THE GLOVE BOX
SECURE ANALYSIS ENVIRONMENTS

WHAT IS GLOVE BOX?

Glove Box is a general, end-to-end, scalable system that allows the creation of complete secure data enclaves (SDE's) for customized collaborative research with data storage, computational resources (including HPC), and analysis capabilities, all connected by software-defined secure networks. The enclaves will be configurable (and reconfigurable) via software, for quick and easy provisioning and customization.

WHY IS GLOVE BOX NEEDED?

The need for secure access to large, sensitive data has become a pressing problem. It is increasingly difficult to support research that requires specialized and dedicated hardware that may be needed only for a short time, and the costs are magnified when these systems have strict security requirements.
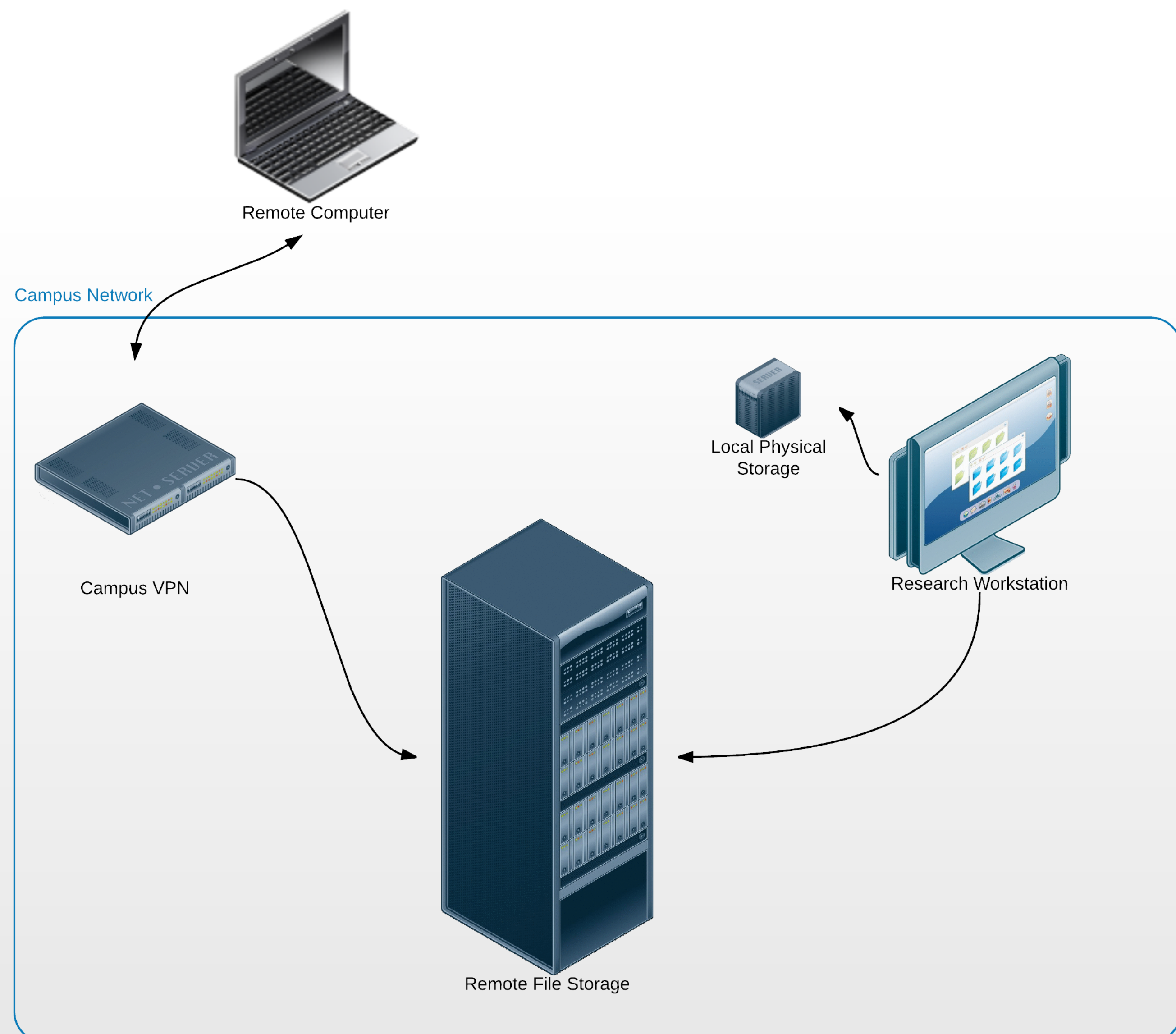
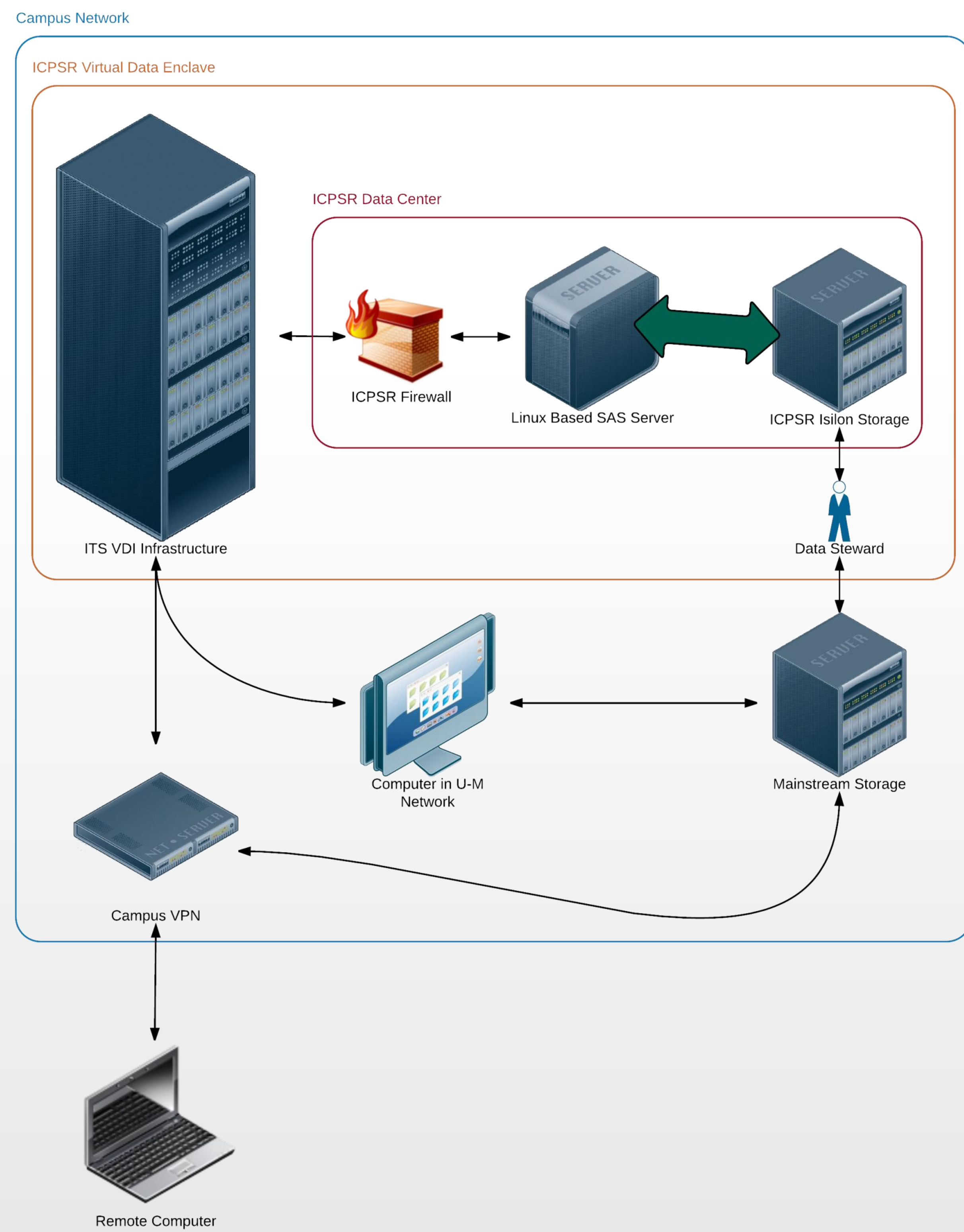Glove Box aims to alleviate these issues by providing:

- ❑ Customized research environments
- ❑ Software-defined private networks with controlled access to the external network
- ❑ Ability to scale resources immediately without administrative interaction
- ❑ "Recipes" that create computing environments that are pre-vetted for different levels of security

THE PAST - HOW WE USED TO MANAGE RESEARCH DATA

- ❑ Data stored on a remote device and accessed over network connection. Far too slow for large datasets.
- ❑ Other option was to store data locally on drives that can be stolen, lost, or damaged.

Remote Computer

Campus Network

Campus VPN

Local Physical Storage

Research Workstation

Remote File Storage

CURRENT STATE – HOW WE MANAGE RESEARCH DATA TODAY

Campus Network

ICPSR Virtual Data Enclave

ICPSR Data Center

ICPSR Firewall

Linux Based SAS Server

ICPSR Isilon Storage

ITS VDI Infrastructure

Data Steward

Computer in U-M Network

Mainstream Storage

Campus VPN

Remote Computer

IMPROVED DATA ANALYSIS EFFICIENCY

While the above setup avoids the issue of processing large files over a slow network, it still requires a great deal of technical knowledge, time, labor, and administrative support to manage.

FUTURE STATE –THE EFFICIENCY OF PRIVATE CLOUD COMPUTING

Glove Box will allow fast and efficient access to data through direct data-to-processing connections. In addition, it also greatly reduces the amount of administrative time necessary for the setup of a secure analysis environment.

Through the use of the open source cloud computing software, OpenStack, this service will be able to offer the following to end users:

- ❑ Windows and Linux Workstations providing interactive desktop replacement services and bastion points of entry via encrypted remote desktop or secure shell (SSH)
- ❑ Windows and Linux Servers to provide long running application support
- ❑ Shared File Servers to ease sharing of data within a secure environment
- ❑ Linux High Performance Computing Clusters allowing for private on-demand clusters
- ❑ Data Intensive Hadoop and Spark Clusters to use the newest methods from the big data community
- ❑ Data-storage resources of different types

Remote Computer

Results

Researcher

Raw Data

Data Administrator

Campus Network

Campus VPN

Research Workstation

Campus Identity (MComm + UMROOT AD)

The Glove Box

Glove Box Instance Defined by Sponsored Project

Software Defined Network

VDI

HPC

Map Reduce

File Storage

Secure Data Bridge

Compute

File Storage

Orchestration