Data Science

Is a broad field that refers to the collective processes, theories, concepts, tools and technologies that enable the review, analysis and extraction of valuable knowledge and information from raw data. It is geared toward helping individuals and organizations make better decisions from stored, consumed and managed data.

Why is this important?

Locating Mirrors or subsets in NRENs will make data science possible in regions with high rates of infection and emerging pathogens.

Two of the most significant obstacles facing the training and development of data scientists and bioinformaticians in low- and middle-income countries such as those in Africa is the bandwidth and reliability of internet access. The recent movement to expand an opportunity to provide access to these essential tools of education and research. The National Institute of Allergy and Infectious Diseases at the NIH is establishing a public-private partnership with private industry, the Research and Education Network of Uganda (RENU), Makerere University and the Infectious Diseases Institute of Uganda to build the second African Center of Excellence in Bioinformatics in Kampala, Uganda. RENU has built a 1 Gbit backbone that connects many of the R&E institutions in Uganda. But internet access is still a bottleneck. The ACE partnership and center will provide reference databases and compute infrastructure across the RENU backbone without needing to use internet gateways. The combination of local infrastructure, local connectivity, and local services for data science will improve the educational and analytical capacity of researchers in Uganda and, through the regional NREN UbuntuNet, across East Africa.

NIH
- Conceptualize the ACE program
- Data center setup and maintenance

Makerere/IDI
- Provide Physical space
- Mentorship and training to students and researchers

ACE

Private industry
- Donate equipment Juniper, HPC equipment, Storage

RENU
- Fiber Connectivity
- Federated Identity and Access Management

The African Centers for Excellence in Bioinformatics demonstrate a way to provision regional research collaborations through the NREN infrastructure.

Why is this important?

Data Science Requires Access to Reference Databases

Size can be a problem

Current Access is limited by distance and bandwidth