Concept Clearance: New Common Fund Program

Harnessing Data Science for Health Discovery and Innovation in Africa

Objective: To explore whether advances in Data Science (DS) applied in the African context can spur new health discoveries and catalyze innovation in healthcare and health research on the continent

1. To advance DS research in Africa and demonstrate the feasibility of DS to improve health
2. To increase capacity to carry out DS research in Africa
3. To explore ELSI issues impacting DS for health from an African perspective and contribute to policy discussion on the continent
4. To catalyze and facilitate development of a trans-African network of Data Scientists; provide access to resources, information, and training opportunities; disseminate information; and serve as a technical resource
5. To catalyze new collaborations, gather stakeholders, and explore the state of the field for DS

Estimated Funds Available: $62.06M over 6 years

Program Duration: 6 years

Council Action: Vote on support of Data Science in Africa

commonfund.nih.gov
## Acknowledgements

### Common Fund Working Group and other NIH contributors

<table>
<thead>
<tr>
<th>FIC</th>
<th>NHLBI</th>
<th>NIEHS</th>
<th>NINDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flora Katz</td>
<td>George Mensah</td>
<td>Gwen Coleman</td>
<td>Stacey Chambers</td>
</tr>
<tr>
<td>Laura Povlich</td>
<td></td>
<td>John Balbus</td>
<td></td>
</tr>
<tr>
<td>Amit Mistry</td>
<td>Ishwar Chandramouli</td>
<td>Charles Schmitt</td>
<td></td>
</tr>
<tr>
<td>Nalini Anand</td>
<td>Darrell Hurt</td>
<td>Bonnie Joubert</td>
<td></td>
</tr>
<tr>
<td>Celia Wolfman</td>
<td>Gray Handley</td>
<td>Kim McCallister</td>
<td></td>
</tr>
<tr>
<td>Michael Cheetham</td>
<td>Chris Whalen</td>
<td>Mike Humble</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ty Lawson</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NCI</th>
<th>NIAID</th>
<th>NLM</th>
<th>NLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Pearlman</td>
<td>Ishwar Chandramouli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tom Gross</td>
<td>Darrell Hurt</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gray Handley</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chris Whalen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NHGRI</th>
<th>NIBIB</th>
<th>NIGMS</th>
<th>NINDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennifer Troyer</td>
<td>Tiffani Lash</td>
<td>Susan Gregurick</td>
<td>Stacey Chambers</td>
</tr>
<tr>
<td>Valentina di Francesco</td>
<td>Grace Peng</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ken Wiley</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NHLBI</th>
<th>NIBIB</th>
<th>NIMH</th>
<th>NLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>George Mensah</td>
<td>Tiffani Lash</td>
<td>Beverly Pringle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grace Peng</td>
<td>Andrea Horvath Marques</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NIGMS</th>
<th>NIMH</th>
<th>OD/OBSSR</th>
<th>OD/DPCPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susan Gregurick</td>
<td>Beverly Pringle</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Andrea Horvath Marques</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NIDDS</th>
<th>NLM</th>
<th>OD/OBSSR</th>
<th>OD/DPCPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stacey Chambers</td>
<td>Valerie Florence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elizabeth Ginexi</td>
<td></td>
</tr>
</tbody>
</table>

| OD/DBPCPSI       |                  |                  |                  |
|------------------|                  |                  |                  |
| Marie Nierras    |                  | Sonynka Ngosso   | Matt Perkins      |
|                  |                  |                  |                  |
Analytic capacity is not keeping up with the volumes of data being collected! Health data scientists needed, especially in LIC, LMIC to conduct policy relevant analysis & monitor progress @CDCgov @MoHFW_INDIA @ICMRDELHI @WHO @DrFrieden @JeremyFarrar

- We propose to develop and adapt techniques of DS in order to address some of the most compelling problems in global health with a focus on Africa
Opportunities in Africa

- Carries disproportionate share of the global burden of disease
- Unique populations, genes, and exposures
- Critical medical workforce shortage – need for clinical decision support
- Extensive mobile phone coverage – opportunities for leapfrog tech!
- Build on NIH investment: ~1600 grants, H3Africa ($170 M), MEPI/HEPI ($190 M)
- Lack of established data infrastructure creates opportunities to develop tools and applications that can be shared, adopted, and harmonized widely
NIH invests significantly in DS for the U.S., but few grants focus on Africa

Expertise in Africa exists but is scattered – Networking could be transformative!

Limitations in data infrastructure are rapidly improving

Ethical, Legal & Social Implications of DS research is key contextual factor
There are many potential partners that could make this effort especially powerful

**Potential NIH Partners interested in DS for Health**
- Wellcome Trust
- Gates Foundation
- World Bank/AfDB
- USAID
- African Academy of Sciences

**Potential Collaborators on grants**
- IBM Research
- Microsoft
- Google
- Safaricom
- African Institute for Mathematical Sciences
- Africa Open Science Platform
- National governments
Propose 5 Initiatives

1. Research Hubs focused on key health problems
2. Data Science Training Programs
3. Ethical, Legal and Social Implications of DS Research
4. Open Science Data Platform and Coordinating Center
5. Symposia (years 1 and 6)
Initiative 1: Six Research Hubs

Each hub will:

• Focus on a critical health research problem
• Bring together multi-disciplinary expertise
• Engage partners from government and private sector

Competencies

- Data science
- Biomedical/Public Health
- Technology & Innovation
- Community Engagement/ELSI

Academic
Academic
Academic
Government
Private sector
NGO
Initiative 1: *Illustrative Hub on Malaria*

**Potential Health Outcomes**
- Quicker diagnosis
- High-accuracy detection
- Better prediction
- Better targeting of public health measures
Initiative 2: Data Science Training

- Training grants to develop and implement new master’s and PhD curricula
- Strengthen faculty
- Train the next generation of data scientists
- Hubs provide some trainees and research projects
Initiative 3: Ethical Legal and Social Implications Of DS Research

• Research studies that address the spectrum of policy, security, and cultural barriers to data science in Africa
• Work with coordinating center to advise on policies and implement research findings
Initiative 4: Open Data Science Platform and Coordinating Center

Scientific Function
• Facilitate data and tool sharing
• Develop portals
• Lead trans-network projects
• Lead curriculum harmonization
• Develop/Deliver short eCourses

Administrative Function
• Facilitate WGs and network meetings
• Network Website
• Communications
Initiative 5: Symposia

Symposia on Data Science for Health Discovery and Innovation in Africa

Year 1
• Assess the state of the field
• Catalyze new collaborations

Year 6
• Assess the state of the field
• Writing projects
Deliverables at the end of 5 to 10 years

- **Recognized centers of excellence** in various fields of data science
- **Advances in policy** surrounding ethical issues around the use of data and data sharing
- **Products**, some of which may spin off into start-up companies
- **New interdisciplinary collaborations**

**Demonstration of feasibility** of DS to improve health in Africa

**Increased capacity** to advance African-appropriate tools and applications that will catalyze new areas of research

**A unique continental network** of scientists interested in advancing this area

**New scientific knowledge** that improves clinical practice & health
## Budget Timetable

<table>
<thead>
<tr>
<th>Init 1 – Research Hubs</th>
<th>Lead IC</th>
<th>FY20</th>
<th>FY21</th>
<th>FY22</th>
<th>FY23</th>
<th>FY24</th>
<th>FY25</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Hubs</td>
<td>FIC, NIBIB, NIMH</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>• Salary/RMS</td>
<td></td>
<td>.36</td>
<td>.73</td>
<td>.73</td>
<td>.73</td>
<td>.73</td>
<td>.73</td>
</tr>
<tr>
<td>Init 2 – DS Training Programs</td>
<td>FIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Training awards</td>
<td></td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Init 3 – Ethical, Legal, &amp; Social Implications of DS Research</td>
<td>TBA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Research Grants</td>
<td></td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Init 4– Open Data Science Platform and Coordinating Center</td>
<td>NIBIB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Center</td>
<td></td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Init 5- Symposia</td>
<td>FIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Symposia in Africa</td>
<td></td>
<td>.192</td>
<td></td>
<td></td>
<td>.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Salary/travel (RMS)</td>
<td></td>
<td>.073</td>
<td></td>
<td></td>
<td>.090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL: $62.06</td>
<td></td>
<td>.625</td>
<td>12.23</td>
<td>12.23</td>
<td>12.23</td>
<td>12.23</td>
<td>12.51</td>
</tr>
</tbody>
</table>

*For African Institutions only*
Examples

Cross-Cutting

- Viral surveillance, tracking, containment
- Precision Public Health
  - Pharmaceutical dosing and pharmacogenetics
  - Risk assessment and biomarkers
- Diagnostics
  - AI for cervical cancer detection on smartphone
  - Digital stethoscope: lung screening thru sound with machine learning
- Predict suicide risk and depression
  - Integrate and analyze data from EHRs, social media, phones sensors, voice recordings
- Exposome
  - Predict outbreaks and understand environmental health risks using data from satellites, geolocation tech, particulate monitors, EHRs