Developing an NIH Strategic Plan for Data Science

- Requested by Congress

- The plan focuses on:

  - Modernizing the data resource ecosystem to increase its utility for researchers and other stakeholders and to optimize its efficiency of operation
  
  - Enhancing data sharing, access, interoperability
  
  - Improving ability to use EHR, clinical, and observational data for research while ensuring data confidentiality
  
  - Modernizing infrastructure, increasing capacity
A Couple of Definitions...

“Data science is an interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data.”

FINDABLE
ACCESSIBLE
INTEROPERABLE
REUSABLE
<table>
<thead>
<tr>
<th>Domains of Data Science</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Infrastructure</td>
<td>Hardware, architecture, and platforms necessary to capture, organize, store, allow access to, and compute on data</td>
</tr>
<tr>
<td>Data Resources</td>
<td>Methods, practices, and associated features needed to increase the value and utility of data beyond its native state</td>
</tr>
<tr>
<td>Advanced Management, Analytics, and Visualization Tools</td>
<td>Algorithms, software, models, and tools necessary to extract knowledge and understanding from data</td>
</tr>
<tr>
<td>Workforce Development</td>
<td>Policies, practices, and programs to train and develop an outstanding data science workforce</td>
</tr>
<tr>
<td>Policy, Stewardship, and Sustainability</td>
<td>The policies and practices necessary for governance, financial management, and sustainable stewardship of the biomedical data science ecosystem</td>
</tr>
</tbody>
</table>
Organization of the Strategic Plan

I. Overarching Goals
   i. Strategic Objectives
      1. Implementation Tactics
         a. Milestones and Performance Measures
Overarching Goal 1: Support Highly Efficient and Effective Data Infrastructure for Biomedical Research

Strategic Objective 1-1: Optimize Data Storage, Access and Security

- Rely on private sector where possible

Strategic Objective 1-2: Connect NIH Data Systems

- Use NIH Data Commons and NLM/NCBI as hubs

https://commonfund.nih.gov/bd2k/commons
Overarching Goal 2: **Promote the Modernization of the Data Resources Ecosystem**

Strategic Objective 2-3: **Leverage Ongoing Initiatives to Better Integrate Clinical and Observational Data into Biomedical Data Science**

**Implementation Tactics (examples):**
- Create efficient linkages among NIH data resources that contain clinical and observational information.
- Develop and implement universal credentialing protocols and user-authorization systems to enforce a broad range of access and patient-consent policies across NIH data resources and platforms.
Overarching Goal 3: Support the Development and Dissemination of Advanced Data Management, Analytics, and Visualization Tools

Strategic Objective 3-1: Support Useful, Generalizable, and Accessible Tools and Workflows

Strategic Objective 3-2: Broaden Use of Specialized Tools

- Example: Algorithms from astronomy adapted for use in cellular imaging.
- Support research for improving methods for using EHRs and other clinical data.
Overarching Goal 4: **Enhance Workforce Development for Biomedical Data Science**

**Strategic Objective 4-1: Enhance the NIH Workforce**
- E.g., data science training and education for NIH staff

**Strategic Objective 4-2: Expand the National Research Workforce**
- Enhance quantitative and computational training for students and postdocs

**Strategic Objective 4-3: Engage a Broader Community**
- E.g., code-athons, bug-bounty programs, contests
Overarching Goal 5: Enact Appropriate Policies to Promote Stewardship and Sustainability

Strategic Objective 5-1: Develop Policies for a FAIR Data Ecosystem

Implementation Tactics (examples):

• Create rational and supportable data-sharing and data-management policies that ensure the security and confidentiality of patient and participant data and comply with applicable law.

• Promote development of community standards that support FAIR principles for data storage.
Next Steps

- The Strategic Plan was delivered to Congress in early May
- The implementation phase has already started and will be ramping up fast
  - Creating implementation teams to plan, execute, coordinate & monitor each implementation tactic or set of closely related tactics
  - Development of performance measures and milestones is key
- Recruitment of NIH Chief Data Strategist
Questions & Comments