Internet2 Portal Project Case Study: Visualizing Big Data Transport in Biomedical Collaborations

This past winter, Internet2 kicked off a web portal initiative to help members see how they are using Internet2 services. The first version of the portal provides a visualization of Internet2 Network traffic—data rates, volumes, and performance measures. In the future, it will also display the use of services, for example, the number of certificates obtained from Internet2’s InCommon Certificates service. The intention is to create a single web portal where members can see their individual or institutional use of Internet2 services in real time. The actual network collaboration described below illustrates the portal’s initial capabilities for network traffic visualization.

The National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) hosts biomedical research data. One major research focus is to identify common variations in the human genome. The 1000 Genomes Project is tackling this challenge by sequencing and comparing the genomes of thousands of people with varying ethnicities. Other research has the goal of improving our ability to diagnose, treat and prevent cancer. The Cancer Genome Atlas (TCGA) is recording the genetic changes found in different kinds of cancer and making that data freely available. Researchers who accept NIH awards agree to make their results public and to upload them to NCBI databases. NCBI assigns unique identifiers and stores the results in genomic databases that contain hundreds of terabytes of sequence data. Other researchers can then search and download known genomic sequences that enable them to analyze new results. The ratio of outbound to inbound data at NCBI is roughly 10 to 1, underscoring the wide use of NCBI data and its importance in genomic research collaborations.

NCBI, European Bioinformatics Institute (EBI), and the DNA Data Bank of Japan (DDBJ) make up the International Nucleotide Sequence Database Collaboration (INSDC). The three INSDC members collaborate to make sure each genomic sequence is assigned a globally unique identifier. They also synchronize data with one another, resulting in big data flows. The Internet2 Network Portal helps visualize this traffic. This map (left) highlights one such flow from EBI in Cambridge, England, to NLM in Bethesda, MD. NCBI plans to use the portal’s capabilities to examine characteristics of its inbound and outbound traffic in order to improve data exchange with researchers. The portal’s capabilities will expand to include additional properties based upon members’ requests.

Help build the Internet2 Portal

Internet2 is looking for community members with needs that require additional portal capabilities. Catch an in-depth demonstration of the portal’s current capabilities at the 2012 Spring Internet2 Member Meeting during the Performance Working Group meeting on Monday, April 23rd, from 11:00 a.m. – 12:00 p.m. in Salon B, and at the Internet2 Advanced Network Services session on Wednesday, April 25th, 3:00 – 5:00 p.m. in Salon J at the Crystal Gateway Marriott. Please communicate your ideas and priorities for new portal features to Jeff Boote at: boote@internet2.edu.