

4b – Research Cyberinfrastructure Networking

Action Item Template Response

General Action Item Information

Lead Division/Office: Research Technologies

Action Item Number: 4b

Action Item Short Name: Research Cyberinfrastructure Networking

Dependencies with other EP Action Items:

Implementation leader (name & email): Dave Hancock (dyhancoc@indiana.edu), Matt Link (mrlink@indiana.edu)

I. DESCRIBE YOUR PLANS FOR IMPLEMENTING THIS ACTION.

Optimizing the functionality of IU's cyberinfrastructure and meeting standards for our continued participation in FutureGrid, PolarGrid, TeraGrid, and beyond, can occur only with a networking infrastructure where, as a rule, each individual research computational or storage node is connected directly to a research network switch, which is in turn connected to the corresponding national (or local) network. This means that as our computational and storage resources expand, so must the research network switch infrastructure that provides our connections to these national networks. In addition, the paired research network switches in the ICTC and the Data Center cyberinfrastructure facilities have become "machine room research backplanes" for the IU advanced research infrastructure. These switches are tied together via the I-Light network, and this overall network infrastructure is the basis for the entire IU research cyberinfrastructure and our disaster resilience strategy for this infrastructure.

Additionally, it is critical to address the ever-increasing high-speed, high-bandwidth needs of our local research community. We continue to see our local researchers saturate the standard, typical campus network connections between their labs and the data centers. This often times severely impacts their ability to conduct research in an efficient manner. We have worked over time to address these individual situations but they have been too few to address the real needs of our largest and most active researchers. We need to work to target researchers across campuses to provide dedicated networking (10Gbps) from their instruments to the data centers.

The term *data deluge* provides a sense of scale to the digital and often automated production of data. However, it does not make clear the complexity and variety of data sources. Well known are the few major instruments that are expected to produce more than a petabyte (PB) per year, such as the Large Hadron Collider and the Large Synoptic Survey Telescope. But even smaller locally-based consortia are creating relatively prolific instruments, such as the WIYN observatory One Degree Imager (500 GB - 4 TB/night), and ice sheet sensing equipment being used in the polar regions (50 TB/year). Instruments such as 454 gene sequencers (3 TB/year), advanced microscopes (2 TB/year), and digital video (94 TB/year for one HD video) are becoming increasingly common in single-investigator labs. The data deluge, when examined in detail, consists of a variety of types of data flows ranging from torrents to trickles, originating from numerous and diverse sources. The data analysis challenges of large-scale instruments are clear. Lack of network capability may make it impossible to move data from more modest instruments

directly to large-scale storage, analysis, and visualization facilities that would allow for the most effective and timely conversion of the data into information and knowledge. Even in circumstances in which data can be effectively managed and analyzed, it remains extremely difficult to capture sufficient metadata to allow for effective reuse of important data later.

II. WHAT ARE THE POLICY AND PRACTICE IMPLICATIONS OF YOUR PLANS

None.

III. IDENTIFY STAKEHOLDERS.

- Research community
 - Faculty
 - Graduate students
 - Laboratories
 - National and international research collaborations
- Creative activities communities
- Outreach and community engagement