HEALTHCARE & LIFE SCIENCES (HCLS) WORKING MEETING

HCLS Requirements

Michael Sullivan, M.D.

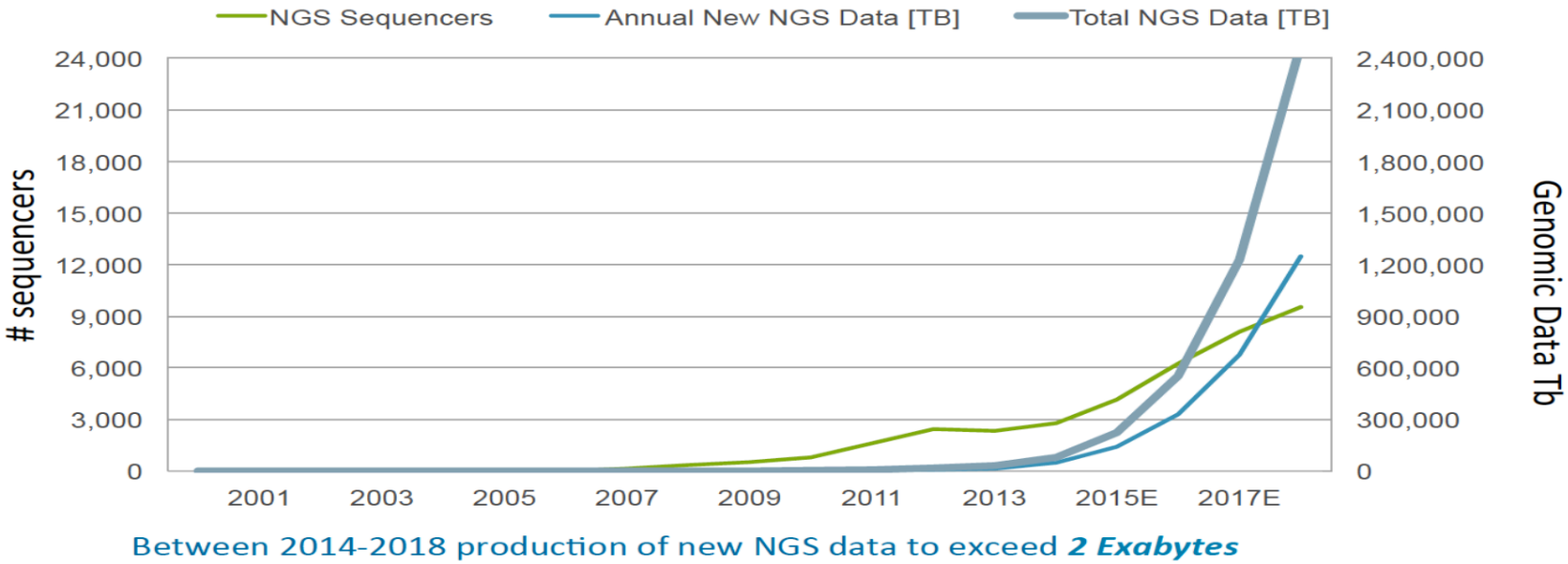
msullivan@Internet2.edu

HCLS Requirements

CONTENTS

- Big Data Driving Requirements
- NCI Strategic Initiatives
- Cancer Moonshot Requirements
- ITCR Funding for IT Development
- Networking Needs Based on SME Interviews
- Strategies

Big Data Driving Requirements



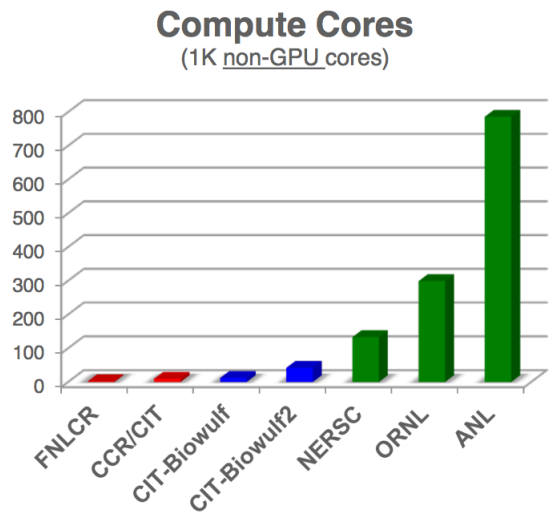
NGS: Next Generation Sequencing
NGS sequencers include machines from Illumina, Life Technologies, and Pacific Biosciences. Human genome data based on estimates of whole human genomes sequenced
Sources: Financial reports of Illumina, Life Technologies, Pacific Biosciences; revenue guidances; JP Morgan; The Economist; Seven Bridges Analysis.

NCI Strategic Initiatives

- NCI-DOE Pilots
- Precision Medicine Initiative
- Cancer Moonshot
- Genomic Data Commons
- Cloud Pilots
- Information Technology for Cancer Research (ITCR)

Compute Cores for NCI-DOE Pilots

US Department of Energy – Leaders in Computing

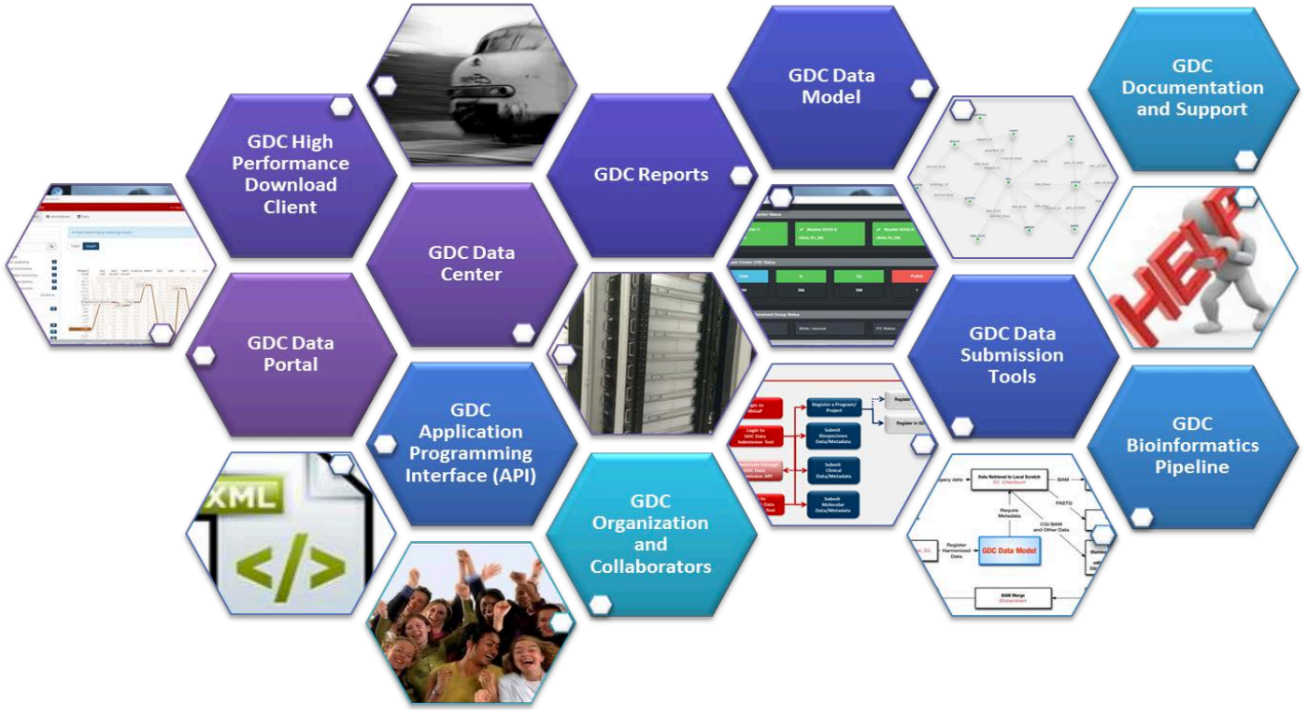


US Department of Energy

- Extreme scale systems
- Network Innovations
- $478M in FY14 into advanced computing research
- Lead for Exascale Computing Initiative

NIH NATIONAL CANCER INSTITUTE

NCI Cancer Genomic Data Commons (GDC)



Cancer Moonshot – Online Request for Ideas

Cancer Moonshot – Enhanced Data Sharing

- Tag resources with RRIDs.
- Mandate timely sharing of all data.
- Set standards for sharing diverse data types.
- Create a repository with data from ALL cancer patients.
- Remove journal paywalls blocking access to published research.
- Develop and deploy open source platforms and tools for researchers.*

* IT development is currently funded by NCI-CBIIT-ITCR.

NCI - ITCR Funding Vehicles for IT Development

- **Algorithm Development (R21)**
- **Prototyping and Hardening (U01)**
- **Enhancement and Dissemination (U24)**
- **Sustainment (U24)**

NCI - ITCR Funding for IT Development – Part I

- data management and analysis
- automation in experiment design and execution
- automation in data collection
- data processing and analysis
- data quality assessment
- data integration
- data presentation and visualization
- text mining and natural language processing
- data compression, storage, organization, and transmission

NCI - ITCR Funding for IT Development – Part II

- Establishing data exchange formats, ontologies, common data elements
- Improving software interoperability and compatibility
- Adapting computational tools for translational, epidemiological, and clinical applications
- Patient-centric laboratory and clinical data coalescence
- Computer-assisted interpretation of experimental results
- Environment for interactive modeling and simulation
- Platform for research collaboration
- Technology for performance evaluation of software tools, algorithms, and data collection methods
- Computational tools for interdisciplinary research training

Networking Needs Based on SME Interviews

Description	Area
• Performance monitoring across a collaborative community.	Monitoring
• Easier configuration of federated IdM and authentication.	Authentication
• Guidelines for attribute sharing for federated IdM.	Authentication
• Convenient GUI for high-performance data transport.	Data Transport
• An off-the-shelf integrated data transport platform.	Data Transport
• Resources to help comply with HIPAA and FISMA.	Security
• Science DMZ design patterns for sensitive data.	Security
• Reliable high-speed connectivity to public cloud resources.	Cloud Computing
• Limited or no public cloud data egress fees.	Cloud Computing
• Access to archival storage in public or private clouds.	Cloud Computing
• Resources to help with IT training for researchers.	Training

Question:

How could biomedical researchers and networking experts come together and work collaboratively to address the many challenges that have been identified?