DATA SCIENCE AND EMERGING TECHNOLOGY WORKING GROUP

PROGRESS UPDATE TO THE NExTRAC

Sachin Kheterpal, MD, MBA
Associate Dean for Research Information Technology & Professor of Anesthesiology,
University of Michigan Medical School

Pilar Ossorio, PhD, JD
Professor of Law and Bioethics,
University of Wisconsin

Meeting of the NExTRAC
July 14th, 2022

PRODUCTS TO MEET THE CHARGE

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:

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

   Phase 1

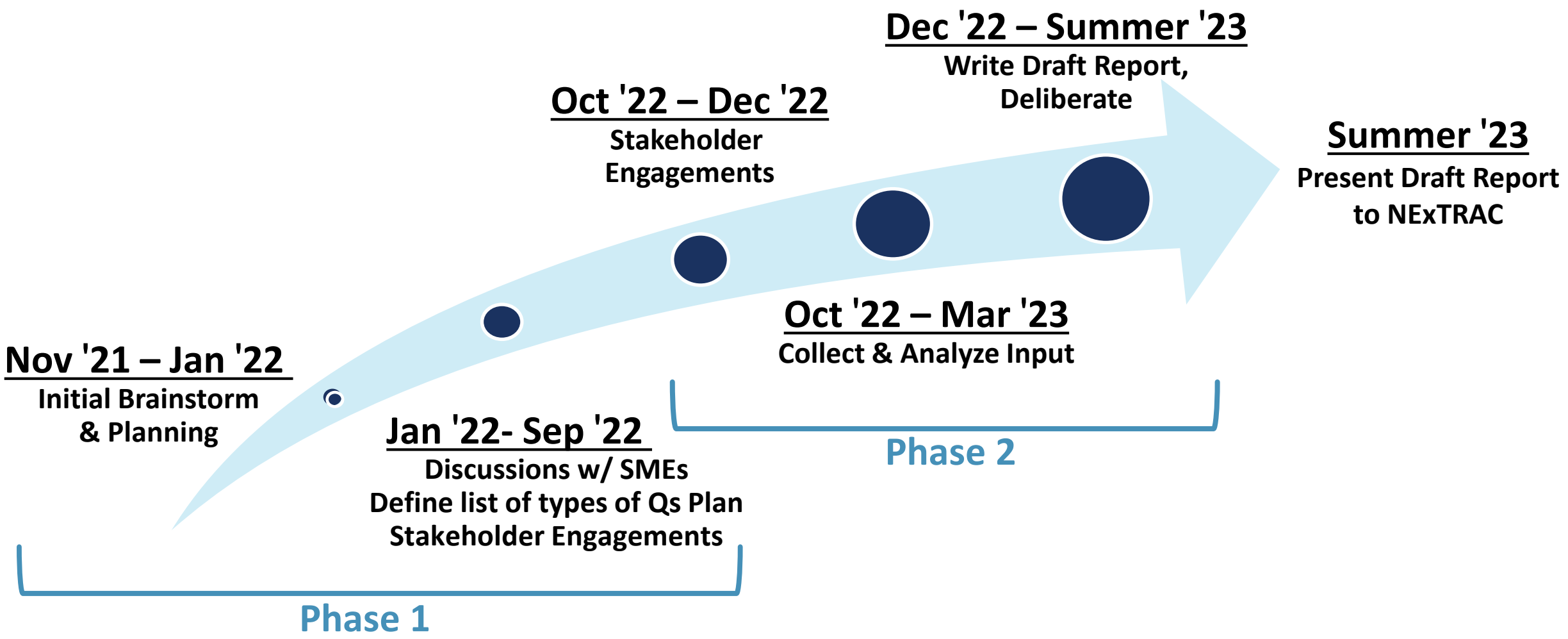
2. For those questions/technologies, consult with stakeholders to discuss and assess the value of and potential implications for individuals, groups, and society:

   Analysis of stakeholders' perceptions of the social and ethical issues raised by such research, how they weigh the inherent risks and benefits and make other value trade offs

   Phase 2

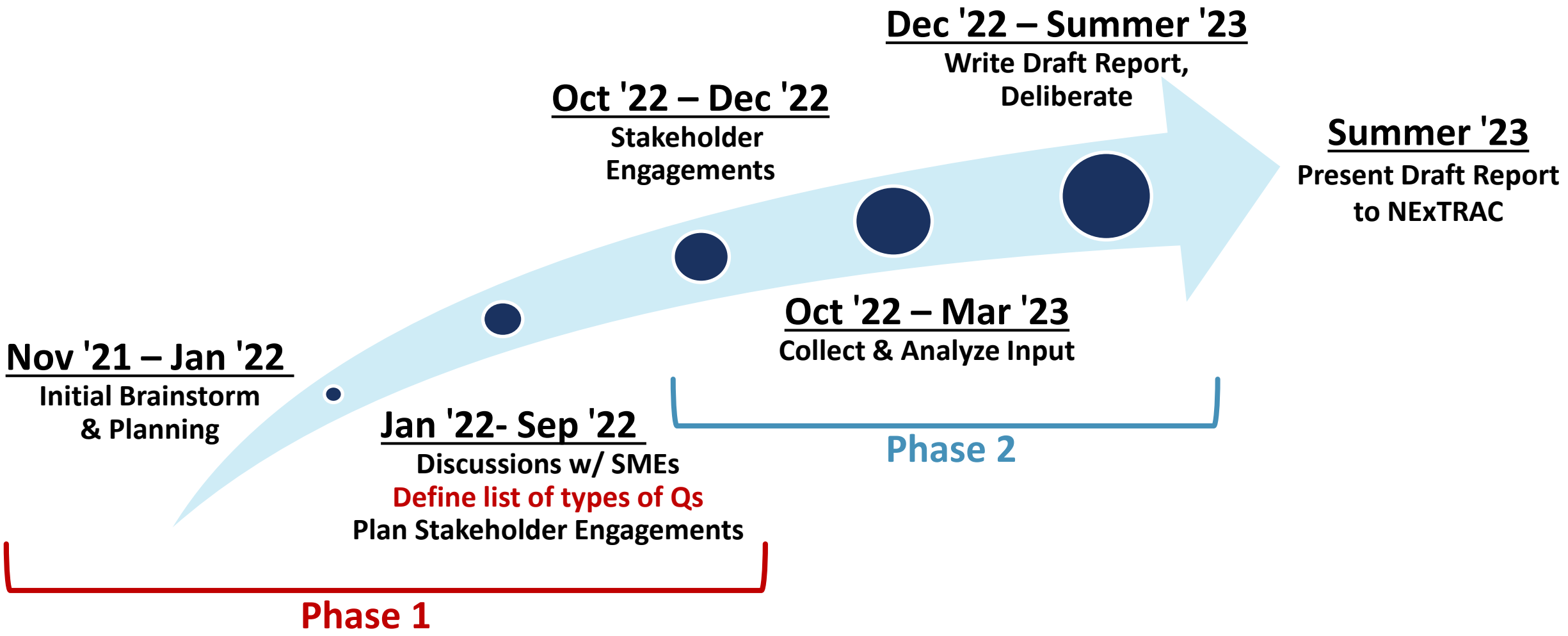
APPROACH

TOWARDS A DRAFT REPORT IN SUMMER, 2023



APPROACH

TOWARDS A DRAFT REPORT IN SUMMER, 2023



A NON-EXHAUSTIVE LIST

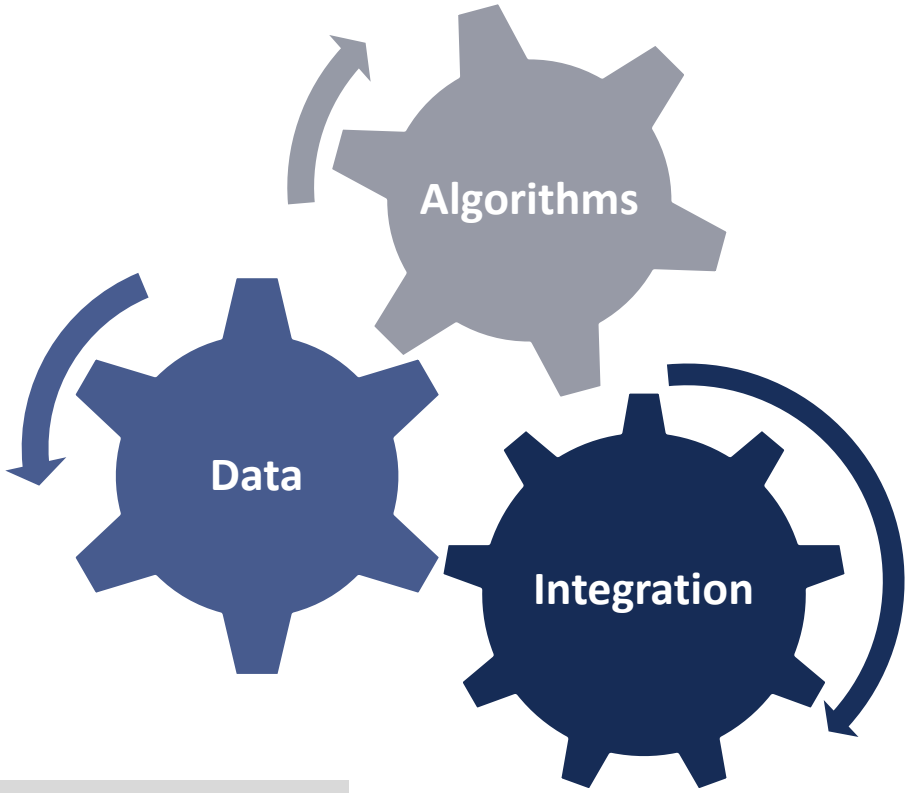
THEMES TO CONSIDER (AND CONSIDERED) THROUGHOUT

- Novelty
- Context
- Complexity
- Sensitivity
- Justice
- Access
- Reliability
- Unintended Consequences
- Control
- Aggregation
- Identifiability
- Privacy, Confidentiality

These themes, and the charge itself, were used to help focus the efforts of the WG to identify specific topics; we acknowledge there are many other research areas and data types that are not the focus of this effort.

FRAMING THE DISCUSSION

THREE OVERARCHING TOPICS



*Note that there is some overlap among these topics

Informing health-related questions through:

- Use of novel **data** from outside of the traditional healthcare system
- Use of **models and algorithms**
- Data **linkage and aggregation** of disparate datasets from multiple sources

For each topic, the Working Group has defined:

1. General types of research questions
2. Exemplar research questions that the NIH might encounter

TOPIC I

USE OF NOVEL DATA FROM OUTSIDE OF THE TRADITIONAL HEALTHCARE SYSTEM

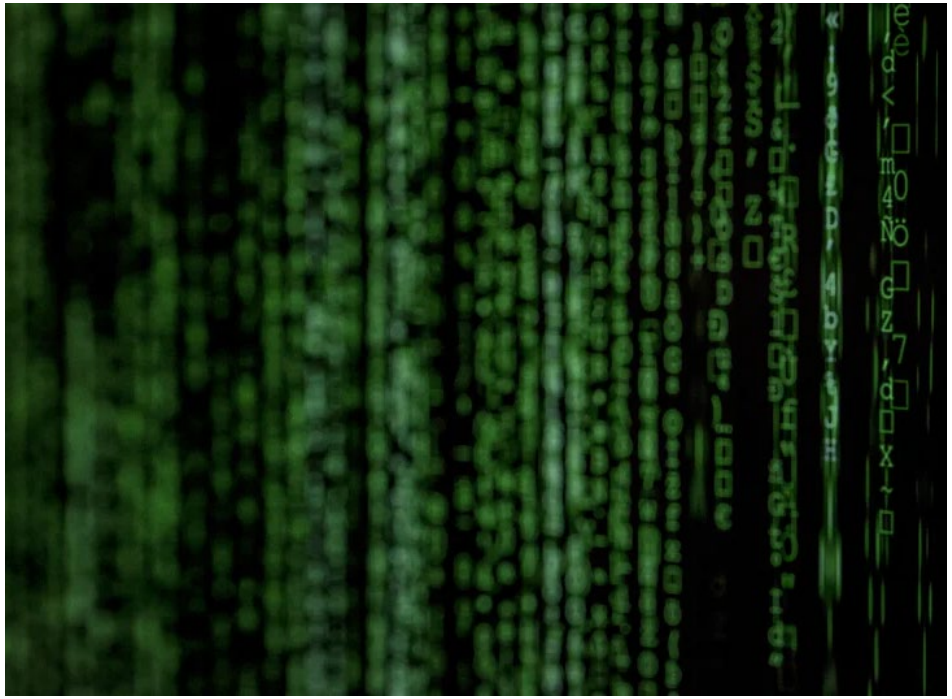
- How are personal health data collected from outside of the traditional health system (e.g., wearables, apps, social media) used to study health-related questions and predict risks?
- How can other consumer and lifestyle data from non-health-specific sources (e.g., credit card, home sensors) be used to study health-related questions and predict risks?
- Can/how can health data be integrated with data on social determinants of health to enable better risk prediction and development of predictive algorithms?

TOPIC 2

USE OF MODELS AND ALGORITHMS

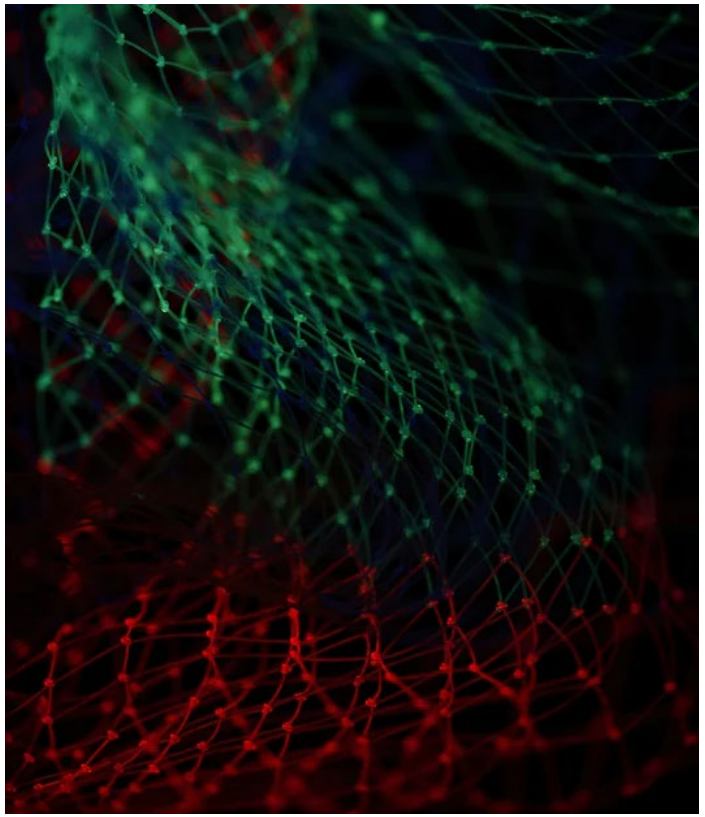
- What is the role of computer-based technologies (e.g., artificial intelligence, machine learning, automated image analysis) in advancing health decision-making?

- Can/how can natural language processing be deployed to analyze data held in health systems (e.g., Electronic Health Records) to provide insights about patient symptoms and disease classification?



TOPIC 3

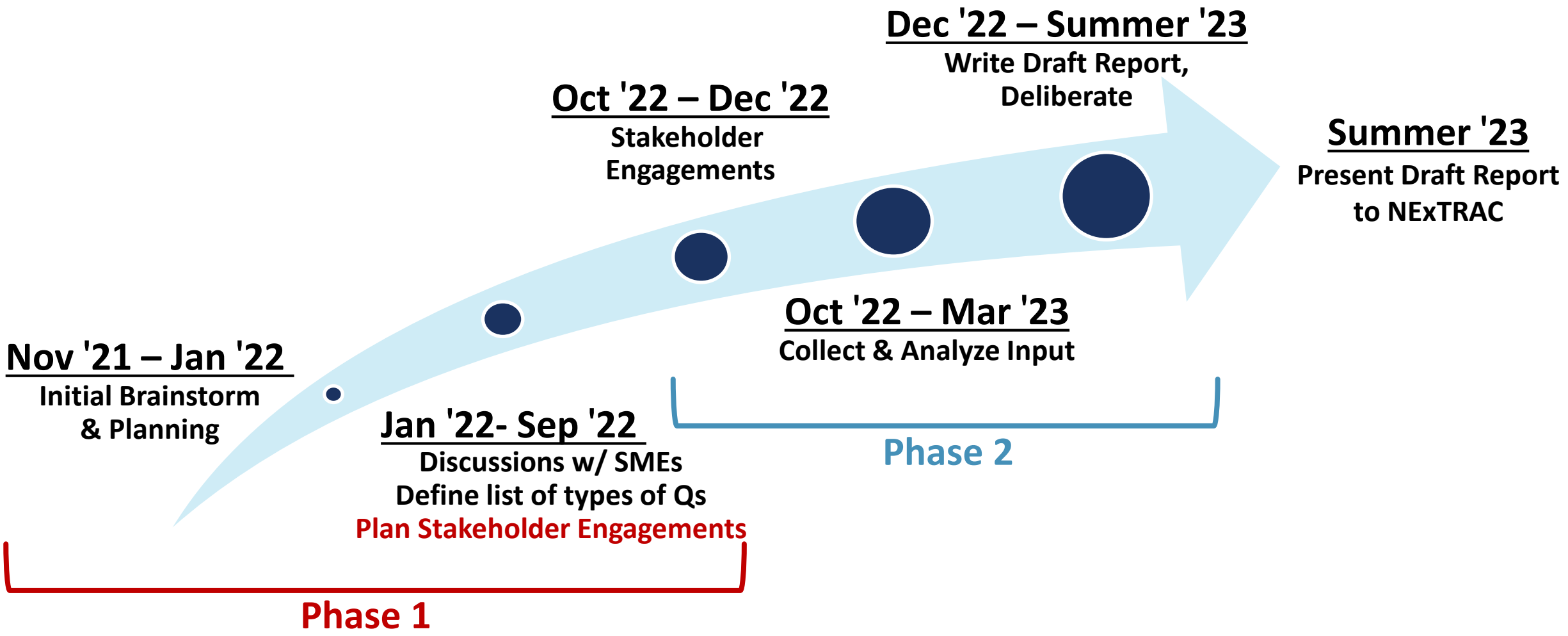
DATA LINKAGE & AGGREGATION OF DISPARATE DATASETS FROM MULTIPLE SOURCES



- Are there opportunities for data standardization so that data from different countries and healthcare systems could be aggregated, linked, and shared across populations?
- Which data sets (e.g., genomics, proteomics, clinical information, clinical imaging, personal health libraries) can be linked and combined with harmonized data aggregators?
- 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?

APPROACH

TOWARDS A DRAFT REPORT IN SUMMER, 2023



PHASE 2

GOALS OF STAKEHOLDER ENGAGEMENT

- Solicit diverse input from a variety of stakeholders with varying perspectives on products from WG charge Phase 1 (informing WG charge Phase 2)
- Engage stakeholders in ways meaningful to them (i.e., outside of the beltway, in communities, etc.) to facilitate bidirectional understanding of the opportunities and challenges in future research directions
- Pilot strategies for engaging individuals and communities in discussions around emerging technologies for further NExTRAC engagements

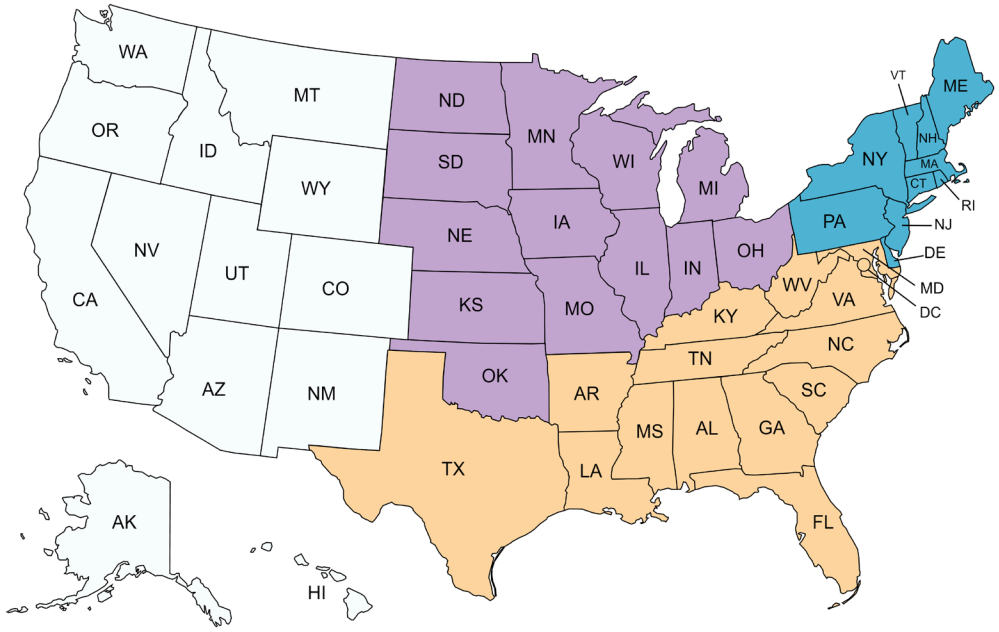
STAKEHOLDER ENGAGEMENTS

PROPOSED FORMAT TO MEET GOALS

- Four engagements that focus on the underlying values and potential implications of the questions identified in Phase 1 for individuals, groups, and society
  - Will take place in the Northeast, Midwest, South, and West (dependent on COVID-19)
- Participants will be selected to incorporate broad and varying perspectives
  - **Engagements 1-3**: direct engagement with research participants, patients, and caregivers
  - **Engagement 4**: discussion with researchers, technology developers, educators, etc.

STAKEHOLDER ENGAGEMENTS 1-3

PATIENT, PARTICIPANT, CAREGIVER ENGAGEMENTS

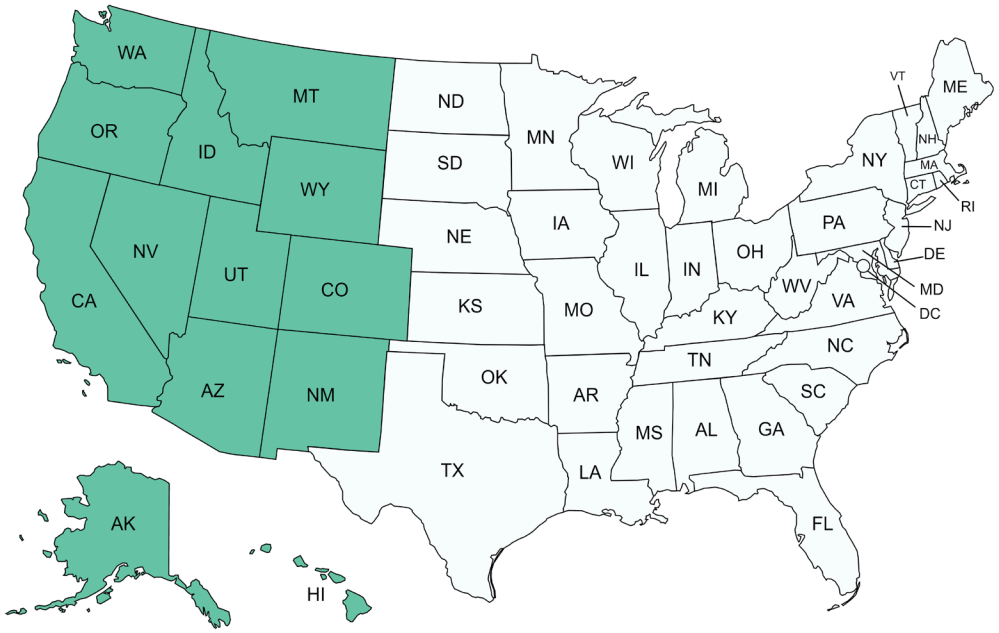


- **Introduction:** WG members present topics and types of questions at high level
- **Facilitated small group discussions via case studies on informing health-related questions through:**
  - Use of novel data from outside of the traditional healthcare system
  - Use of models and algorithms
  - Data linkage and aggregation of disparate datasets from multiple sources
- **Outcome:** Synthesis of discussions on stakeholder perspectives on the types of research questions and issues raised
- The same session format will be repeated three times in different regions and with different groups of people
- *Virtual options provided to maximize accessibility and participation of diverse stakeholders*

STAKEHOLDER ENGAGEMENT 4

SCHOLAR, EDUCATOR, TECHNOLOGY DEVELOPER EVENT

- **Introduction:** WG members present topics and types of questions at high level
- **Summary of main themes of engagements 1-3**
- **Panel discussions on informing health-related questions through:**
  - Use of novel data from outside of the traditional healthcare system
  - Use of models and algorithms
  - Data linkage and aggregation of disparate datasets from multiple sources
- **Outcome:** Synthesis of researcher/scholar/other expert opinions on how different groups perceive different types of data use and the impact on themselves, their family, and their communities
- *Virtual options provided to maximize accessibility and participation of diverse stakeholders*



TOWARDS A DRAFT REPORT

MEETING THE CHARGE AND SYNTHESIS

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:

A list of types of research questions of most relevance to the NIH over the next ~5-10 years → Questions Developed

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

Analysis of stakeholders' perceptions of the social and ethical issues raised by such research, how they weigh the inherent risks and benefits and make other value trade offs → Draft Stakeholder Engagement Plan Developed

3. Potentially: Recommendations for new policy/ethical frameworks that may be needed to face the tensions articulated in meeting the charge (e.g., high level guidance on what future policies addressing data, algorithms, and data integration must address)

NExTRAC DISCUSSION – 45 MINUTES