Information Power Grid

A Vision for Large Scale Computing and Data Management

William E. Johnston, Project Manager

Arsi Vaziri, Deputy Project Manager

Leigh Ann Tanner, Implementation Task Manager

The NAS Division

NASA Ames Research Center

William J. Feiereisen, Division Chief

William Thigpen, Engineering Branch Chief,

http://www.ipg.nasa.gov

Application Classes Motivating Widely Distributed Computing Environments

- ***Computational modeling, multi-disciplinary simulation, and scientific data analysis with a world-wide scope of participants*** **– e.g. aviation safety, observational cosmology, climate modeling**
- ***Real-time data analysis and collaboration involving on-line instruments*****, especially those that are unique national resources – e.g. wind tunnels, turbomachine test cells, Mars sample laboratory**

Motivating Applications

- ***Generation, management, and use of very large, complex data archives*** **that are shared across global science communities – e.g. Earth environment data (EOS), human genome data**
- ***Collaborative, interactive*** **analysis and visualization of massive datasets by multi-Center teams – e.g. wind tunnel data, air/space frame design data**

Multi-disciplinary Simulations

Multi-disciplinary simulations provide a good *example of a class of applications that are very likely to require aggregation of distributed computing, data, and intellectual resources*.

Such simulations – e.g. whole system aircraft simulation and whole system living cell simulation – *require integrating applications and data that are developed by different teams of researchers* frequently in different locations.

The research teams are the only ones that have the expertise to maintain and improve the simulation code and/or the body of experimental data that drives the simulations. This results in an inherently distributed computing and data management environment.

Consider a vision for Aviation Safety:

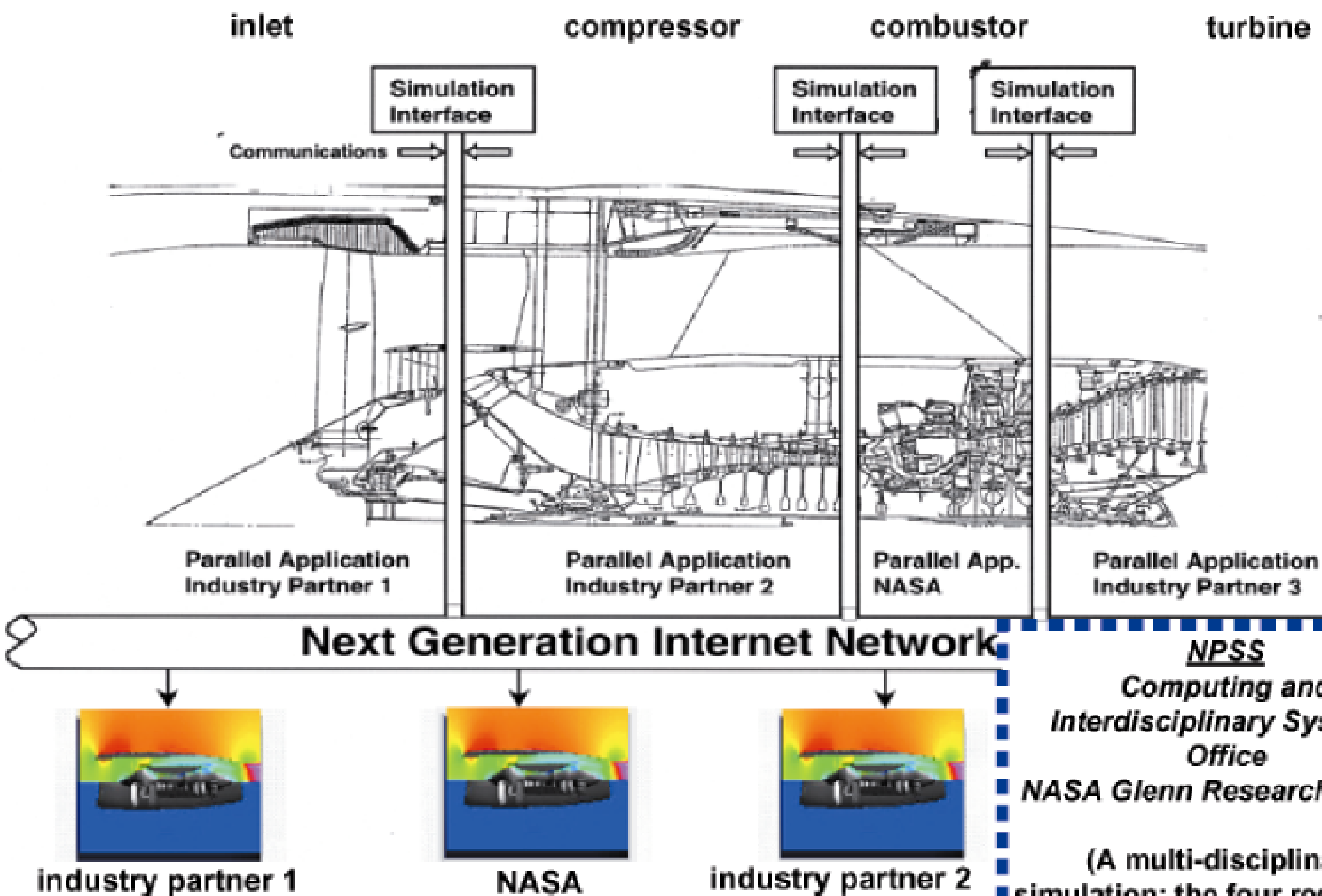
How do we simulate the entire commercial airspace of the country?

(Yuri Gawdiak (VNAS) and Bill McDermott, NASA Ames, John Lytle and Gregory Follen, NASA Glenn (NPSS)).

This vision is being approached through a set of increasingly complex and computationally intensive integrations:

- **Component simulations are combined to get a system simulation.**

Multi-Component Simulation



full simulation viewed at all collaborating sites

NPSS
Computing and Interdisciplinary Systems Office
NASA Glenn Research Center

(A multi-disciplinary simulation: the four regimes of an operating jet engine.)

Information Technology

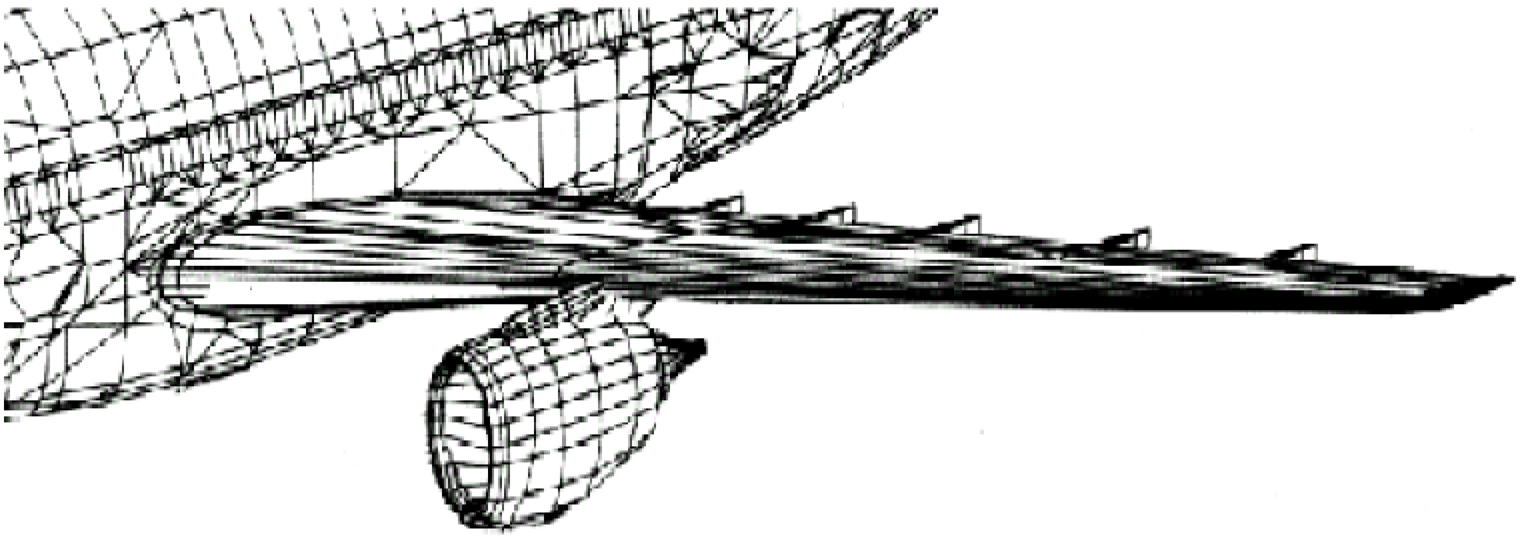
Jet Engine System

Issues for combining component simulations

- **wrapping the simulation code**
- **composing these codes**
- **coordinating resources for executing the multiple components**

¨ **Multiple system simulations are coupled to represent pieces of a device.**

Multi-System Simulation



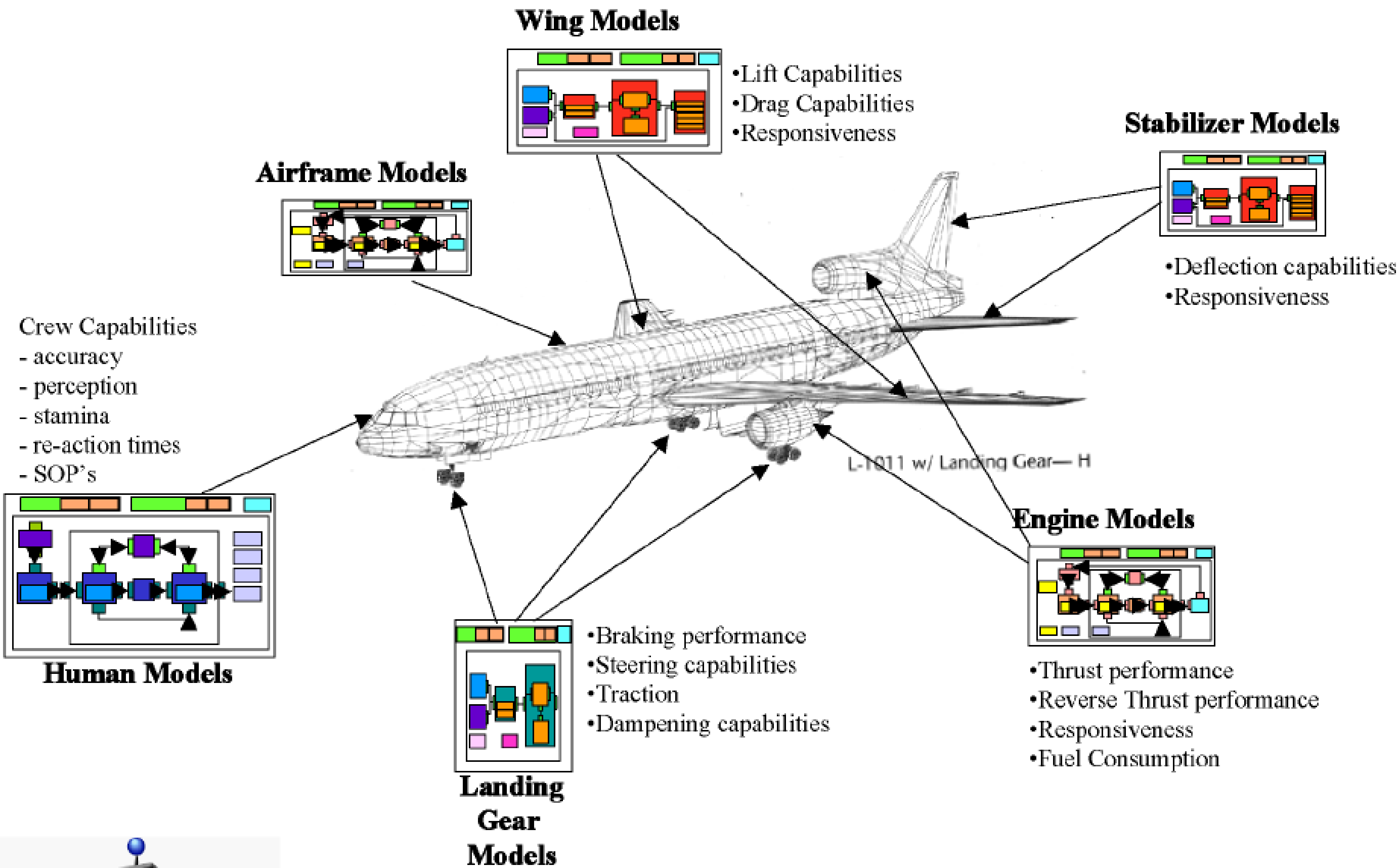
Engine System + Wing System

Multi system simulation issues:

- **multi-Center interactions - component parameters maintained by discipline experts**
- **shared compute and data resources**

¨ **Whole device simulations are produced by coupling all of the subordinate system simulations.**

Performance Characteristics for Fast-Time Simulations



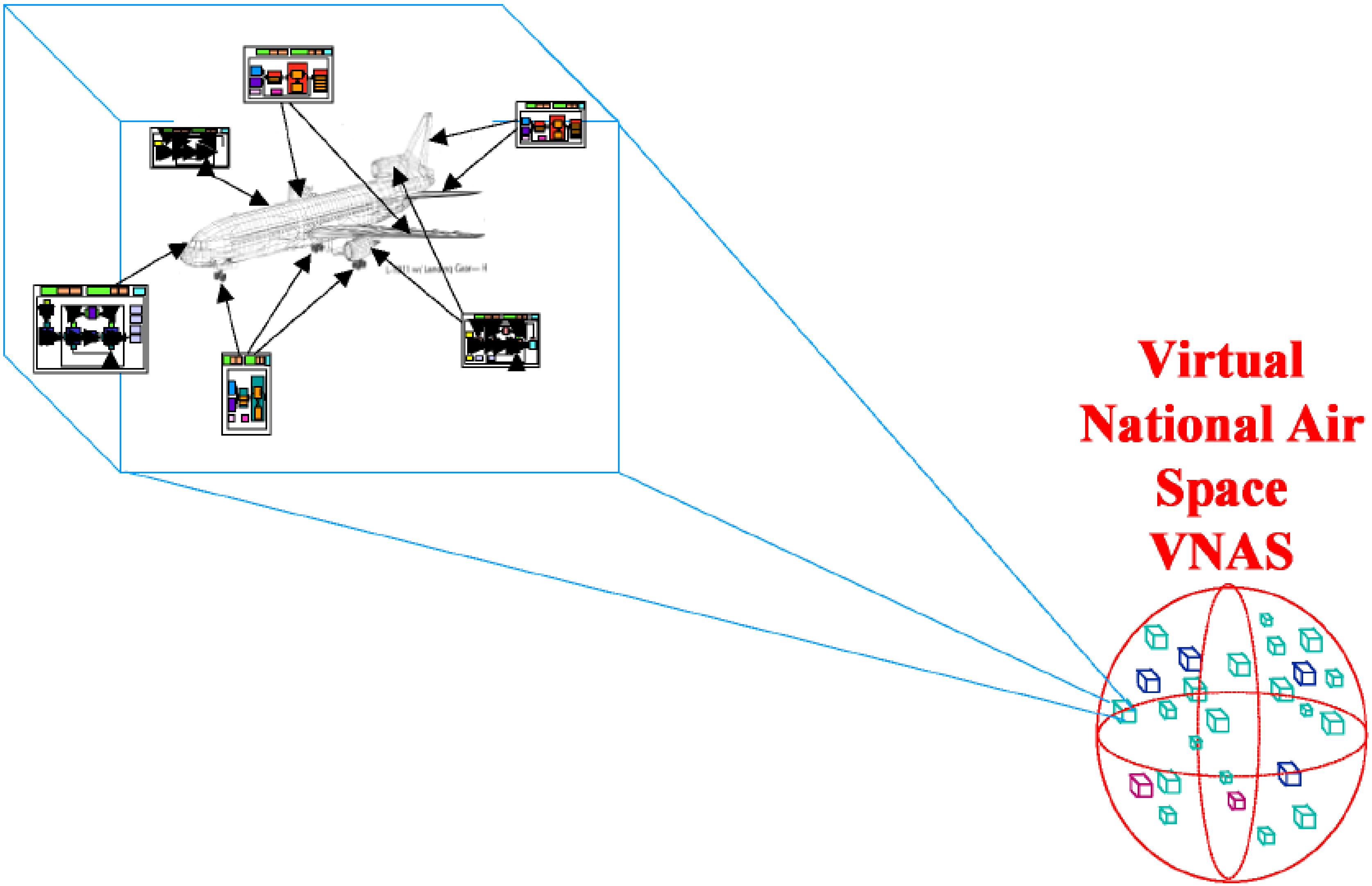
Information Technology

Whole device simulation issues:

- **increasingly complex interaction of models and data**
- **scaling of computing and networking capacity by O(10)**

¨ **Devices are inserted into a realistic environment.**

National Air Space (NAS) Simulation Environment

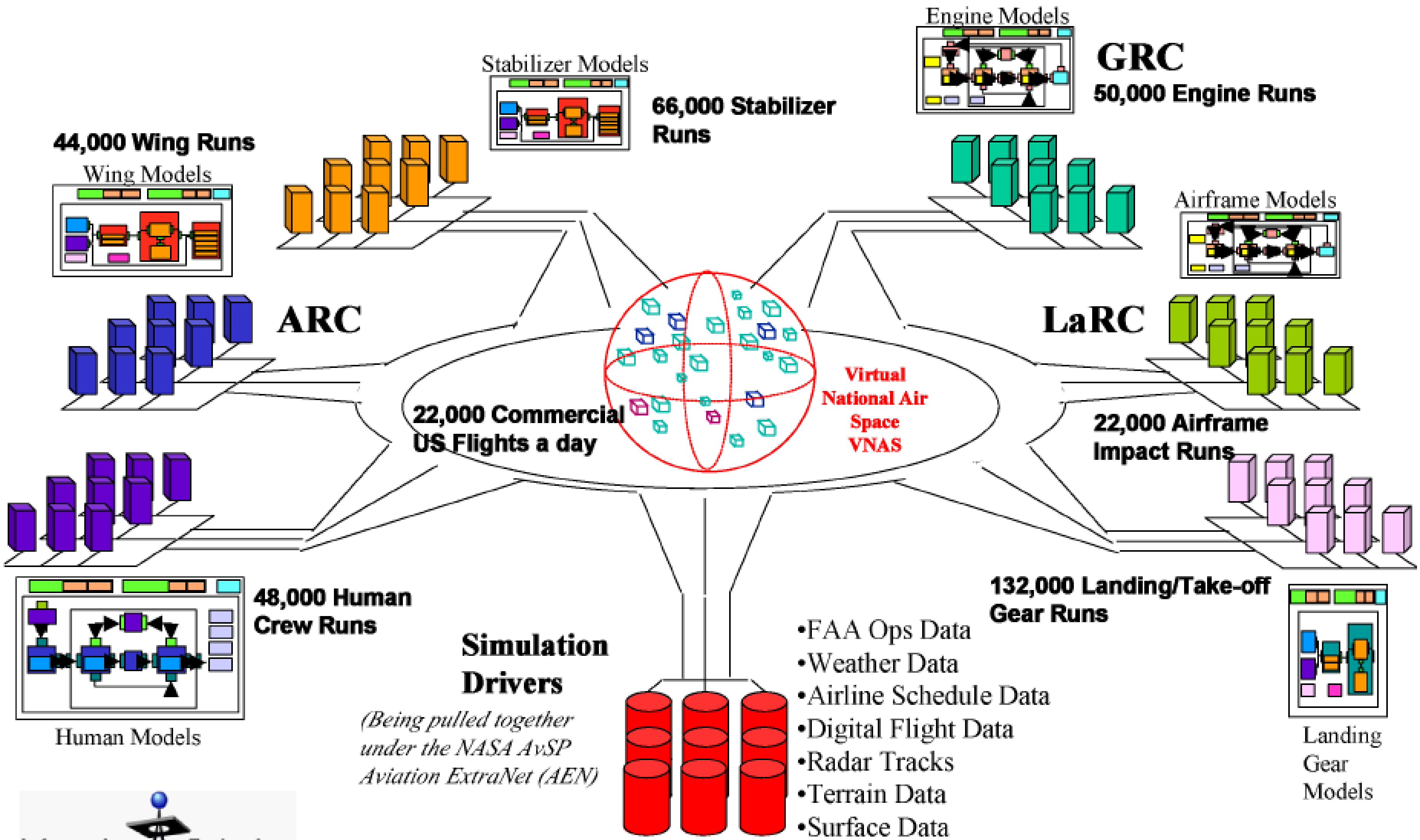


Issues:

- **the operating environment must be incorporated into the simulation, which effectively established further coupling of the system simulations**

¨ **Devices and environment are combined for operational systems simulation.**

Daily NAS Simulation Baseline Generation



Information Technology

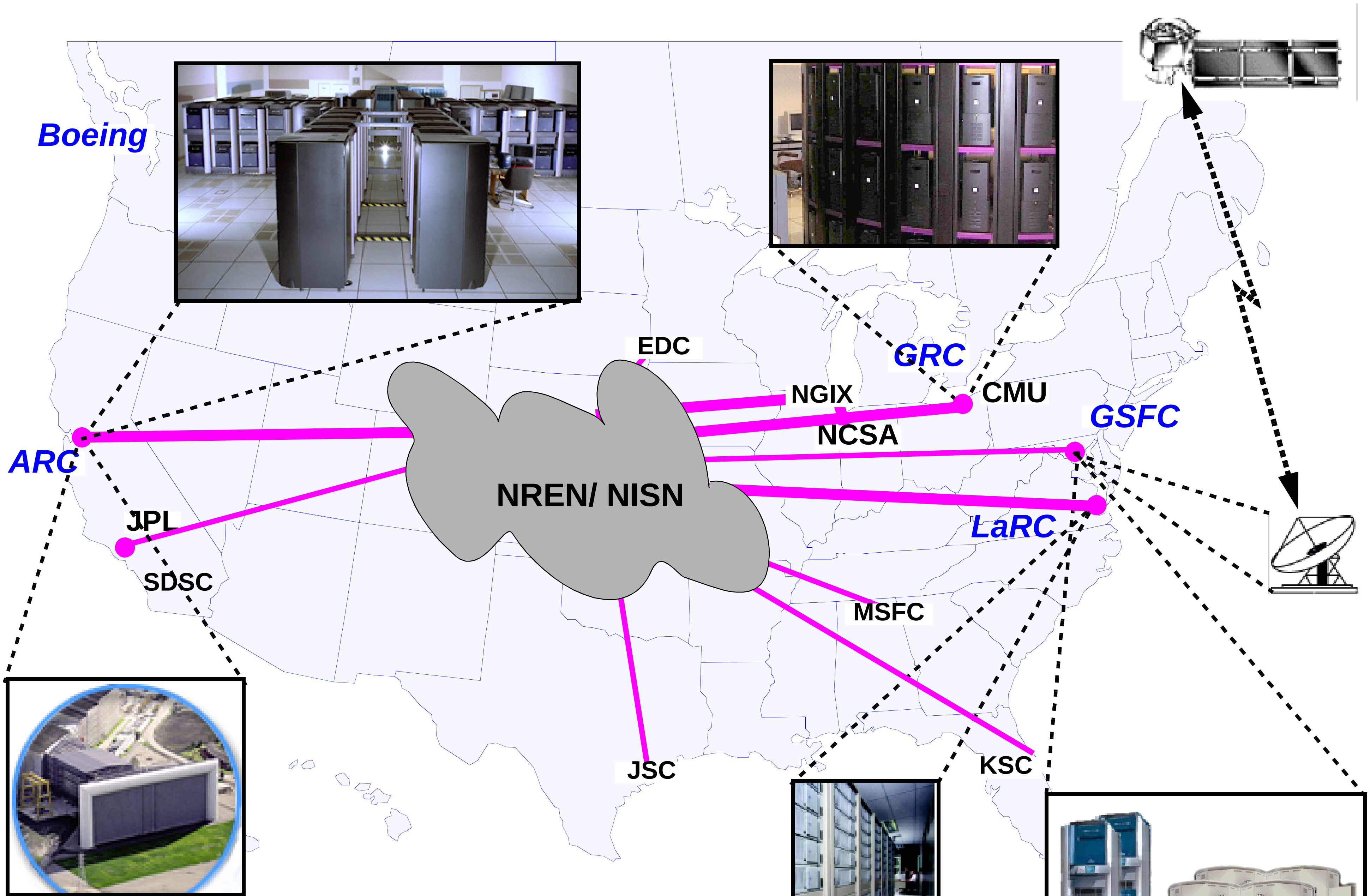
Issues:

- **overall system operational models and parameters must be incorporated, which introduces "indirect device" interactions**
- **scale computing and networking by O(10000)**

¨ **Clearly such simulations will need to use aggregated computing, data, instrument, and intellectual resources across multiple NASA Centers.**

Current Environment

- Some good tools, with others being developed
- Very little persistent common infrastructure

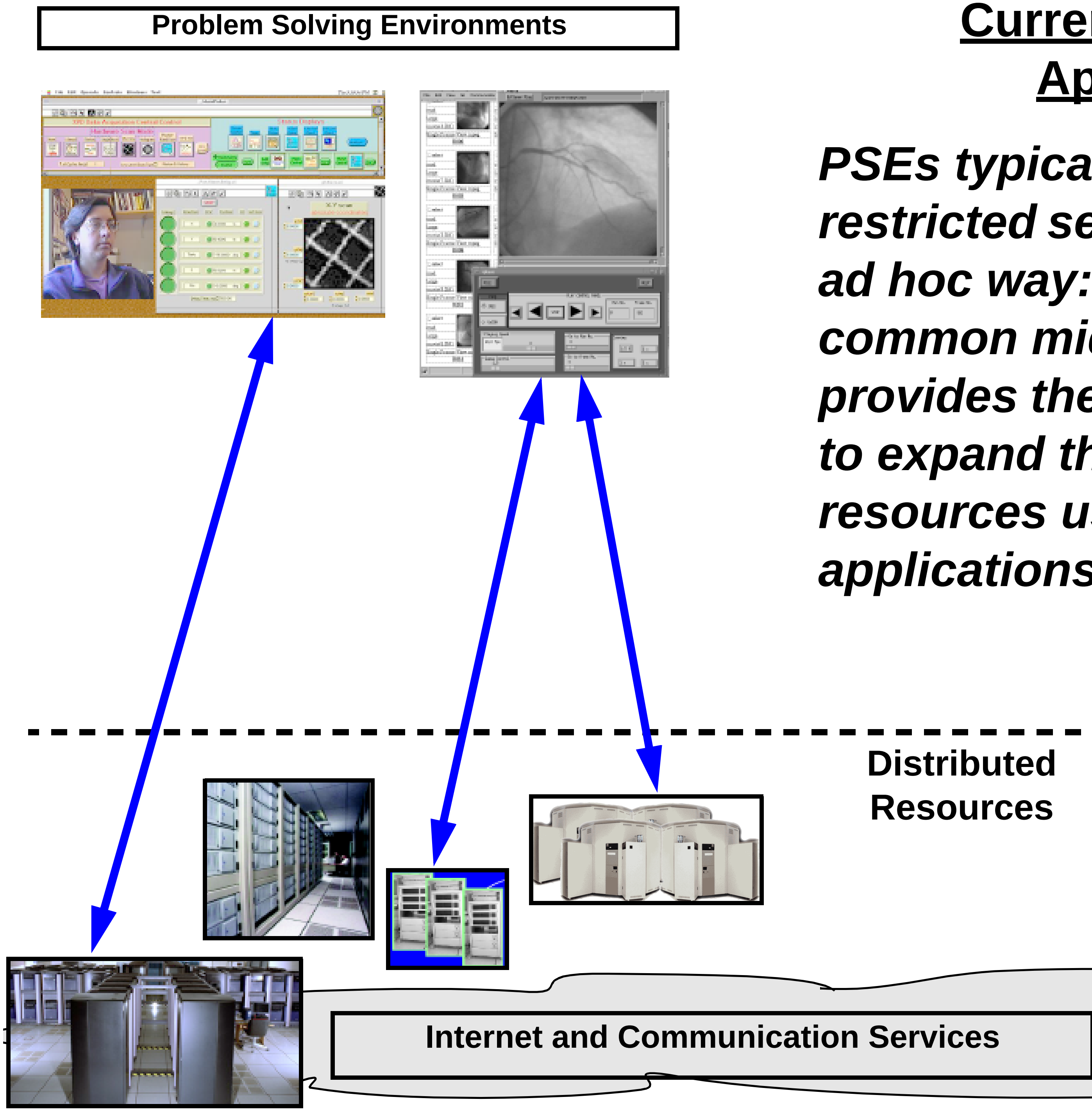


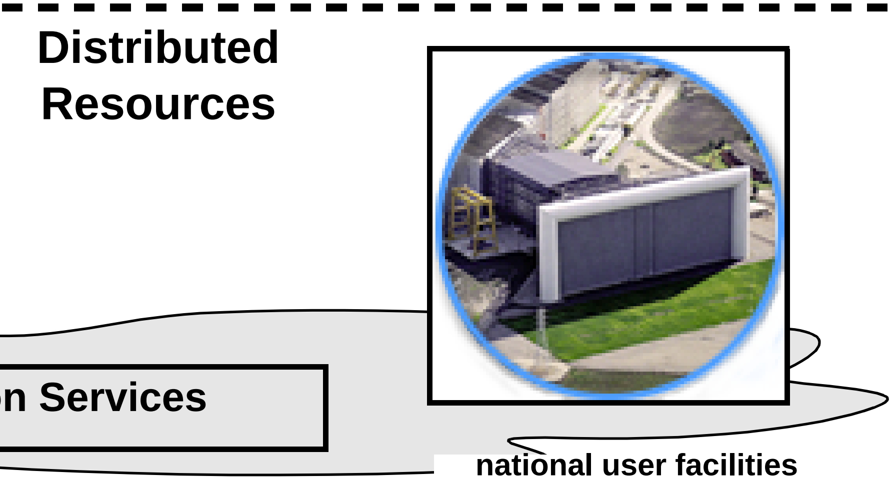
The NASA Environment Today

- **Result is that most distributed application environments deal with a restricted set of resources in an ad hoc way, and every project re-invents this scenario**

Current Distributed Applications

PSEs typically deal with a restricted set of resources in an ad hoc way: There is no common middleware that provides the uniformity needed to expand the scope of resources used by these applications.

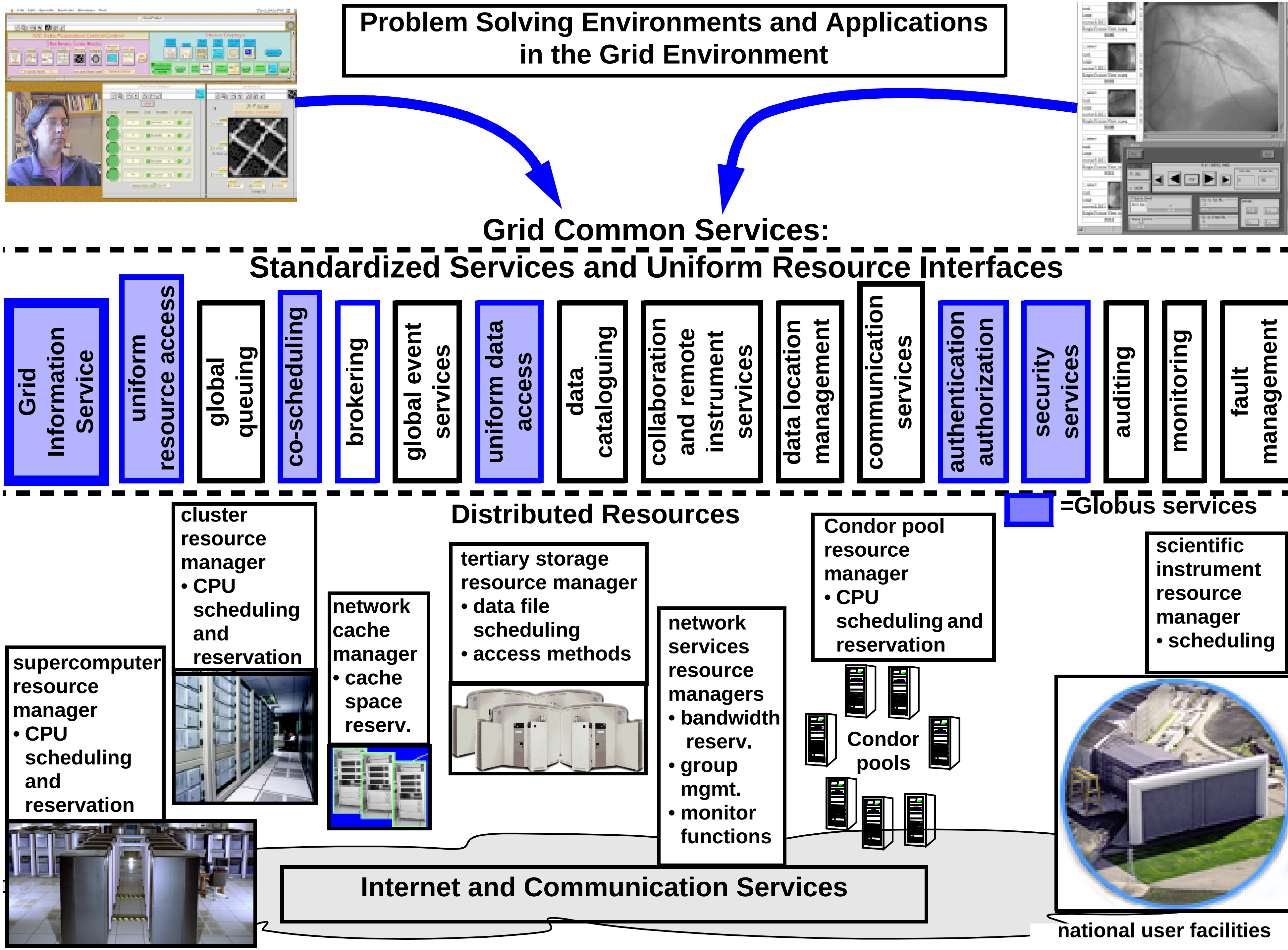




Vision for the Information Power Grid

The vision for “Grids” is to revolutionize the use of computing in science and engineering by making the construction and use of large scale systems of diverse resources as easy as using today’s desktop environments.

- **It is the role of Grids – tools and middleware for widely distributed systems – to enable the degree of scalability in scientific and engineering computing necessary for NASA to address very large scale simulation and data analysis problems**



Vision for the Information Power Grid

Approach and Goals for IPG

- **Grids are built through collaborative efforts, and at the same time facilitate collaboration: IPG is a collaboration among several NASA Centers and the NSF Supercomputer Center consortia (PACIs), with the Grid Forum providing "coordination" of man institutions world wide**

- **Deployment of existing technology (Globus [1], Condor [23], Grid portals [13], etc.) will provide for relatively rapid impact – IT/ANCS computing and storage resources are providing the prototype production IPG environment**

Approach and Goals

- **NAS will provide the development and support for critical Grid services – this will ensure persistent and usable infrastructure across the NASA Centers**
- **Grid support for building collaboration services will both facilitate construction of the Grid and more readily provide collaboration tools to users**
- **The IPG operational model provides for easy user access across Centers and for local control of resources that are connected to IPG**

Approach and Goals

- **Strong security will be provided from the start in order to address authentication, authorization, and infrastructure assurance in open science networks for both applications and Grid services**

- **As Grid services are debugged and validated they will be offered to CoSMO as a means of providing a uniform supercomputing environment – the production IPG represents a new service delivery model for NASA computing, data, and instrument resources**

Expected Outcomes

Grids will present a *uniform usage and management interface* to computing, collaboration, storage, and instrument systems, and provide the *capability of dynamically and scalably connecting* these into large, on-demand systems. This should lead to:

- **Increased mobility of experience and access to computing and data by computational scientists**
- **Routine collaboration among NASA Centers and their university partners through ready and secure access to collaboration tools, remote instruments, and petabyte size data sets**

Expected Outcomes

- **Easily used, application oriented, user interfaces (problem solving environments / workbenches) that provide access to powerful, diverse, and widely distributed resources**
- **Large-scale problem solving systems that involve coupling multiple computational simulations and data archives will be built dynamically from aggregated resources to support scientific and engineering computing and data based activities that are not steady state - i.e. those that may require a different resource mix for every different problem**

- New approaches to laboratory science through the coupling of large-scale computing and storage systems to instrument systems in order to provide real-time analysis of experiment data and feedback based experiment control
- Standardized services and tools that make it easier to incorporate new computer architectures, data systems, and instruments into a usable application environment
- A pool of resources (not just within NASA) that has standardized capabilities, aggregation strategies, and management so that large-scale systems could be quickly built for emergency response situations

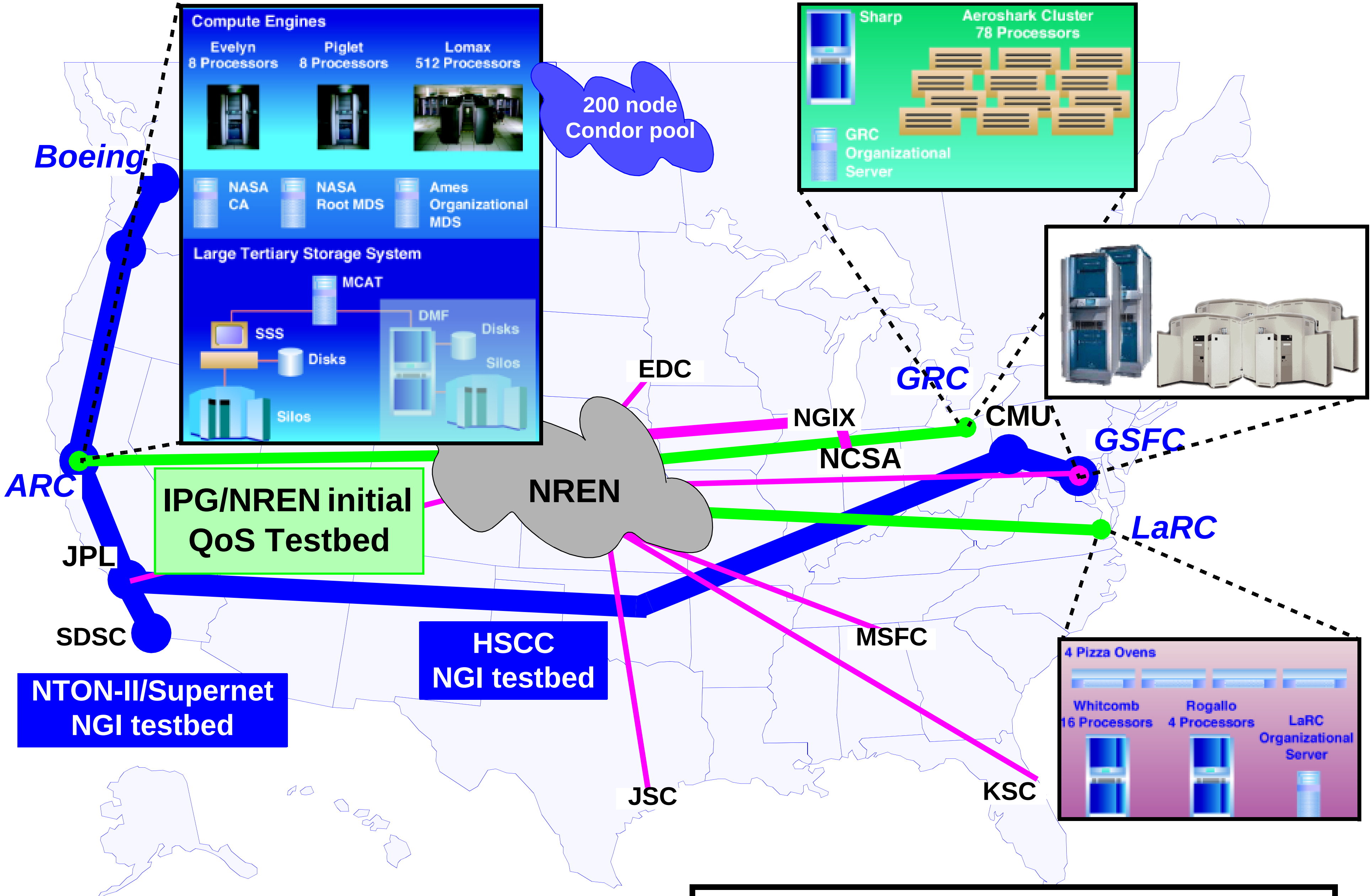
The State of IPG

10/2000 Baseline Operational System

- Computing resources: ≈600 CPU nodes in half a dozen SGI Origin 2000s and several workstation clusters at Ames, Glenn, and Langley, with plans for incorporating Goddard and JPL, and ≈200 nodes in a Condor pool

- Wide area network interconnects of at least 100 mbit/s

- Storage resources: 30-100 Terabytes of archival information/data storage *uniformly* accessible from all IPG systems via SDSC's MCAT and SRB



IPG Baseline System and High Data-Rate (DX) Testbed

State of IPG: Baseline Operational System (cont.)

- **Globus providing the Grid common services**
- **Programming and program execution support**
  - **Grid MPI (via the Globus communications library)**
  - **CORBA integrated with Globus**
  - **global job queue management**
  - **high throughput job manager**
  - **Condor [23] ("cycle stealing" computing)**
- **A stable and supported operational environment**
- **Several "benchmark" applications operating across IPG (multi-grid CFD code, parameter study)**
- **Multi-Grid operation (applications operating across IPG and NCSA)**

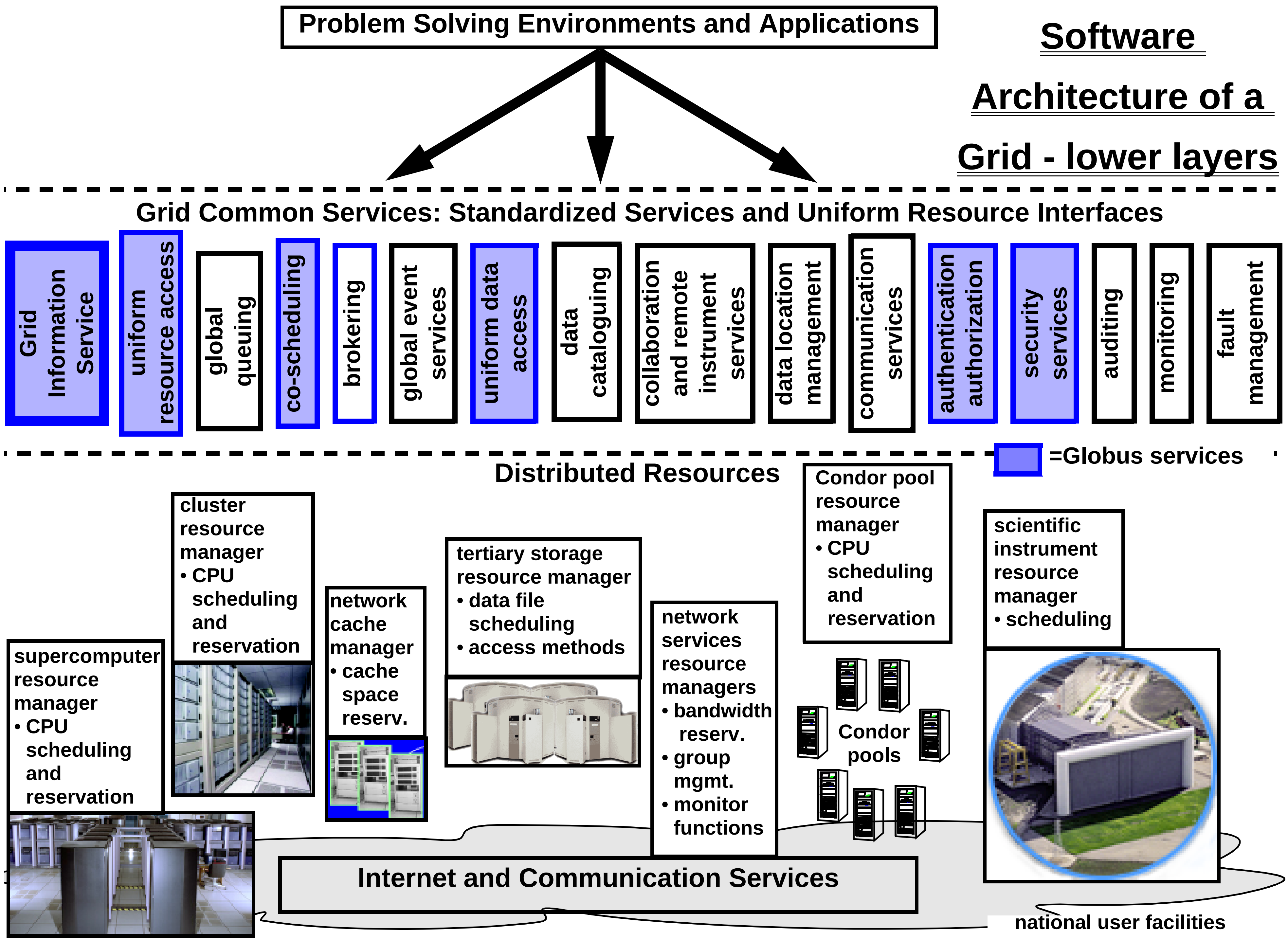
Notes on The Information Power Grid Architecture

¨ **Problem Solving Environments are the user interface to Grids, and are supported by Grid toolkits for**

- **job submission, control, and tracking services**
- **workflow management for specific classes of applications (e.g. physics data analysis frameworks or aircraft design parameter study managers)**
- **policy based access control, etc.**

- **Application development tools and services support various styles of programming in the Grid environment, as well as the development of Grid services themselves. E.g.:**
  - **uniform data access methods developed in the DataGrid project [14] will form the foundation for global storage management services such as MCAT/SRB [22] and the Storage Access Coordination System (STACS), HRM (HPSS Resource Manager) [15]**
  - **Globus I/O enabled MPI library provides coordinated, MPI communication between processes on separate systems**

- The “Grid Common Services” locate, schedule, and provide uniform views of the underlying resources. E.g.: resource access, naming and location, and co-scheduling for computing, networking, and instrument systems
- Resource managers provide the basic functionality and access for the actual resources
  - some already exist - e.g. batch schedulers - however do not always support the required functionality (e.g. for advance reservation)


Software Architecture of a Grid - lower layers
Problem Solving Environments and Applications
Grid Common Services: Standardized Services and Uniform Resource Interfaces
Grid Information Service
uniform resource access
global queuing
co-scheduling
brokering
global event services
uniform data access
data cataloguing
collaboration and remote instrument services
data location management
communication services
authentication authorization
security services
auditing
monitoring
fault management
=Globus services
Distributed Resources
supercomputer resource manager
• CPU scheduling and reservation
cluster resource manager
• CPU scheduling and reservation
network cache manager
• cache space reserv.
tertiary storage resource manager
• data file scheduling
• access methods
network services resource managers
• bandwidth reserv.
• group mgmt.
• monitor functions
Condor pool resource manager
• CPU scheduling and reservation
Condor pools
scientific instrument resource manager
• scheduling
Internet and Communication Services
national user facilities

Software Architecture of a Grid - upper layers

Problem Solving Environments

- Tools to implement the human interfaces
- Mechanisms to express, organize, and manage the workflow of a problem solution
- Access control
- E.g. SciRun [24], Ecce [25], "portals", WebFlow [26],...

Applications and Supporting Tools

- application codes
- visualization toolkits
- collaboration toolkits
- instrument management toolkits
- data publication, subscription, and access toolkits

Grid enabled libraries

Application Development and Execution Support Services and Systems

- Globus MPI
- CORBA
- Condor-G
- Legion
- Java/Jini
- OLE/DCOM

Grid Common Services

Distributed Resources

Information Power Grid Vision

IPG will provide uniformity, location independence, and capabilities for building complex, on-demand, large scale distributed systems from computing, data storage, instruments, and intellectual resources that are spread across the NASA Centers and their partners.

This will lead to revolutionary new capabilities for solving large-scale science and engineering problems.

References and Acronyms

[1] Globus is a middleware system that provides a suite of services designed to support high performance, distributed applications. Globus provides:

- Resource Management: Components that provide standardized interfaces to various local resource management systems (GRAM) manage allocation of collections of resources (DUROC). All Globus resource management tools are tied together by a uniform resource specification language (RSL).
- Remote Access: Components that enable remote access to files (GASS and RIO) and executables (GEM).
- Security: Support for single sign-on, authentication, and authorization within the Globus system (GSI) and (experimentally) authorization (GAA).
- Fault Detection: Basic support for building fault detection and recovery into Globus applications.
- Information Infrastructure: Global access to information about the state and configuration of system components of an application (MDS).
- Grid programming services: Support writing parallel-distributed programs (MPICH-G), monitoring (HBM), etc.

www.globus.org provides full information about the Globus system.

[2] *The Grid: Blueprint for a New Computing Infrastructure*, edited by Ian Foster and Carl Kesselman. Morgan Kaufmann, Pub. August 1998. ISBN 1-55860-475-8. http://www.mkp.com/books_catalog/1-55860-475-8.asp

[3] "Grids as Production Computing Environments: The Engineering Aspects of NASA's Information Power Grid," William E. Johnston, Dennis Gannon, and Bill Nitzberg. Eighth IEEE International Symposium on High Performance Distributed Computing, Aug. 3-6, 1999, Redondo Beach, California. (Available at http://www.nas.nasa.gov/~wej/IPG)

[4] "Vision and Strategy for a DOE Science Grid" - http://www.itg.lbl.gov/~wej/Grids

[5] See www.nas.nasa.gov/IPG for project information and pointers.

[6] See http://www-itg.lbl.gov/NGI/ for project information and pointers.

[7] The Particle Physics Data Grid has two long-term objectives. Firstly: the delivery of an infrastructure for very widely distributed analysis of particle physics data at multi-petabyte scales by hundreds to thousands of physicists. Secondly: the acceleration of the development of network and middleware infrastructure aimed broadly at data-intensive collaborative science. http://www.cacr.caltech.edu/ppdg/

[8] Tierney, B. Lee, J., Crowley, B., Holding, M., Hylton, J., Drake, F., "A Network-Aware Distributed Storage Cache for Data Intensive Environments", Proceeding of IEEE High Performance Distributed Computing conference (HPDC-8), August 1999.

[9] "Real-Time Generation and Cataloguing of Large Data-Objects in Widely Distributed Environments," W. Johnston, Jin G., C. Larsen, J. Lee, G. Hoo, M. Thompson, and B. Tierney (LBNL) and J. Terdiman (Kaiser Permanente Division of Research). Invited paper, International Journal of Digital Libraries - Special

Issue on “Digital Libraries in Medicine”. May, 1998. http://www-itg.lbl.gov/WALDO/

[10] MAGIC: “The MAGIC Gigabit Network.” See: http://www.magic.net

[11] TerraVision-2: VRML based data fusion and browsing - www.ai.sri.com/TerraVision

[12] “A Monitoring Sensor Management System for Grid Environments,” Brian Tierney, Brian Crowley, Dan Gunter, Mason Holding, Jason Lee, Mary Thompson. To appear, HPDC-9, July, 2000. Available at http://www-didc.lbl.gov/JAMM/

[13] A collaborative effort to enable desktop access to remote resources including, supercomputers, network of workstations, smart instruments, data resources, and more - computingportals.org

[14] “The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets.” A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, S. Tuecke, (to be published in the Journal of Network and Computer Applications).

[15] “Storage Access Coordination Using CORBA,” A. Sim, H. Nordberg, L.M. Bernardo, A. Shoshani and D. Rotem. Proceedings of the International Symposium on Distributed Objects and Applications. See http://gizmo.lbl.gov/sm/

[16] The Clipper Project: Computational Grids providing middleware that supports applications requiring configurable, distributed, high-performance computing and data resources. See http://www-itg.lbl.gov/~johnston/Clipper

[17] The Grid Forum (www.gridforum.org) is an informal consortium of institutions and individuals working on wide area computing and computational Grids. Current working groups include Security (authentication, authorization), Scheduling and Resource Management, Grid Information Services, Application and Tool Requirements, Advanced Programming Models, Grid User Services and Operations, Account Management, Remote Data Access, Grid Performance

[18] "New Capabilities in the HENP Grand Challenge Storage Access System and its Application at RHIC" http://rncus1.lbl.gov/GC/docs/chep292lp1.doc
"STACS is ... responsible for determining, for each query request, which events and files need to be accessed, to determine the order of files to be cached dynamically so as to maximize their sharing by queries, to request the caching of files from HPSS in tape optimized order, and to determine dynamically which files to keep in the disk cache to maximize file usage."

[19] "DeepView: A Collaborative Framework for Distributed Microscopy." IEEE Conf. on High Performance Computing and Networking, Nov. 1998. See http://vision.lbl.gov/ (projects -> collaborative computing)

[20] **Akenti**: "**Certificate-based Access Control for Widely Distributed Resources**," Mary Thompson, William Johnston, Srilekha Mudumbai, Gary Hoo, Keith Jackson, Usenix Security Symposium '99. Mar. 16, 1999. (See http://www-itg.lbl.gov/Akenti)

[21] GAA: "**Generic Authorization and Access control API**" (GAA API). IETF Draft. http://ghost.isi.edu/info/gss_api.html)

[22] Storage Resource Broker (SRB) provides uniform access mechanism to diverse and distributed data sources. http://www.sdsc.edu/MDAS/

[23] Condor is a High Throughput Computing environment that can manage very large collections of distributively owned workstations. http://www.cs.wisc.edu/condor/

[24] SCIRun is a scientific programming environment that allows the interactive construction, debugging and steering of large-scale scientific computations. http://www.cs.utah.edu/~sci/software/

[25] Ecce - www.emsl.pnl.gov

[26] WebFlow - A prototype visual graph based dataflow environment, WebFlow, uses the mesh of Java Web Servers as a control and coordination middleware, WebVM. See http://iwt.npac.syr.edu/projects/webflow/index.htm

[27] "QoS as Middleware: Bandwidth Reservation System Design." Gary Hoo and William Johnston, Lawrence Berkeley National Laboratory, Ian Foster and Alain Roy, Argonne National Laboratory and University of Chicago. To appear, Eighth IEEE International Symposium on High Performance Distributed Computing, Aug. 3-6, 1999, Redondo Beach, California. (See http://www-itg.lbl.gov/Clipper/QoS)