Presentation of the Draft Report of the NExTRAC Working Group on Data Science & Emerging Technology

Sachin Kheterpal, MD, MBA

Pilar Ossorio, PhD, JD
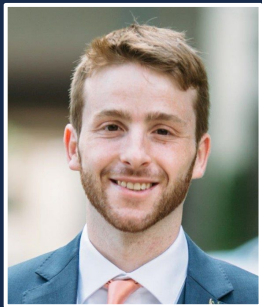
Meeting of the Novel & Exceptional Research & Technology (NExTRAC) Advisory Committee

August 29, 2023

Data Science & Emerging Technology Working Group

Novel and Exceptional Technology and Research Advisory Committee (NExTRAC)

Staffing Support (NIH)

Jesse Isaacman-Beck

T.J. Kasperbauer

Members

Lorraine M. Albritton — Columbia University
Suzanne B. Bakken — University of California, San Diego
Cinnamon Bloss —
Alan I. Leshner — AAAS
Husseini K. Manji — University of Oxford
Debra J.H. Mathews — Johns Hopkins University
Margaret F. Riley — University of Virginia
University of Tennessee

Co-Chairs

Pilar Ossorio
University of Wisconsin, Madison

Sachin Kheterpal
University of Michigan Medical School

NExTRAC Working Group on Data Science and Emerging Technology

1. Define and characterize the types of research questions that require increasing granularity and aggregation of data about individuals that are likely to be addressed through emerging technologies:

   Phase 1   A list of types of research questions of most relevance to the NIH over the next ~5-10 years

2. For those questions/technologies, consult with stakeholders to discuss and assess the value of and potential implications for individuals, groups, and society:

   Phase 2   An analysis of how stakeholders perceive these questions and weigh the inherent risks and benefits

NExTRAC Working Group on Data Science and Emerging Technology (continued)

1. Define and characterize the types of research questions that require increasing granularity and aggregation of data about individuals that are likely to be addressed through emerging technologies. Please consider but do not limit the scope to:

   - Goals of such research studies and how they advance the NIH mission;

   - Emerging technologies that may generate potentially sensitive datasets;

   - Data types generated and their sources (e.g., digital health devices, EHR platforms) with an emphasis on exploring new data types or unique sources; and

   - Data science platforms and tools that facilitate data access, combination, and analysis (e.g., artificial intelligence, cloud computing).

Phase 1, Section III: Types of research questions

WG deliberation, discussion with subject matter experts

Phase I Products

List of types of biomedical research questions involving new types of data, new types of analysis, & data linkage & aggregation

Specific types of research questions

- How are personal health data collected from outside of the traditional health system (e.g., wearables, apps, social media) used to study health-related questions and predict health risks?

- How can other consumer and lifestyle data from non-health-specific sources (e.g., sensors in the home and credit card and consumer rewards data) be used to study health-related questions and predict health risks?

- What is the role of computer-based technologies (e.g., artificial intelligence, machine learning, automated image analysis) in advancing health decision-making?

- How can new methods for integrating data while maintaining the security of private information (such as Privacy Preserving Record Linkage) be used for precision medicine and public health?

- How should the research context (e.g., clinical, public health) and participants' consent status affect data linkage and aggregation?

5

A note…

Generative AI
(e.g., Large Language
Models like ChatGPT) had
yet to emerge for broad
public use and awareness.

Findings

Emerging technologies to capture, store, and analyze personal health data (e.g., wearables, apps, & artificial intelligence) offer rich resources for research. However, more policy and ethics work is needed to responsibly realize their full benefits.

Phase I

Emerging methods to harmonize and aggregate data from different studies and across the world show great promise for precision medicine and public health. Still, refinements are needed to limit variation that can reduce significance, accuracy of results, and generalizability of conclusions.

Phase 1, Section III: Types of research questions

WG deliberation, discussion with subject matter experts

Phase I Products

List of types of biomedical research questions involving new types of data, new types of analysis, & data linkage & aggregation

Phase 2, Section IV: Observations from Public Engagement

Community conversations and webinars

WG summary of themes by research topic

Workshop with scholars, developers, researchers

Phase II Products

All summary feedback, observations and analysis considered by WG

8

NExTRAC Working Group on Data Science and Emerging Technology

2. For those questions/technologies, consult with stakeholders to discuss and assess the value of and potential implications for individuals, groups, and society. Please consider, but do not limit the scope to:

   - Attitudes and perspectives about sharing participant data to advance biomedical research, specifically through the lens of balancing research risk (e.g., privacy, autonomy) with research deliverables; and

   - How these perspectives may evolve depending on the context of who is to benefit or assume risk, whether it be at the individual level, through the community, or broader society's expectations for public health advancement.

   In addressing this charge, **the working group shall convene consultations with stakeholders including, but not limited to, research participants, patient groups, ethicists and privacy experts, data scientists, technology developers across sectors, and public health officials.**

Harlem, NY
Nov 9, 2022

Bronx, NY
Nov 10, 2022

Alamosa, CO
Dec 5, 2022

Jackson, MS
Dec 14, 2022

24 Attendees
Union Settlement

25 Attendees, en Español
Coalición Mexicana

23 Attendees
Trinidad State College

25 Attendees
Jackson State University

Who we have engaged:
Community Conversations, Webinars with Advocates

Dearborn, MI
Jan 18, 2023

Rural & Rare Disease Organizations
Jan 24, 2023

Santa Clara, CA
Jan 24, 2023

American Indian, Alaska Native
Advocates; Feb 16, 17

24 Attendees
Arab Community Center for
Economic and Social Services

15 Attendees
Virtual Webinar

20 Attendees
Silicon Valley Network,
Ding Ding TV

13 Attendees
Virtual Webinar

To all who joined us in their communities and generously shared perspectives:

What Community Members said: High Level Themes

- Support for personal data use and analysis dependent on:
  - Who is receiving the data (e.g., for-profit vs. academia)
  - Whether they are trustworthy
  - Whether there would be a significant return in terms of health benefits to the individual or community

- Concerned about:
  - How little they know about how their data are used
  - How little control they seem to have over those uses

- Generally want more:
  - Transparency in what uses of their data would entail
  - Reassurances that the benefits would in fact be worth the potential risks

Public engagement as a throughline
Workshop with ethicists, developers, researchers

Using Public Engagement to Inform the Use of Data in Biomedical Research

352 Views

February 27, 2023, UC San Diego

- 4 Panels on Research Topics
- Reflections on Participant Feedback
- Q/A open to the public

Workshop materials

bit.ly/3kv8Uc9

Munmun De Choudhury, Georgia Institute of Technology

Jenna Wiens, Univ. of Michigan School of Medicine

Kendra Nervik, Univ. of Wisconsin, Madison

Luca Foschini, Evidation, Sage Biosciences

George Hripcsak, Columbia Univ., OHDSI

Jennifer Wagner, Penn State Univ.

Monica Rodriguez, Pyxis Partners

Kolbi Brown, Pyxis Partners

Deven McGraw, Invitae

Jay Schnitzer, MITRE

Pyxis Partners

Shaun Grannis, Indiana Univ., Regenstrief Institute

Lucila Ohno-Machado, Yale School of Medicine

Rupa Valdez, University of Virginia

Krystal Tsosie, Arizona State Univ.

Maya Sabatello, Columbia Univ.

Lisa Lehmann, Harvard Univ.

Ronnie Tepp, Pyxis Partners

Mildred Cho, Stanford Univ.
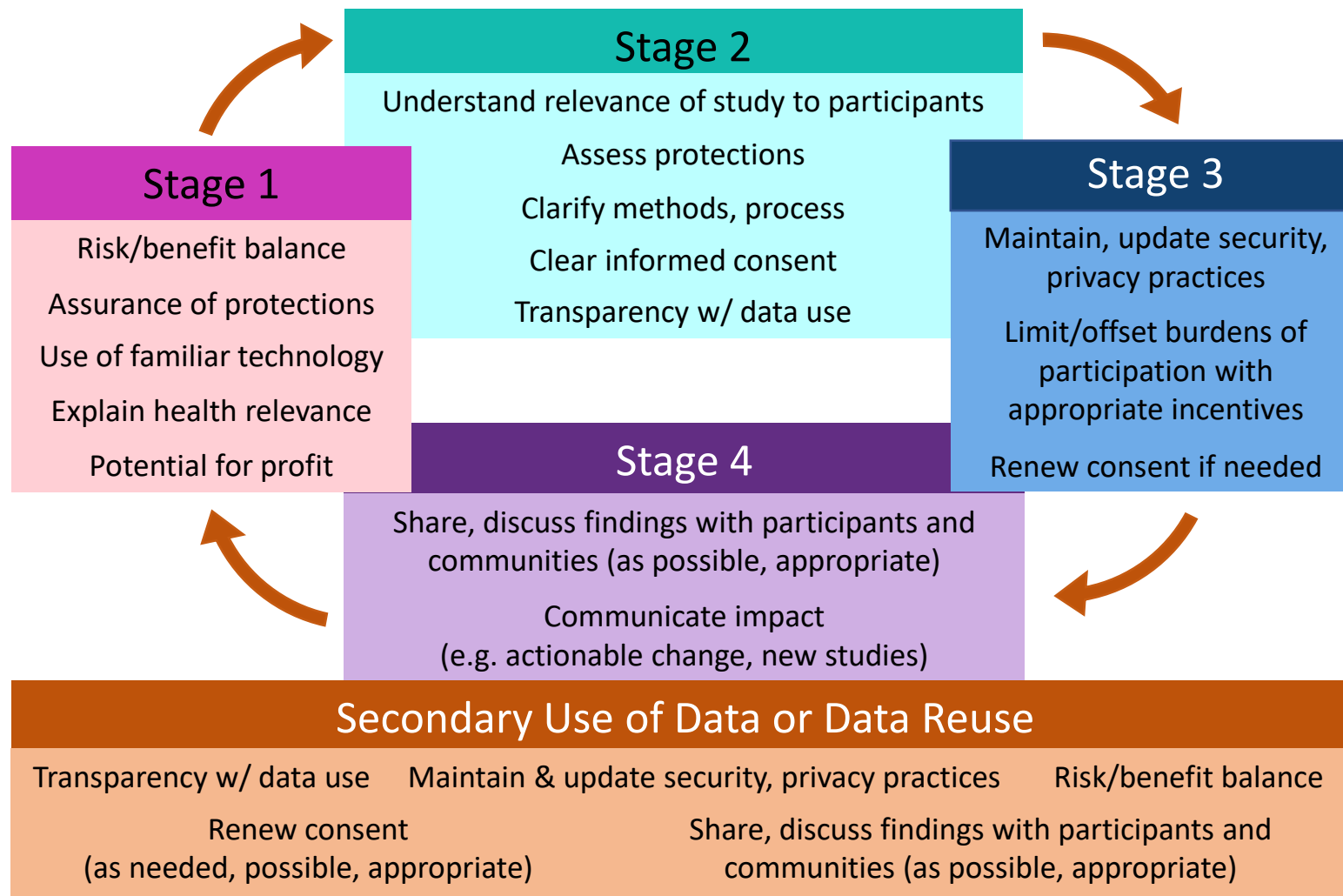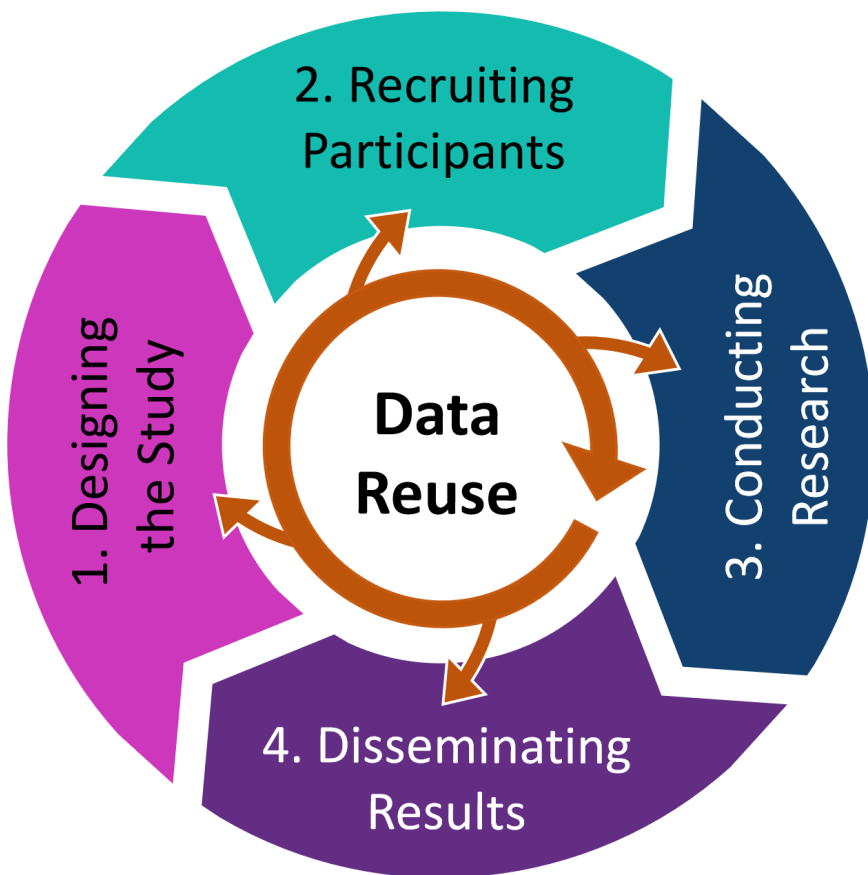
Bud Tribble, Apple (former)

Aaron Coleman, Fitabase

Palladian Partners

To all facilitators, scholars, and developers who contributed to this effort

Opportunities to engage with participants in biomedical research with personal data

Research and Innovation Pipeline

Data Reuse

1. Designing the Study
2. Recruiting Participants
3. Conducting Research
4. Disseminating Results

Stage 1
- Risk/benefit balance
- Assurance of protections
- Use of familiar technology
- Explain health relevance
- Potential for profit

Stage 2
- Understand relevance of study to participants
- Assess protections
- Clarify methods, process
- Clear informed consent
- Transparency w/ data use

Stage 3
- Maintain, update security, privacy practices
- Limit/offset burdens of participation with appropriate incentives
- Renew consent if needed

Stage 4
- Share, discuss findings with participants and communities (as possible, appropriate)
- Communicate impact (e.g. actionable change, new studies)

Secondary Use of Data or Data Reuse
- Transparency w/ data use
- Maintain & update security, privacy practices
- Risk/benefit balance
- Renew consent (as needed, possible, appropriate)
- Share, discuss findings with participants and communities (as possible, appropriate)

Opportunities to engage with participants in biomedical research with personal data (continued)

Research and Innovation Pipeline

Data Reuse

1. Designing the Study
2. Recruiting Participants
3. Conducting Research
4. Disseminating Results

Stage 1

Stage 2

Stage 3

Stage 4

Key Concepts That Emerged
Trust & Trustworthiness
Transparency & Respect
Informed Consent & Notification
Return of Results & Return of Value

Secondary Use of Data or Data Reuse

Phase 1, Section III: Types of research questions

WG deliberation, discussion with subject matter experts

Phase I Products

List of types of biomedical research questions involving new types of data, new types of analysis, & data linkage & aggregation

Phase 2, Section IV: Observations from Public Engagement

Community conversations and webinars

WG summary of themes by research topic

Workshop with scholars, developers, researchers

Phase II Products

All summary feedback, observations and analysis considered by WG

Section V: Findings, Recommendations

Observations from feedback and discussions

Findings based on observations

Overarching Topics for Recommendations to NIH

1. **Trustworthiness** and **transparency** in research, **respect** for participants and communities
2. Issues with **informed consent** for, control of data sharing & linkage
3. **Return of value** on participation in research, data sharing, & linkage

Findings and Recommendations: Phase II

Recommendation

1 NIH should catalogue, evaluate, and as appropriate, coordinate efforts among current initiatives addressing public engagement, ethical, and technical issues in data science.

Trustworthiness through honest engagement & transparency

Findings and Recommendations (cont.)

Finding

There is misalignment between the expectations of some members of the public and some NIH policies and programs on the use and analysis of personal health data for biomedical research.

Recommendations

2

When NIH invests in public engagement, especially around emerging technologies, the agency should be prepared to update processes and infrastructure and to revise its programs and policies in accordance with public expectations and social justice principles, or clearly explain to the public why the agency may not revise its course.

3

NIH, in consultation with other federal agencies as appropriate, should align its policies on data and research governance with widely held public values.

Community engagement throughout research & innovation
Findings and Recommendations (cont.)

Finding

It is important to consult with research participants, individuals whose data are used in secondary research, families, & communities throughout the biomedical research process when novel types, uses, & analyses of data are involved.

Recommendations

4
NIH should help establish long-term relationships between members of the public and research communities to facilitate sustained public engagement on evolving issues regarding the use, analysis, management, and reuse of personal health data in biomedical research.

5
NIH should encourage investigators to prospectively address bioethical considerations in emerging areas of data science that are particularly important to participant communities, as reflected in public engagement, in their research programs and proposals.

Better communication around personal data use & consent
Findings and Recommendations (cont.)

Findings

More transparency is needed around the research process, how data will be collected, analyzed, and used, and the risks and benefits to participants.

Current regulations and procedures for the use of biomedical research data provide ways for data to be reused in research without consent or adequate oversight.

Linkage of data in the absence of robust informed consent raises concerns about data privacy, transparency, and respect for participants.

Recommendation

6

NIH, in consultation with other federal agencies as appropriate, should develop and disseminate best practices on informing participant communities about how their personal health data are being managed, reused, and linked to other data when consent especially when these best practices extend beyond regulatory requirements.

Return of value, including return of results
Findings and Recommendations (cont.)

Findings

Beyond potential promises of precision medicine, research participants deserve greater return of value for themselves and their communities.

Providing participants with financial benefits and/or increased control over personal data may bolster community support for biomedical research conducted by for-profit entities or in public - private partnerships.

Recommendations

7 NIH should leverage prior research on return of results and support further public engagement to inform best practices for return of results from biomedical research using emerging data science and to determine implementation infrastructure needs.

8 NIH should support exploration of methods for providing benefits and return of value beyond returning research results to participants and their communities, when appropriate and as consistent with current regulations.

1 The NIH should catalogue, evaluate, and coordinate efforts among current initiatives addressing public engagement, ethical, and technical issues in data science as applied in biomedical research.

2 When the NIH invests in public engagement, especially around emerging technologies, the agency should be prepared to update processes and infrastructure and to revise its programs and policies in accordance with public expectations and social justice principles, or clearly explain to the public why the agency may not revise its course.

3 The NIH, in consultation with other federal agencies as appropriate, should align its policies on data and research governance with widely public values.

4 The NIH should help establish long-term relationships between members of the public and research communities to facilitate sustained public engagement on evolving issues regarding the use, analysis, management, and reuse of personal health data in biomedical research.

5 The NIH should encourage investigators to prospectively address bioethical considerations in emerging areas of data science that are particularly important to participant communities, as reflected in public engagement, in their research programs and proposals.

6 The NIH, in consultation with other federal agencies as appropriate, should develop and disseminate best practices on informing participant communities about how their personal health data are being managed, reused, and linked to other data when consent or other such notification extends beyond regulatory requirements.

7 The NIH should leverage prior research on return of results and support further public engagement to inform best practices for return of results from biomedical research using emerging data science and to determine implementation infrastructure needs.

8 The NIH should support exploration of methods for providing benefits and return of value beyond returning research results to participants and their communities, when appropriate and as consistent with current regulations.