NRO: Overview of Enterprise Requirements

LTJG Robert Grzybowski, USN

IMINT/RTS - Enabling Technologies

Contractors:

Marti Bancroft, Shomo Technical Systems

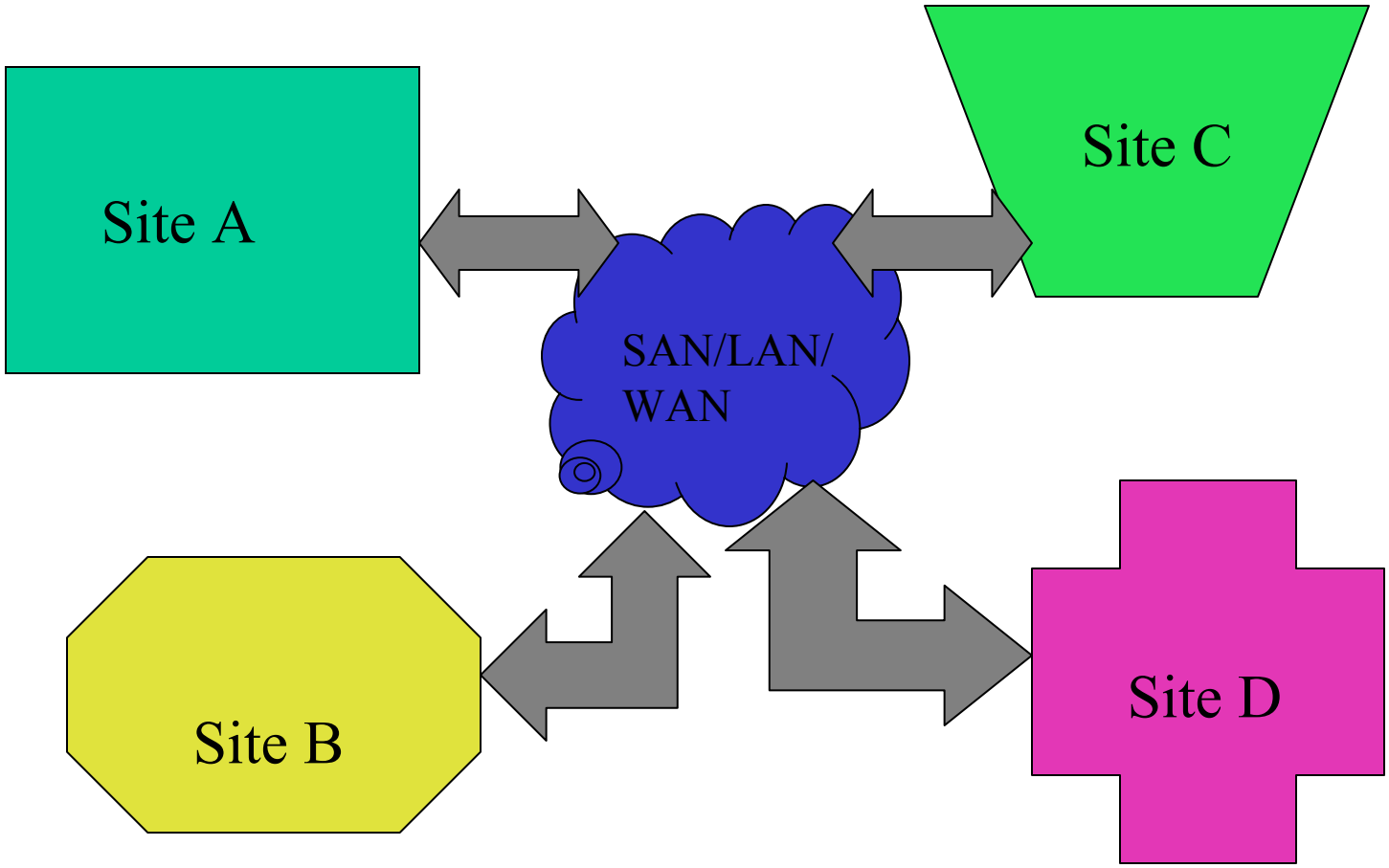
NRO Overall: Our Community

- Multiple data centers and labs
- Separated by campus, metro, or continental distances
- All developing/prototyping/deploying core enterprise applications
- Also expected to soon handle enterprise-wide data sharing and resource sharing

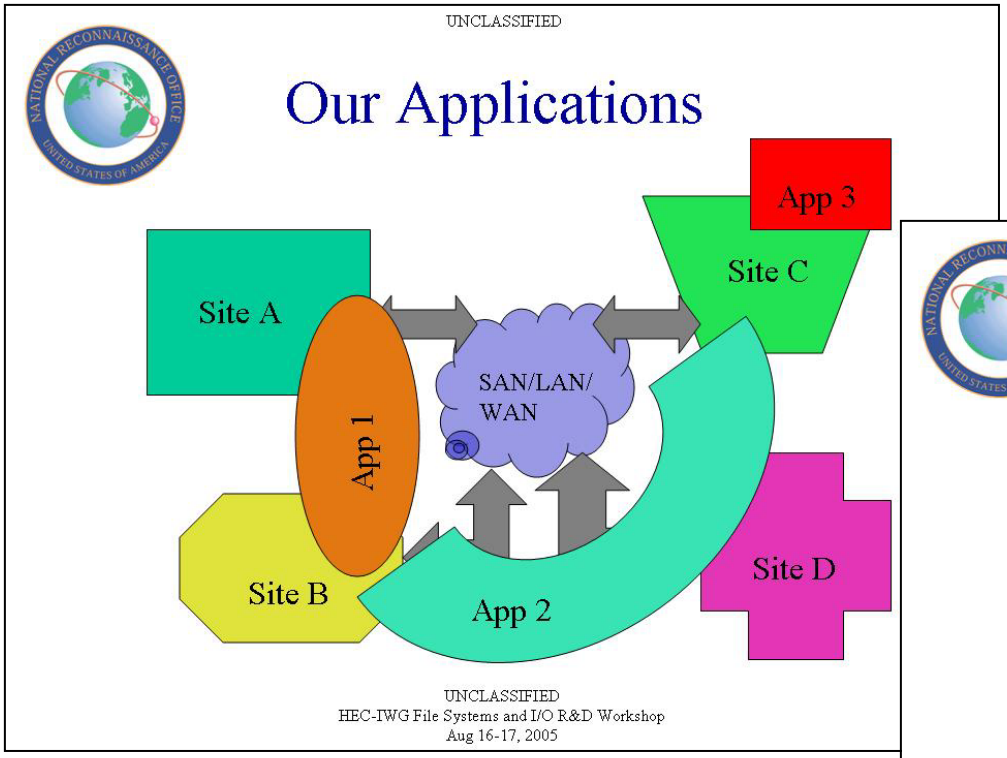
Our Emphasis

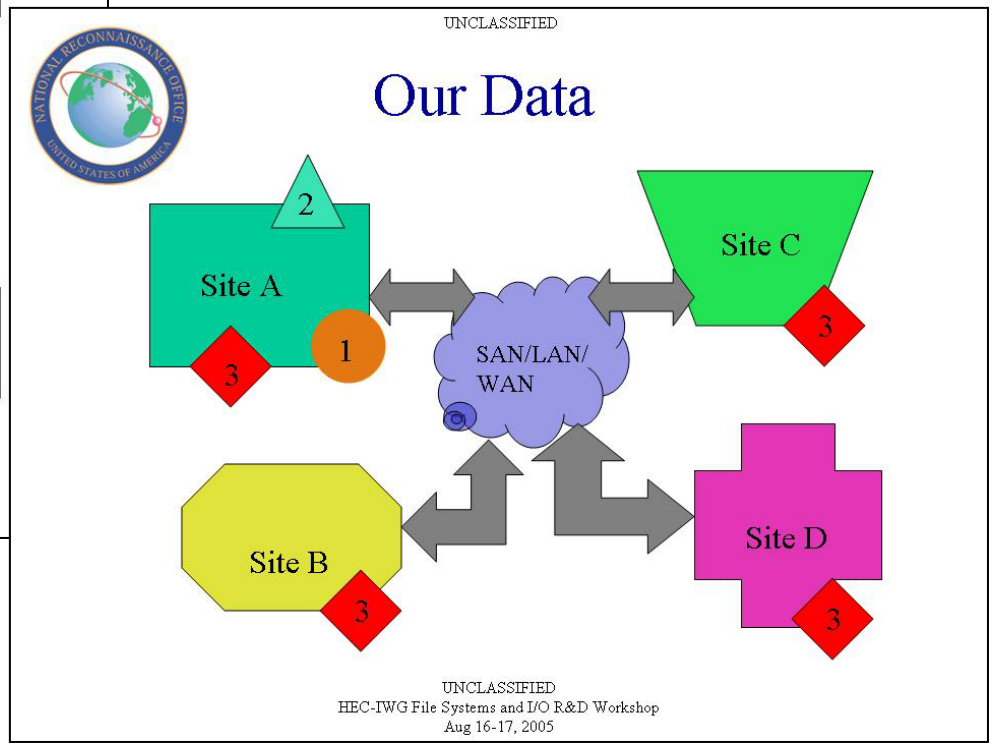
- Time critical / Time to solution!
- Datasets are huge
  - Priority is variable
- Time-Critical Problems are global

Our Enterprise Community

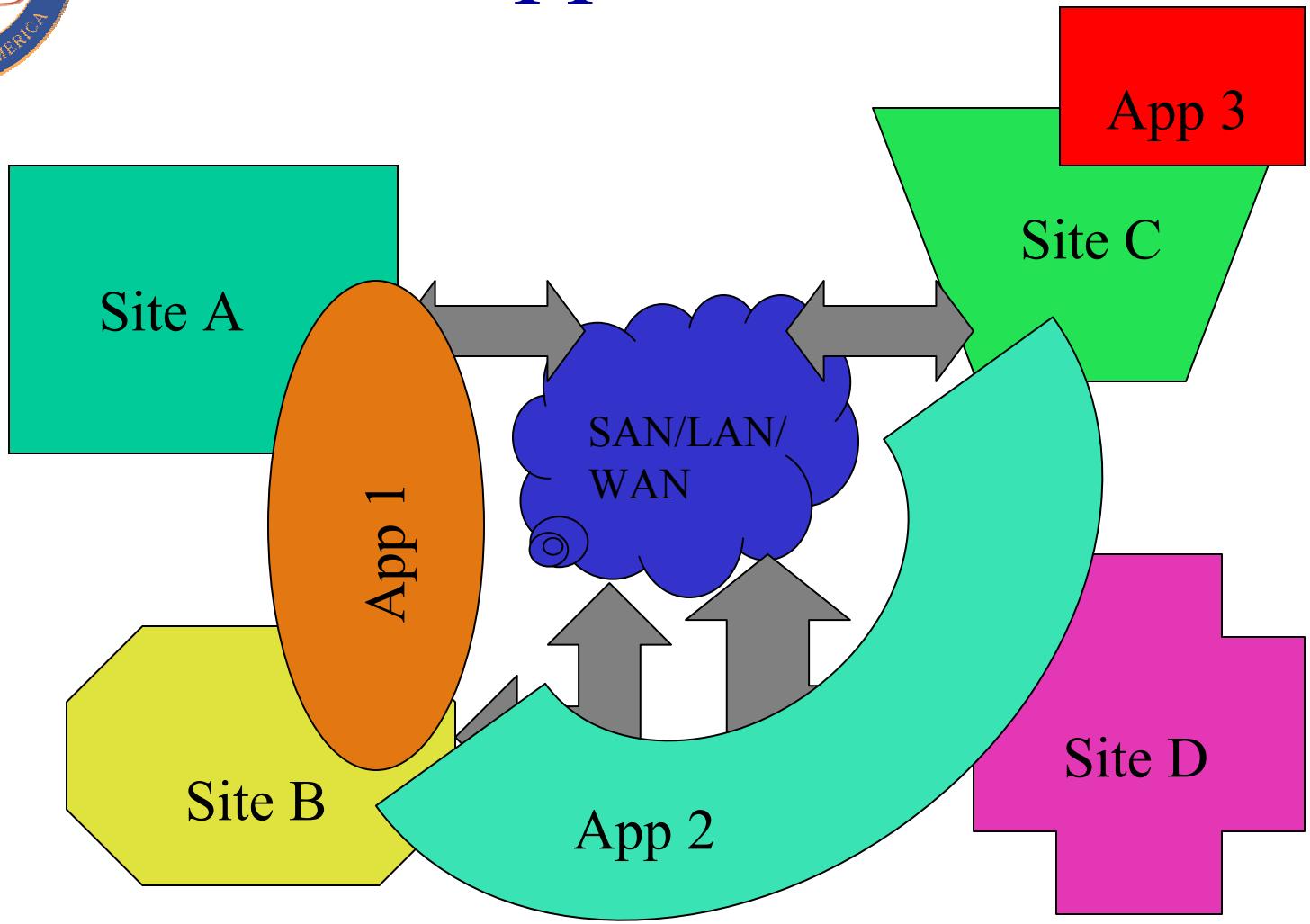


Data Management Support to Applications

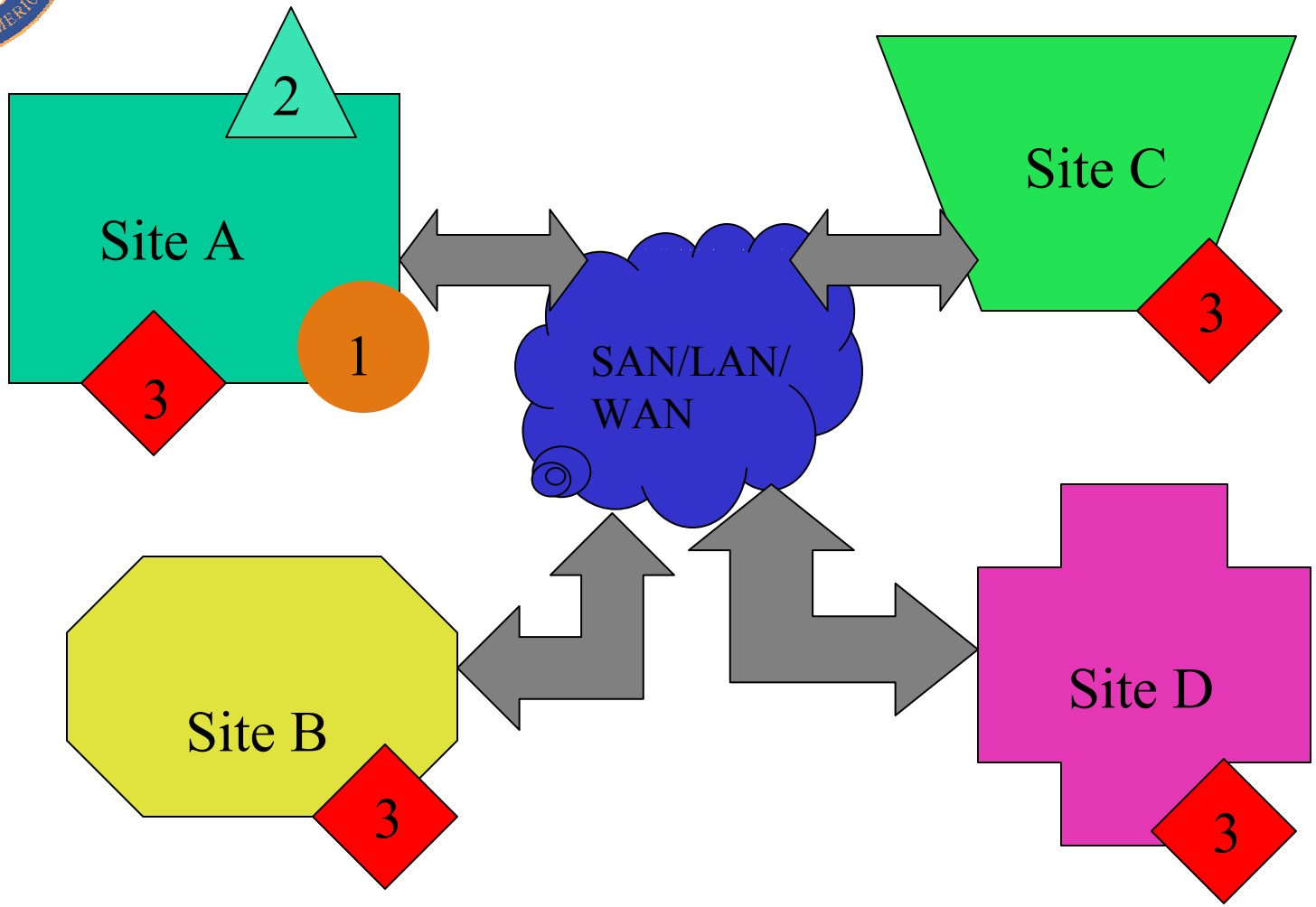




Our Applications



Our Data



Organizational Goals

- Combine geographically diverse data centers into a single virtual data center
- Maximize utilization
  - Currently, systems are dedicated to specific applications and that lowers utilization of total enterprise resources
  - Integrate old systems alongside of newer, resource sharing
- Run applications that may access data anywhere in this virtual data center
- Provide failover capability across data centers
- Self-federate to add, subtract assets dynamically
  - Data storage, data source, data sync
- Scale in all dimensions to meet evolving enterprise needs

I/O Requirements Continue to Increase

- 3.5 - 4.0 PetaBytes per day ingest (envisioned)
- Some data sets very large (many GB to even TB)
  - Some processing cannot be broken into sub-processes
  - Data occasionally must be reassembled during processing
- Single streams up to 30 Gigabytes/second
- Internal ground station bandwidth 5-10x or more the ingest rates to ensure parallel processing chains and asynchronous access
- Restart tasks rather than checkpoint/restart

NRO's 5-10 Year View

- HPC performance requires HPC I/O performance
  - Mega to Gigabytes/second sustained for large datasets
  - Efficiency - let the CPUs and memories be used for compute, not shoving bytes around
    - DMA
    - Very large requests
- Key shared file systems features for HPC I/O

  We're not sure COTS development is going to fully realize our needs in these areas.
  - Striping to increase single-file rates
  - Means to increase aggregate file system rates without dropping MTBF numbers
  - Expandability without dump/restore
  - Recoverability/resiliency
  - Small files/metadata scaling to HPC space

Current Research Efforts

Combination research with prime contractors and in-house

- Geared specifically towards our missions
- Hardware technology development
- Mission related algorithms
- Some research may trickle down to commercial

Current HPC I/O research in modeling current and future I/O systems, no development into new file systems and I/O gap areas; NRO previously anticipated I/O evolution to occur in sync with COTS HPC development.

Proposed Future Research Focal Points

- Data transparency
  - Our current systems are heterogeneous, need to consider future architecture goal
  - Goal is 30 GBytes/sec, today we have trouble reaching 500 MBytes/sec
  - XML, XDR, HDF5… efficiency?
  - Long term archives?
  - Have metadata include information about following binary data
- QoS and Determinism
  - Large files are IMPORTANT to us
  - Mitigation of adverse effects of transport on different data types

Proposed Future Research Focal Points

- How can you easily/automatically add resources to the distributed environment?
  - Storage (adding storage or configuring compute resources to use storage) is a challenge
    - Easier with NAS but at a huge performance and efficiency penalty
    - Investigating newer storage technologies
  - Storage & compute resources need to self-federate
- Data awareness in scheduling - WAN will be enterprise internal bandwidth bottleneck for "follow the sun" processing

NRO: Overview of Enterprise Requirements

Questions?