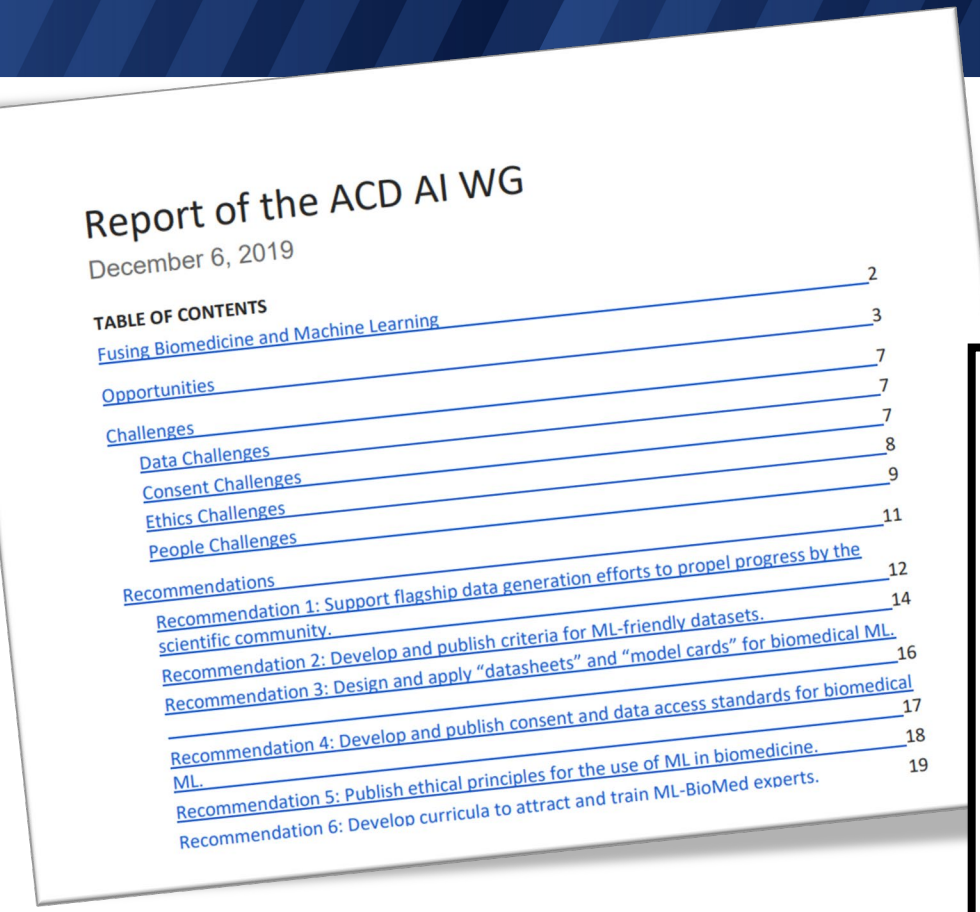
Laura Biven, Ph.D.
Technical Lead for Data Science
NIH Office of Data Science Strategy

Council of Councils | *May 20, 2021*

ODSS Concept Clearance

- **Concept Clearance**: New ODSS Program
- **Title**: Developing Experts for Better Biomedical and Behavioral Research Data: FAIR and AI/ML-Ready Data
- **Objectives**: To develop new experts able to make biomedical and behavioral research data FAIR and AI/ML-ready, and to advance the field of data science for AI in biomedicine and health.
- **Funds Available and Anticipated Number of Awards**:
  - FY22-26: $10M/year to support
    - ~10 multi-institutional, 5-year centers;
    - One 5-year coordination center and workshops

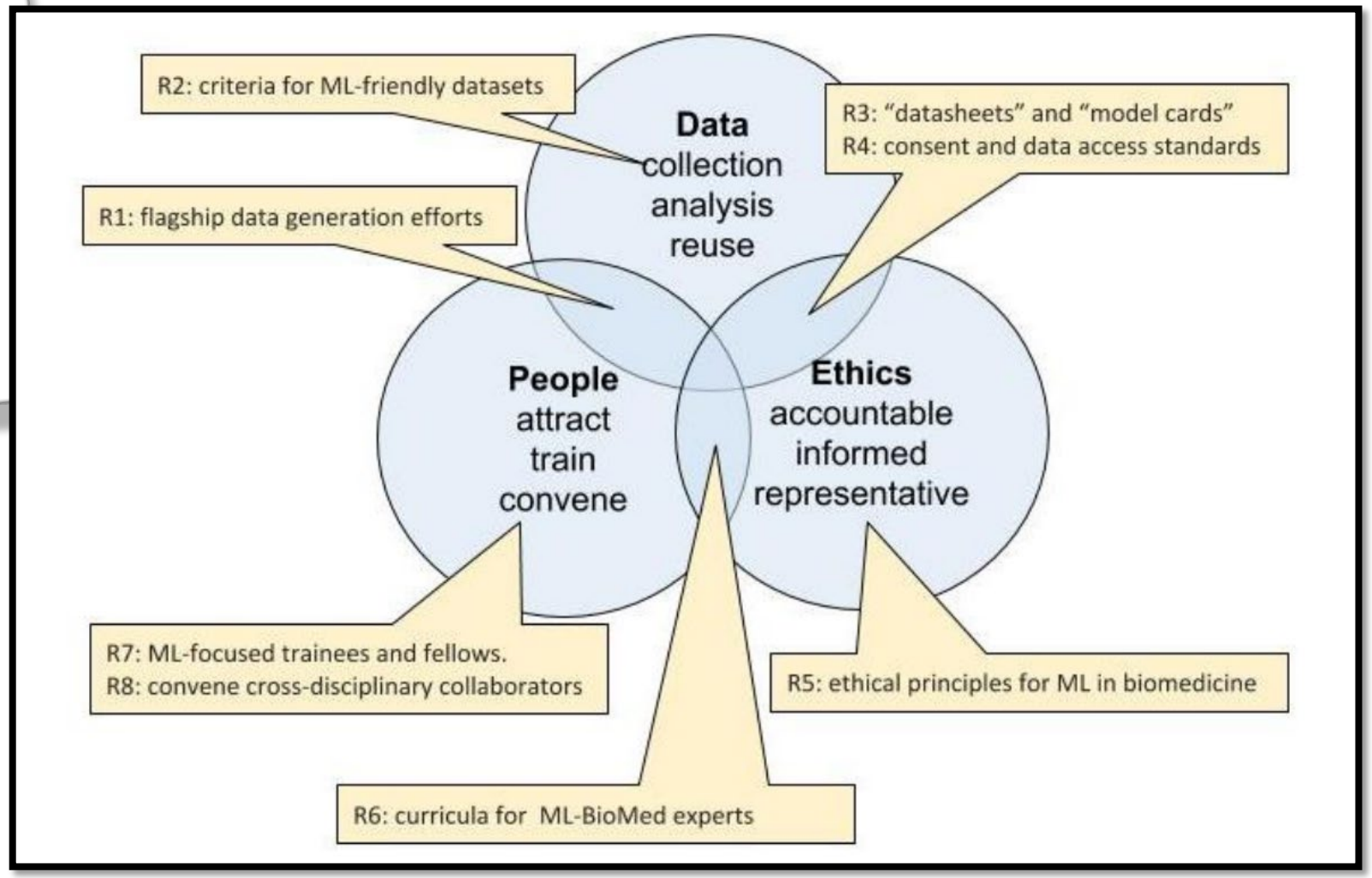
NIH ACD AI Working Group Report



Report of the ACD AI WG

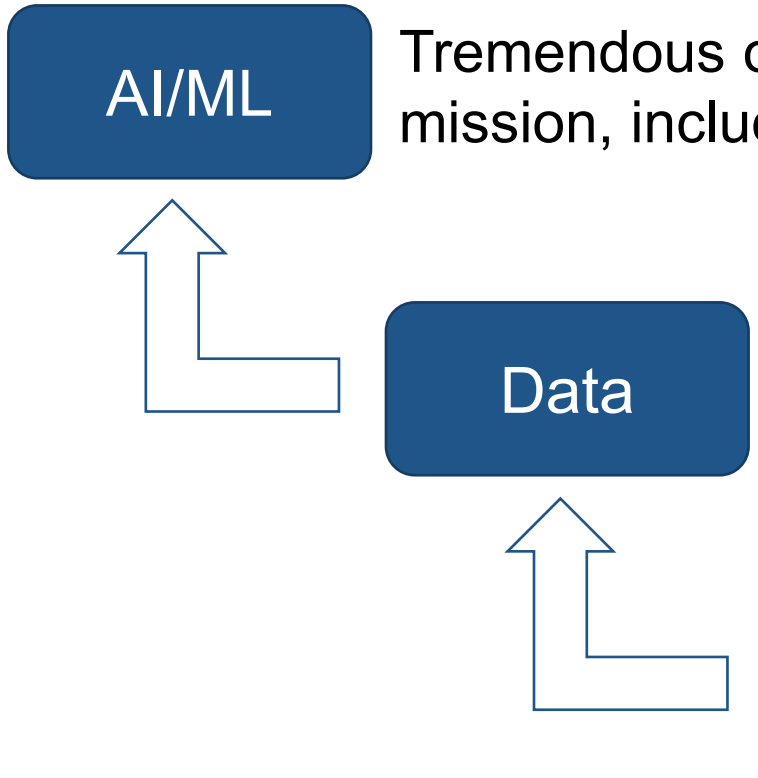
December 6, 2019

TABLE OF CONTENTS



https://acd.od.nih.gov/documents/presentations/12132019AI_FinalReport.pdf

Motivation

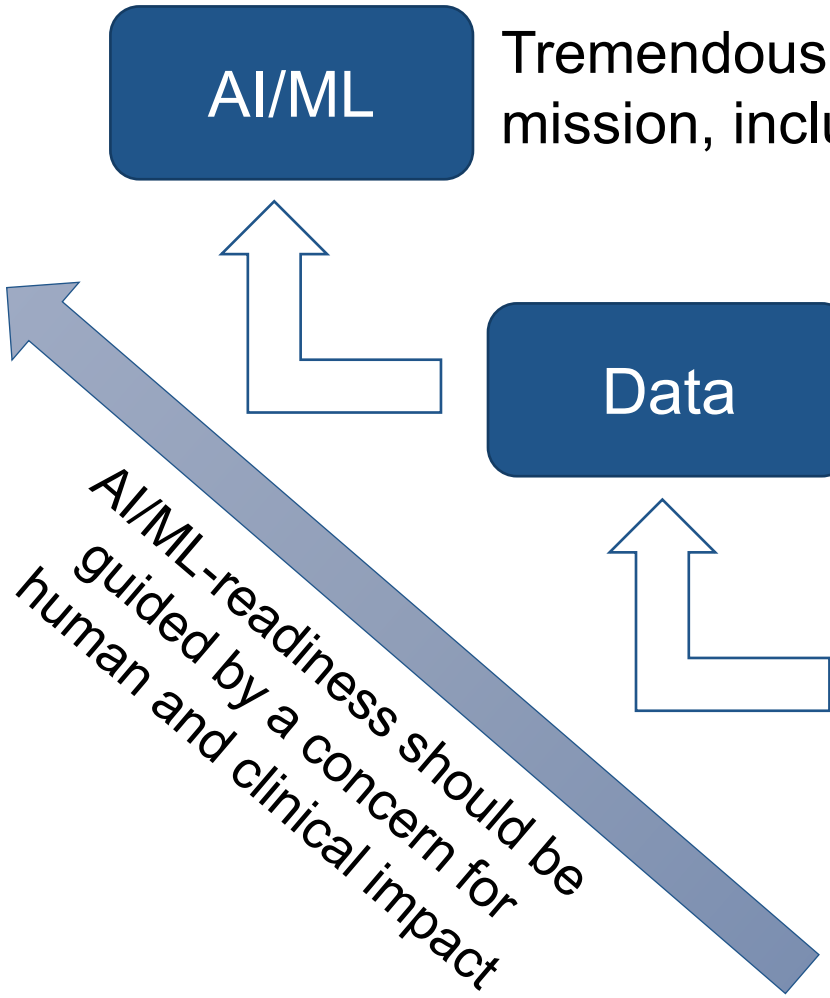


Tremendous opportunity for data-driven discovery across NIH mission, including from AI/ML technologies

Requires high-quality, reusable (FAIR) and AI/ML-ready data

Making data FAIR and AI/ML-ready requires interdisciplinary skills not typically held by PIs in biomedical and behavioral research and not typically addressed by current data science training

Motivation



Tremendous opportunity for data-driven discovery across NIH mission, including from AI/ML technologies

Requires high-quality, reusable (FAIR) and AI/ML-ready data

Making data FAIR and AI/ML-ready requires interdisciplinary skills not typically held by PIs in biomedical and behavioral research and not typically addressed by current data science training

Components of the Proposed Initiative

I. Interdisciplinary Education and Training Centers

II. Coordinating Center

Interdisciplinary Education and Training Centers

5-year awards to…

Develop and deliver new curriculum and training offerings including ethics and the attention to culturally sensitive data and use cases

- Develop outreach plans and partnerships to promote diversity among the trainees and diversity of data
- Develop and implement a plan for sustaining the training programs after the 5-year award
- Develop and implement plans for sustaining the workforce and incorporating developed expertise into the research culture at participating institutions
- Evaluate the impact of their offerings, in part by tracking the career trajectories of trainees and eliciting feedback

Interdisciplinary Education and Training Centers

Reflect the interdisciplinary skills and competencies through multi-department partnerships.

Recruit new, diverse talent with multi-institutional partnerships.

Feature relevant and culturally sensitive data and use cases with multi-institutional partnerships.

Interdisciplinary Education and Training Centers

- **Funding Mechanism:** U54 applications for linked cooperative agreements in Training (TL1), Research projects (RL1) and administrative activities (UL1)

Coordinating Center

- Organize workshops and other events to **share best practices** and strengthen the network of Centers
- Broker **cross-center training** opportunities for students
- Coordinate annual events and meetings for trainees to develop a **community of experts**

Funding Mechanism: U01 Cooperative Agreement

- **Available funds**: $10M /year from ODSS
  - Partner ICs may also contribute funding, but this is not required
  - Partner ICs include: **NHLBI, NIA, NIAMS, NIBIB, NICHD, NIDA, NIDCD, NIDDK, NIEHS, NIGMS, NIMH, NLM,**

- **Number of awards**:
  - ~10 Center awards, 5-year duration
  - 1 Coordinating Center award, 5-year duration

- **Number of experts trained**:
  - ~ 125 students per year

- **Funding mechanism**: Linked combination of U, R, and T mechanisms

Sept. 2021: Publish Centers FOA

July 2022: Centers awarded

Year 1: establishing partnership, developing and testing curricula; some part-time enrollment and online offerings

Year 2: part-time enrollments and online offerings to grow; full-time programs to begin

Year 3: all programs to grow, some assessment to begin

Years 4-5: some assessment of impact

FY26: Consider renewal of the program

Award period: July 2022 – July 2026

Sustainability and Culture Change

This initiative provides funding to establish new training programs and catalyze culture change. (Renewals of the 5-year award should not be assumed.)

Sustainability of the training beyond the 5-year award

Career paths for the workforce and incorporation of experts into research culture

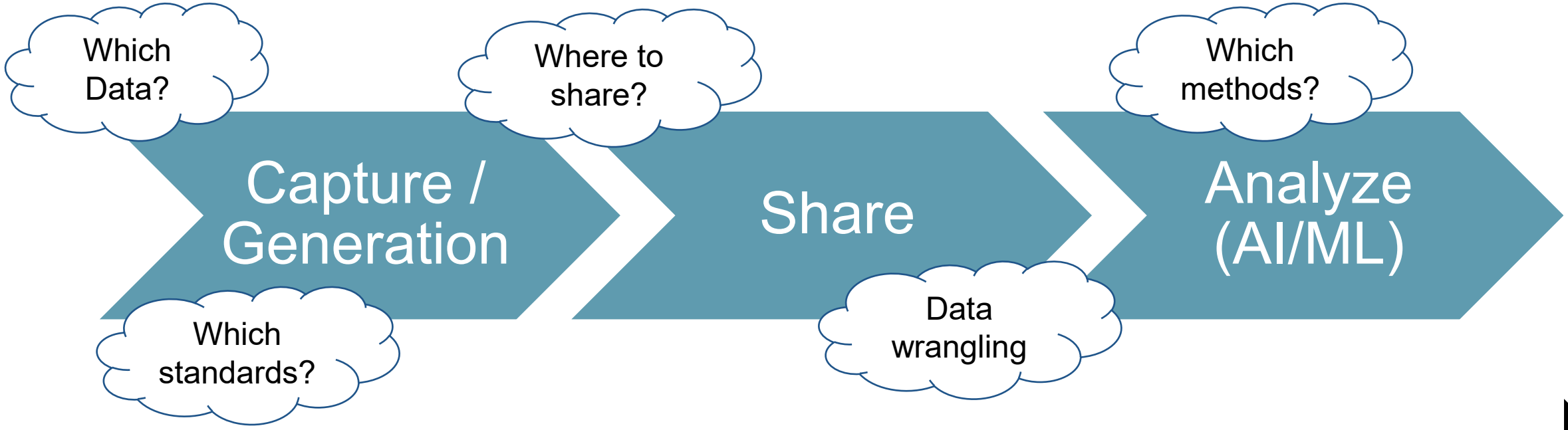
Future of the NIH program after 5 years

NIH Working Group

- Laura Biven (ODSS)
- Jennifer Collins (NIEHS)
- Allissa Dillman (ODSS)
- Lisa Federer (NLM)
- Michele Ferrante (NIMH)
- Alison Gammie (NIGMS)
- James Gao (NEI)
- Kenneth Gibbs (NIGMS)
- Elizabeth Ginexi (OD/OBSSR)
- Anthony Kirilusha (NIAMS)
- Jessica Mazerik (OD/HEAL)
- Asif Rizwan (NHLBI)
- Carol Shreffler (NIEHS)
- Alyssa Tonsing-Carter (ODSS)

Backup

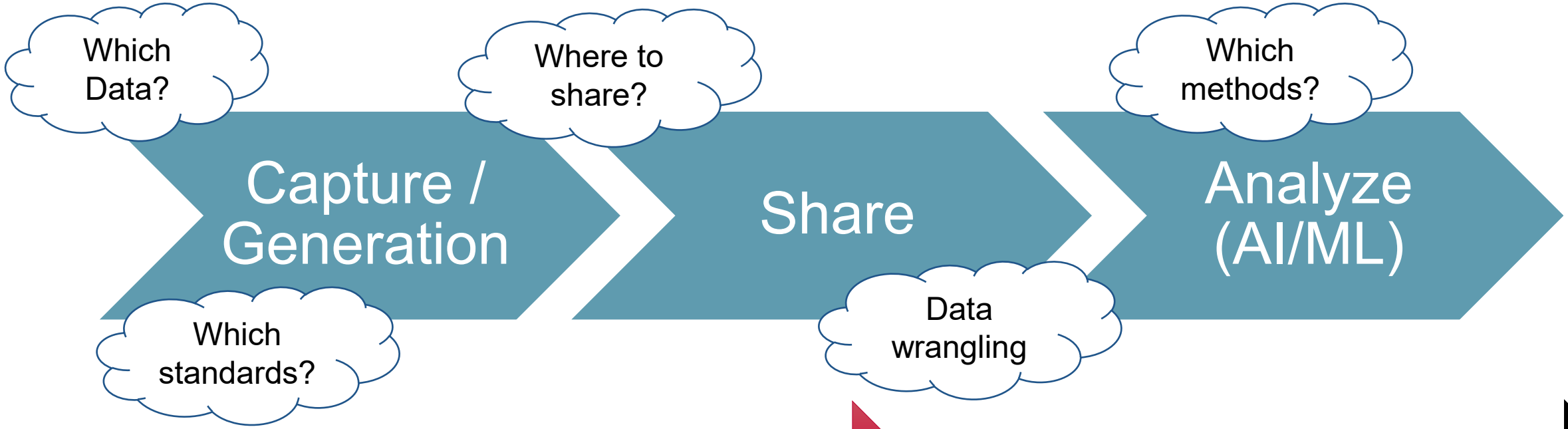
Oversimplified Data Lifecycle



Current AI and data science training

- Multi-disciplinary: Math / stats; data visualization; coding; database and software management; data wrangling; AI/ML; … ethics;…

Oversimplified Data Lifecycle



Focus of ODSS effort

Current AI and data science training

- Multi-disciplinary: Information sciences; data stewardship; computer science; AI/ML; ethics; …

- Multi-disciplinary: Math / stats; data visualization; coding; database and software management; data wrangling; AI/ML; … ethics;…

What will these experts do?

Engage with biomedical research teams throughout the data lifecycle, beginning at research planning

Lead and develop FAIR data management practices, community standards, etc.

Work across domains and with experts from areas such as data, legal, and ethics

Facilitate engagement with AI/ML community

Gain unique insights into the needs of biomedical researchers and AI/ML community

Work with interdisciplinary teams to develop new practices and capabilities (e.g., machine actionable provenance; ethics for machines…)

What will these experts become?

- Leaders of NIH-funded efforts for data interoperability and stewardship
- Co-investigators on multi-disciplinary activities
- Co-authors on publications and datasets
- Innovators of new capabilities and best practices
- Part of biomedical-focused AI industries, government agencies, and non-profits

This is new and there is no set career path. This will need to evolve.

Synergies between Bridge2AI and ODSS Activities

Bridge2AI

- Core focus is flagship data generation efforts
  - multidisciplinary teams oriented around a general problem domain, with co-designed tools, data standards, ethical framework, and pilot data collection in a single award.
  - Cross cutting activities in data management, tools, and ethics support the data generation efforts

ODSS AI Workforce

- Builds on foundational activities toward a FAIR data ecosystem
- Addresses strategic workforce gap
- Seeks to build community-wide capabilities

Complementarity and Synergy:

- ODSS Workforce curriculum development can be used and tested in B2AI
- B2AI awardees will contribute to the development of the program through joint workshops and partnerships

ODSS and CF Coordination:

- Regular meetings
- Agreement on terminology
- Coordination on focus and scope of activities
- Planning NIH AI Website

FY21 Supplements: NOT-OD-21-079

- Administrative supplements for workforce development at the interface of information sciences, artificial intelligence and machine learning (AI/ML), and biomedical sciences
  - Existing institutional training grants are eligible
  - Supplements will be up to $80,000 in direct costs
- Purpose: To support the development and implementation of curricular or training activities to develop the competencies and skills needed to make biomedical data FAIR AI/ML-ready.
  - Curriculum development: exportable training modules and integrated training plans
  - Training: events or other educational experiences where the structure and outputs are shared
- Targets people from a variety of backgrounds and career stages.
- Recognizes that different disciplinary areas and data types may need personnel with different sets of expertise and training experiences, including basic, clinical, population, and behavioral and social sciences. Programs are not expected to cover all of these disciplines, but the educational experiences should prepare those who complete the program to work with a range of data types.
- Must be aligned with NIH's interest in diversity (NOT-OD-20-031)
- https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-079.html