SBE CYBERINFRASTRUCTURE NEEDS AND REQUIREMENTS:

CURRENT STATUS AND A LOOK FORWARD

SBE ADVISORY COMMITTEE

OCTOBER 30, 2014

CIF21 INVESTMENT - SBE

SBE investment in cyberinfrastructure is substantial, but proportionately smaller than that of other directorates with larger budgets

CIF21 Funding by Directorate

(Dollars in Millions)

Directorate/Office	FY 2013 Actual	FY 2014 Estimate	FY 2015 Request
Biological Sciences	$3.75	$6.16	$3.75
Computer and Information Science and Engineering	57.03	85.00	80.00
Education and Human Resources	-	-	2.50
Engineering	7.00	12.00	10.00
Geosciences	10.25	15.25	11.00
Mathematical and Physical Sciences	25.60	21.00	11.50
Social, Behavioral, and Economic Sciences	4.50	6.00	6.00
International and Integrative Activities	1.00	-	-
Total	**$109.13**	**$145.41**	**$124.75**

Totals may not add due to rounding.

From FY 2015 Budget Request

CIF21 PORTFOLIO: THREE INVESTMENT FOCI (FROM FOUNDATION-WIDE CIF21 LEADERSHIP GROUP)

- Community Building
- Data-Intensive Science
- Computational Models, Methods, and Algorithms

SBE PROGRAMMATIC INVESTMENTS IN CYBERINFRASTUCTURE PROGRAMS

- **Building Community and Capacity for Data Intensive Research in SBE and EHR (BCC)**
  - Successor to BCC, forthcoming soon
- **Critical Techniques and Technologies for Advancing Big Data Science and Engineering (BIGDATA)**
- **Data Infrastructure Building Blocks (DIBBs)**
- **Others:**
  - Meta data for Long-standing Large-Scale Social Science Surveys (META-SSS)
  - Software Infrastructure for Sustained Innovation (SI^2)
  - Integrative Graduate Education and Research Traineeship Program – CIF 21 Track (IGERT-CIF21); National Science Foundation Research Traineeship (NRT) Program

LOOKING FORWARD

Types of data and consequentially types of cyberinfrastructure challenges somewhat different for SBE science than for other sciences

Historically, "big data" in much of SBE tended to be *survey data* and *administrative data*

Both surveys and administrative data are what SBE AC member Bob Groves has described as "designed data"

LOOKING FORWARD

The "Big 3" national surveys -- GSS, PSID, ANES

- The Big 3 are not very big in comparison with some data sets from other sciences
  - ANES 70,000-80,000 respondents
  - GSS data from fewer than 60,000 respondents (although 5,500 variables)
  - PSID 70,000 individuals and 9000 families
- But, PSID goes back to 1968, GSS goes back to 1972, ANES back to 1977 (or, 1948)
- So substantial data and meta-data management issues

LOOKING FORWARD

Administrative data bases, primarily government statistics from the federal statistical agencies or state agencies

- Economic, demographic, geographic and spatial science research using Census, BLS, school systems data, and so on
- Administrative data often provide bigger data sets (sometimes much bigger) than survey data, but do not routinely stretch capabilities of present-day state-of-the-art cyber tools

LOOKING FORWARD

A big, on-going change in SBE is the rapidly growing importance of big data sets that are organic (not designed), as Bob Groves describes it

- Financial data, social media, digitized text, transactional data, records of all sorts of computer mediated behaviors -- think Google, Amazon, eBay
- A lot going on in that area, but much is being done privately, not by NSF funded researchers
- Work with such data sets will require advances in cyberinfrastructure, as well as analytical methods

LOOKING FORWARD

Understanding the Brain

- Amount of data produced by modern-day fMRI research already places extraordinary demands on analysis and cyberinfrastructure capabilities
- The BRAIN Initiative is sure to lead to even greater requirements for cyberinfrastructure

Intersection of genetics and the SBE sciences

- Growing importance of genetic data in Anthropology; Psychology and Cognitive Sciences; Forensics, and other SBE sciences

LOOKING FORWARD

Other areas where there will be increasing demands for high-performance cyberinfrastructure:

- Network analyses
- Simulation, agent-based and other forms of computational behavioral and social science
- Visualization
- Analyses of texts and communications
- Linguistics, translation, textual analyses
- Cyber-GIS and Spatial Modeling
- Digitized video and audio databases
- Object-based databases (archaeological artifacts)
- More sophisticated analytical, management and statistical capabilities to deal with these data

CONCLUDING OBSERVATION

SBE science, methods, and data are placing ever increasing demands on cyberinfrastructure

At the same time, many SBE communities are still working to take advantage of current CI capabilities

- This was rationale for BCC:
  Historic lack of access and opportunity to learn about and take advantage of capabilities of high performance computing and cyberinfrastructure

Next few years

- an exciting, dynamic period of great change
- increasing needs and demands on cyberinfrastructure
- unexpected challenges, opportunities, developments, innovations