

I. COMPLETE TABLE 1

TABLE 1. Program Overview

Program Title	Harnessing Data Science for Health Discovery and Innovation in Africa
Program Goal:	<i>To explore whether advances in Data Science applied in the African context can spur new health discoveries and catalyze innovation in healthcare and health research on the continent.</i>
IC Director(s) or Senior IC Staff who provide general scientific leadership	<i>Roger Glass (FIC), Bruce Tromberg (NIBIB), Joshua Gordon (NIMH)</i>
Program Coordinator(s): IC staff who provide daily oversight and leadership for the program	<i>Flora Katz (FIC), Laura Povlich (FIC, currently H3Africa Lead for bioinformatics training), Margaret Lentz (FIC), Tiffani Lash (NIBIB), Beverly Pringle (NIMH) Planning working group has 28 members from 12 ICOs (including representatives of the H3Africa working group), who may also manage individual grants</i>
Total # of Proposed Initiatives and their names:	<i>5 initiatives: Research Hubs Data Science Training Programs Ethical, Legal, and Social Implications (ELSI) Research Studies Open Data Science Platform and Coordinating Center Symposia</i>
Total Common Fund \$\$ Requested for ALL Years (millions)	<i>\$62.06mil over 6 years</i>

II. DESCRIBE BACKGROUND INFORMATION

We are experiencing an explosion of data relevant to human health, from both biological and non-health sources, that currently exceeds our ability to capture and interpret. This gap applies to every field of biomedical and behavioral research. If we don't invest in the data itself, applying known and developing new Data Science (DS) approaches, we are wasting an incredible opportunity to improve Human Health. This set of initiatives addresses that gap for Global Health research and Public Health application in low resource settings.

In the next decade, rapid advances in DS, including new approaches to the description, collection, storage, integration, and analysis of large, heterogeneous, structured and unstructured data sets, and new computational methods such as advanced deep learning, digital phenotypes, machine/artificial intelligence, and 3D imaging are expected to **transform** biomedical and behavioral research and lead to improved health for individuals and populations. Traditional datasets (e.g. national health systems, surveillance, surveys) are becoming deeper and richer while new sources of data based on new technologies and sensors (e.g. social media, geospatial data, mobile phones, wearables, electronic medical records, bioimaging, and genomics) are emerging that may be of greatest value when linked to DS. Progress in development of huge new data sets and advanced methods for mining them underpin advances in diagnostics, technology development, and the potential for precision public health. This initiative examines whether advances in DS developed and/or applied in the African context can be used to spur discoveries and innovations that ultimately promote significant improvements in health for African individuals, communities, and populations. We define DS as "the 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" (NIH Strategic Plan for Data Science).

While advances in the United States and other high-income countries (HICs) are being actively explored and increasingly supported in both the public and private sectors, applications that are relevant, affordable, acceptable, and scalable in low- and middle-income countries (LMICs) are largely undeveloped. Moreover, in most cases applications and tools cannot simply be adapted from HICs for use in LMICs as they are generated from data collections biased towards European ancestry populations and may not apply to other genetic backgrounds (this is true, for example, for GWAS studies, clinical imaging data used as a basis for machine learning, and major disease biomarkers). However, we can leverage the momentum and knowledge generated in this field in HICs to stimulate new investment, knowledge generation and innovations for LMIC populations.

We focus this initiative in Africa for several reasons: First, despite recent progress, Africa carries a disproportionate share of the global burden of disease. DS has the potential to significantly impact both quantitative and qualitative research and health on the continent. Second, the African population is growing faster than other world regions and some African leaders are eager to transition to knowledge-based economies. Extensive mobile phone coverage in Africa has led to major innovations in banking, agriculture, and other sectors and has the potential to "leap frog" in health care delivery systems, bringing the clinic to the patient through Point of Care (POC) technologies and self-management systems, with applications to rural and underserved populations worldwide. Third, this initiative would be **synergistic** with and leverage the substantial investments NIH has already made in research and research training in

Africa. NIH funds more than 1,600 active grants working in Africa and has catalyzed a well-developed clinical and academic research network through prior investments by the CF and its partners in the Medical- and Health-professional Education Partnership Initiatives (MEPI and HEPI, \$194 mil). **Importantly, this initiative would provide opportunities to access, build on, and translate the resources and expertise developed through H3Africa (\$170 mil by CF and its partners) into products and policies impacting health.** Finally, the lack of an established data infrastructure in most African countries provides a rare opportunity to create tools and applications that can be shared, adopted, and harmonized widely [e.g., OpenMRS, an innovative open-source electronic medical record (EMR) that grew out of a modest FIC investment to the AMPATH program in Western, rural Kenya, is now under consideration as a national EMR in 12 countries and is used in many African countries in support of PEPFAR programs].

External feedback was obtained through individual (18) and group (6) consultations around the topic of DS in Africa with extramural US and African DS experts, and working group members attended 8 relevant meetings, all of which contributed to the design of the current proposal. The main conclusions from these consultations, NIH portfolio analysis, and our landscape analyses are the following:

- **Expertise exists but it is scattered and needs a coordinated effort to catalyze a new field of health DS in Africa adapted to African populations:** Many of the enabling factors for a catalytic impact are already in place, including government support, budding private sector engagement, and increasing interest from academic institutions. There are a small number of DS efforts in Africa that are specifically relevant to health, and there are others that are more generally related to DS, bioinformatics, AI, or other related fields. However, there is not a coordinated effort across the continent focused on health. As such, a CF effort, with support of key partners across academic, government, and private sectors, could have a transformative impact on advancing the use of DS for health discovery and innovation.
- **These initiatives would provide an opportunity to build on the network and expertise established through H3Africa:** H3Africa investigators, infrastructure, and coordinating platforms will be competitive assets in applications for the new initiatives and enrich the network. **Recommendations from the H3Africa working group have been incorporated in this proposal.**
- **There are many potential partners that could make this effort especially powerful:** The **African Academy of Sciences** demonstrated strong interest through a recent Grand Challenges Africa initiative on *Data Science Approaches to Improve Maternal and Child Health in Africa*, in partnership with the **Bill & Melinda Gates Foundation**. The **Wellcome Trust** recently announced a planning phase for a new program around health data innovation, including potential hubs in South and East Africa, committing £75 million/\$94.5M over five years. Their timeline may well coincide with this CF effort. Private sector corporations are investing in DS for Health in Africa, including **IBM Research**, which supports two offices on the continent, **Google**, which recently launched an AI lab in West Africa, and **Microsoft**, which is engaged in health-related research in Kenya. **Safaricom's** innovation lab is involved in health research using their data and networks and has a platform through which researchers can submit ideas for partnerships.
- **Infrastructure and capacity are rapidly improving:** Internet and broadband access, computing capacity, stable electricity, and expertise are limited in many parts of Africa, but they are rapidly improving and there are ongoing efforts focused on strengthening capacity. For example, **Microsoft** recently launched enterprise-grade data centers in South Africa, becoming the first global provider to deliver cloud services from datacenters on the continent. The **World Bank's** African Centers of Excellence program and a growing number of universities are offering degree programs in DS, bioinformatics, and bioengineering. There has been an increase in more robust cable connections to major African cities and universities, extension of National Research and Education networks (NRENS), increasing satellite connections, and investments in infrastructure through H3Africa and the African Centers of Excellence in Bioinformatics and Data Intensive Sciences (NIAID/FNIH). Many countries are developing national strategies for eHealth/Health ICT reflecting an evolving infrastructure and high level commitment to developing this sector (e.g., Nigeria, Uganda and Rwanda). Moreover, an alternative infrastructure for data exchange is already in place: mobile phones, with 4G networks in some countries like Rwanda and Kenya and 5G networks anticipated. The Ethical, Legal, and Social Implications (ELSI, borrowing the term from genomics) of DS research are potential barriers to the advancement of DS in Africa and a key contextual factor to be addressed by the CF program. Policy challenges in security, privacy, open-source code and sharing of data that are of global concern will be addressed through initiatives 1 and 4, below.
- **NIH and USG landscape:** While NIH broadly is investing heavily in research related to DS, very little of it is directed towards Africa and only a small percentage involves African investigators. The following **NIH ICOs** have been a part of the working group developing this proposal, including staff involved in DS and global health: FIC, NIBIB, NIMH, NHLBI, NICHD, NHGRI, NCI, NINDS, NIEHS, NIAID, OBSSR, and NLM. With support from NIH and other USG agencies, biomedical engineering hubs in the US are reaching out to partners in Africa and may provide expertise for the Hubs. **USAID** has funded programs related to DS and informatics in Africa and has demonstrated growing interest in data innovation and artificial intelligence for global health and other sectors. The **CDC** in Africa has invested in EMRs that might eventually be used as research platforms.

Why is a trans-NIH strategy needed to achieve these goals? The unique contribution of this CF initiative is to link the relatively modest and uncoordinated investments in DS across the African continent to create a powerful network that will support the emergence of this new field. This was one of the major outcomes of the MEPI CF programs: leading medical schools, which had been isolated, formed a community over the course of the 10-year program that strengthened all schools and their partners and now is open to all institutions in sub-Saharan Africa. Many resources developed by the DS network will be public and thereby create a public good. The core developments

in DS can be applied to all fields of research at NIH, so the hubs will serve **cross-cutting** interests across NIH ICs. No single IC can take on an initiative of this variety and research breadth and build a continental network, possibly the first of its kind in the world, so this will not happen without CF support. However, once the Centers are established, the ICs can pick up projects at the end of CF support in their individual fields. As cooperative agreements, the consortia will also benefit from the diverse expertise provided by NIH ICO staff.

III. STATE GOALS

The **Overarching Goal** of the program is to advance DS and related innovations in Africa to create an ecosystem that can begin to provide local solutions to countries' most immediate public health problems through advances in research. **Deliverables:** We propose to do this by creating a **unique** continental network that would, in partnership with the international academic community, governments, the private sector, and other funders, develop new data collection and analytic systems, applications, and tools with attention to usages that are population-relevant, affordable, acceptable, and scalable. In addition, we will enhance the capacity to carry out DS research by training a new generation of interdisciplinary African scientists in DS. At the end of 5 to 10 years, we expect there will be recognized centers of excellence in various fields of DS that can be regional and continental resources; 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; considerable increased capacity to advance African-appropriate tools and applications that will **catalyze** new areas of research; demonstration of the feasibility of advanced DS to improve health in Africa; new interdisciplinary collaborations; and new scientific knowledge. We also anticipate the emergence of trans-network data repositories, such as selected medical image repositories to be used as a basis for machine learning, which could serve as a shared resource for scientists across the globe and which would be coordinated through the Open Data Science Platform. **Milestones** will be the aggregate milestones of the individual initiatives. NIH has experience building broad networks in Africa that have been functional and are perceived as transformational (e.g. the MEPI network, now embodied in AFREhealth, and the H3Africa consortium). While we anticipate requesting a second five years from CF for this initiative, we feel Stage 2 should be shaped by what we learn in the first five years in this rapidly changing field. We limit our discussion here to Stage 1. Target completion dates indicated in parentheses.

We propose 5 initiatives:

Initiative 1. Research Hubs: Goals: To advance DS research in Africa and demonstrate the feasibility of DS to improve health.

Deliverables: Each Consortium will focus on a critical health research area to be addressed through DS approaches and include at least two partners outside the applicant institution. Through the FOA (and its review criteria), applicants will be highly encouraged to engage African governmental and private partners (industrial and NGOs) as collaborators, in addition to African and US academic partners, and it is expected that these partnerships will expand over the course of the award. Local or national governmental stakeholders should have a defined role in the project. Each Hub will include several interdependent component sub-programs or demonstrated competencies, as appropriate to the project proposed, including DS, Biomedical Domain Knowledge, Technology and Innovation, and Community Engagement/ELSI. Applicants should propose at least two research projects. One research project must be solution-oriented through the development or adaptation of a product (software or hardware) or new technology that is either needed for the collection of new data types or translates initial DS research into a new health innovation. While new data may be collected, it cannot be the exclusive focus of the research project and projects should incorporate re-use of data as much as possible. A third arm of each Hub should include locally-determined pilot grants that will be shorter term and vary throughout the period of award, allowing the grantees to take advantage of new directions stimulated by their projects and to provide support to younger investigators to undertake DS-related research. These would undergo peer review within the Hub. Hubs would also participate in cross-network technical working groups (TWGS), with topics suggested by the investigators. The standard of FAIR (findable, accessible, interoperable, and reusable) should be a goal of the projects, with a focus on data science methods to gather, verify, integrate, curate, store, use and protect health data that works in typical African settings.

Milestones (unless noted, all milestones will be evaluated annually; as cooperative agreements, NIH could also negotiate grant-specific milestones after award): Engagement and/or strengthening of partnerships outside the institution, particularly including governmental and industrial collaborations; research products including publications, software, technology, patents, and others; engagement of key stakeholders, from Ministries to communities; participation in trans-network working groups, workshops, and other activities; engagement of interdisciplinary teams, e.g., through the participation of Schools or Departments that do not normally interact; contribution to infrastructure within Hub institutions to support and sustain DS research; demonstration of health impact through DS-driven research products (completion date 2025).

There are many areas of DS that may be engaged by the Hubs and most projects would be expected to employ a variety of approaches (see above). We envision project proposals in every ICO area, such as:

- genomics (e.g., pharmacogenetics, biomarkers, and development of polygenic risk scores for defined African populations);
- viral surveillance with field deployable diagnostics for multiple viruses simultaneously;
- computational and voice assessment of depression;
- development of shared platforms and potential repurposing of EMRs for harmonizing clinical data, digital clinical trial recruitment, and discovery of emergent multi-morbid patterns as well as longitudinal risk, such as childhood exposure and later pathology;

- image-, sound-, and genome-based development of POC diagnostics or mobile medical decision support (particularly relevant in Africa where most diagnoses are made by non-specialists and even non-physicians) for pre-cancer and cancer, pulmonary disease, infectious disease, dermatological diagnosis, and mental health among others;
- integration of sensor technologies with satellite data, demographics, exposure biomarkers, and other data for exposure assessment and intervention;
- modelling applications to predict disease outbreaks, pathogen exposure, infection dynamics, and susceptibility, along with modelling of markers for Precision Public Health (use of Big Data to predict public health risk and customize treatments for more specific and homogeneous subpopulations, focusing resources where there is the most urgent need), among many others.

We expect each Hub to be interdisciplinary (biomedical science, computer science, engineering, social and behavioral science, and other fields), including diverse faculties from Hub institutions, and to include multiple institutions from Africa, the United States and/or globally that can partner to provide synergistic expertise, collaborate on research, and support DS linked to new technologies. The grants would be direct awards to African institutions. We anticipate six awards from CF.

Initiative 2. DS Training Programs: Goal: Increase capacity to carry out DS research in Africa. **Deliverables:** Train a cohort of investigators in DS that will have the skills to become independent investigators, research leaders, and research collaborators and contribute to DS research in Africa. Strengthen faculty at participating institutions to teach in this field. Develop, accredit and implement new curricula. Develop and deliver short courses to reach a wider community at the host and partnering institutions. Participate in a network working group on DS training, which might cover topics such as core competencies, curriculum sharing, and supporting Hubs. **Milestones (all will be evaluated annually against targets set in the applications):** Number, demographics, prior disciplines, and degrees obtained of students trained; integration of trainee projects with funded research, including consortia programs; research outputs of students, including publications, software, technology, patents, and others; number of faculty-enhancement activities and number of faculty participating; number and variety of new courses created and number institutionalized (which would contribute to sustainability), both full courses and short courses. Participation in trans-network activities such as working groups. The applicants may be either African institutions or US institutions in collaboration with African institutions. All trainees must be African. We anticipate five awards.

Initiative 3. Ethical, Legal, and Social Implications (ELSI) of DS Research: Goal: To explore ELSI issues impacting DS for health research from an African perspective and contribute to the policy discussion on the continent. **Deliverables:** Carry out research on ELSI issues from a broad or comparative African perspective through one or more research projects central to the conduct of DS research and its applications to public health in Africa. Direct African awards with collaborators from Africa or the United States. **Milestones:** Research products (publications, white papers, presentations); number of events involving engagement with stakeholders, from Ministers to the community; participation in and contributions to trans-network working groups and activities. We anticipate four awards.

Initiative 4. Open Data Science Platform and Coordinating Center: Scientific functions: Goals: To catalyze and facilitate the development of a trans-African network of Data Scientists; to advance the field through providing access to resources, information, and training opportunities that will serve as global public goods; to disseminate information about the network; to serve as a technical resource for the initiatives. **Deliverables:** Will convene and provide scientific leadership to the DS working group(s), made up of representatives from each Hub; facilitate and implement cross-network projects (such as a common image repository for selected diseases, which would require coordinated data standards and diagnostic criteria, and likely uniform data collection instruments, but would be an enormous resource for AI development in Africa); develop avenues for access to cloud computing environments; develop and deliver DS short courses through distance learning available to the entire network and archive courses for non-synchronous access; host hackathons and/or special workshops associated with network meetings, including training in team science; develop a web portal for relevant publicly available data resources and databases to make data discoverable and accessible (for example, databases and tools from USG resources such as NIH/NLM, NOAA, and NASA; tools such as IHME; existing African and other repositories through re3data.org; and eventually data generated through the network); provide and continually update a web portal for training and grant opportunities in DS available to African investigators and students; actively seek new partners for the network; and serve as a helpdesk. There would be great value in developing mutually agreed upon standardized data elements and sets that could be integrated and mined using data science techniques. The Center should work towards developing a “knowledge network”, a means of linking currently non-interoperable and siloed datasets to allow cross-dataset analyses, data re-use, and emergent insights. This would be an explicit goal of a Stage 2 program. [see [NSTC Workshop Report](#)]. Work with the DS training programs to establish core competencies and harmonize curricula. Act as mediator between outside groups and network for access to network resources. **Milestones (evaluated annually unless indicated):** leadership of DS working group, meetings convened and topics discussed; initiation of one or more trans-network DS projects (completion data 2025); development of a portal to make resources (scientific tools and databases, training and grant opportunities) findable and accessible and number of resources offered and added annually; number of online short courses of relevance to the network developed and delivered; number of workshops/hackathons delivered; partnerships initiated; technical assistance provided to other members of the network.

Administrative functions: Goals: Coordinate and facilitate administrative and logistical functions for the network. **Deliverables:** organize two annual network meetings each year to be held in Africa at rotating locations (thereby also allowing site visits for most lead network

institutions over five years), maintaining a website on network related news and accomplishments, facilitating training exchanges among participating members of the network, and facilitating logistics of meetings for all working groups (DS, training, ELSI, and others).

Milestones (annual unless indicated): 2 Network meetings per year organized and executed (includes travel for advisory committee); website developed and updated; # training exchanges facilitated; number of working groups facilitated. Having both scientific and administrative functions in one entity may be more cost effective and help with coordination of all tasks. We anticipate 1 award.

Initiative 5. Symposia and writing project: Goals: Catalyze new collaborations, gather stakeholders, and explore the State of the Field for DS in Africa at the beginning and end of Stage 1. **Deliverables:** Two Symposia and a supplement publication on DS in Africa. **Milestones:** Symposium on the State of the Field of DS in Africa, to be held in Africa. This would bring together the scientific communities required (biomedical, computational, engineering, social science and others), representatives of government and multilateral organizations such as African Academy of Sciences, private sector organizations exploring DS in Africa, community representatives, and other major funders. This would also serve to establish a baseline against which the progress of the CF program overall might be measured (completion date: 2020). Symposium and writing project on the state of the art of DS in Africa, to be held in Africa, at the end of year six (completion date: 2025). This can serve as a measure of the rate of change and accomplishments of the field over the course of the first five years and an opportunity to disseminate this information widely. The Center for Global Health Studies at the Fogarty International Center, with ICO CF partners, would undertake this and has significant experience carrying out writing projects with grantees resulting in supplement publications in journals such as Nature, Lancet, and BMJ.

IV. DESCRIBE PROGRAM MANAGEMENT

The proposed lead ICs, budgets, and timeframe are given in Table 2. Overall coordination of the initiatives will be through the program staff of the lead ICs (FIC, NIBIB, and NIMH), in consultation with the participating ICO working group (currently 12 ICOs). Within each initiative, individual grants may be managed through any of the implementing group HSAs, consistent with the topic and available expertise and according to the terms of the award. Collectively, the managers for all the grants in each initiative will meet on a regular basis as members of a steering committee for that initiative. Scientific officers for the award will be assigned from NIH extramural or intramural staff who have expertise in the appropriate area. Both the administrative and scientific leads will be expected to attend site visits and network meetings. The Symposia will be developed and managed by the FIC Division of Policy and Planning through the Center for Global Health Studies, in partnership with the participating ICOs. The implementing ICO working group will meet at least twice a year to review progress and make suggestions. An External Advisory Group will be convened by the NIH partners in Year 1 and will be expected to attend annual network meetings and provide an annual written report on recommendations to the NIH. They will also deliver a final report in year 6 to help inform a Stage 2 proposal. Applicants will be restricted to extramural investigators, although intramural investigators may be brought into the research teams. Initiatives 1, 3, and 4 must be led by African institutions. Initiative 2 can be led by either US or African institutions.

V. PROVIDE BUDGET (\$\$/year in millions)

TABLE 2. F&A to foreign institutions or for training grants to US institutions is fixed at 8% and included in these figures. In FY20 we will write and publish the FOAs, hire any additional personnel needed to lead these efforts, hold a Symposium in Africa, and engage with potential funding partners to strengthen the network. The FOA should be published at least 4 months before the receipt date to get quality applications.

	Lead IC	FY20	FY21	FY22	FY23	FY24	FY25
Init 1 – Research Hubs							
• Hubs	FIC, NIBIB, NIMH		6.0	6.0	6.0	6.0	6.0
• Salary/Travel		.36	.73	.73	.73	.73	.73
Init 2 – DS Training Programs							
• Training awards	FIC		2.0	2.0	2.0	2.0	2.0
Init 3 – ELSI Research grants							
• Research grants	TBA		2.0	2.0	2.0	2.0	2.0
Init 4 – Open Data Science Platform and Coordinating Center							
• Center	NIBIB		1.5	1.5	1.5	1.5	1.5
Init 5- Symposia							
• Symposia in Africa	FIC	.192					.192
• Salary/travel (RMS)		.073					.090
TOTAL: \$62.06 mil		.625	12.23	12.23	12.23	12.23	12.51