

A Tour of Research Computing at Genentech

Reece Hart, Ph.D.
Scientific Manager
Research Computing & Informatics
Genentech, Inc.

April 27, 2009
Bio-IT World Expo
Boston, MA

Slides available at <http://harts.net/reece/pubs/>

Genentech
IN BUSINESS FOR LIFE

Organization

Research needs and Corporate needs must be balanced.

Research Needs

- **Modern infrastructure through rapid evolution and agile processes.**

Corporate Needs

- **Operational efficiency through standardization and consolidation.**

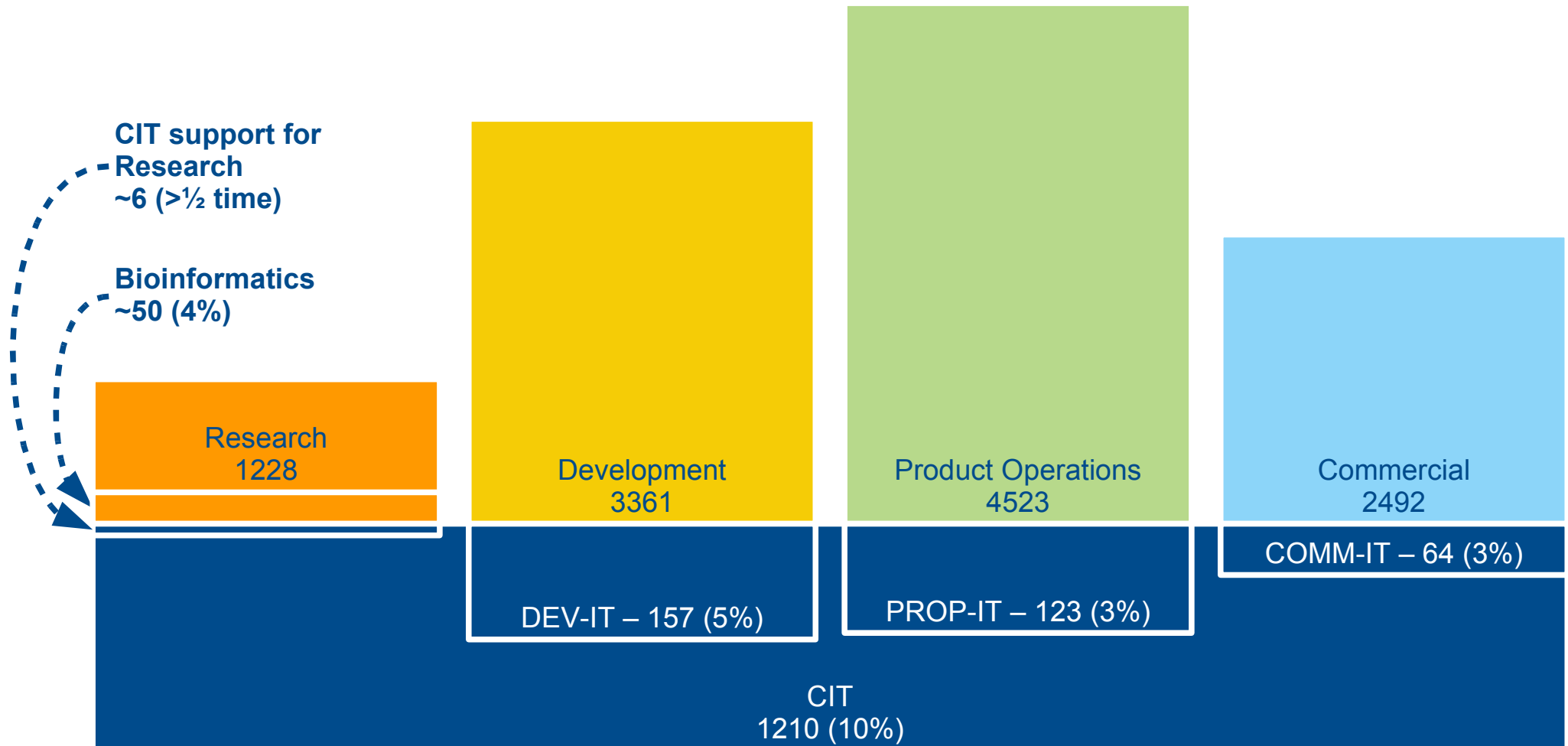


Genentech IT is centralized.



- **Centralized and standardized IT architecture.**
- **Centralized IT operations.**
- **Centralized IT accounting.**
- **Centralized support for project management, legal review, security.**

Divisions have unique needs and dedicated teams.



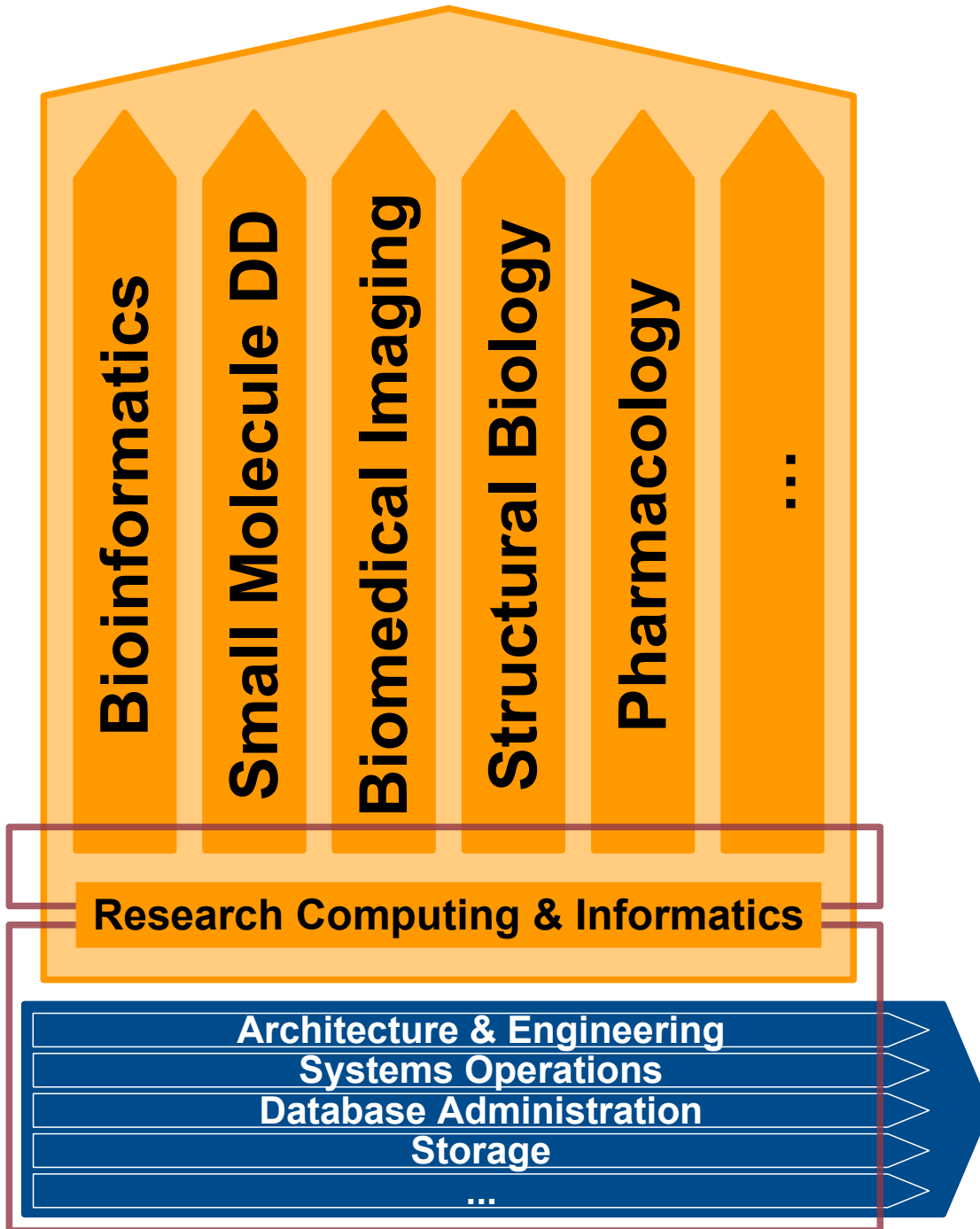
Dynamic needs
 Agile development
 Lots of custom dev.
 Non-validated systems
 Technical challenges
 Scaling challenges

Stable, predictable needs
 Large, long-term projects
 Mostly purchased applications with in-house integration
 Validated systems
 Reliability challenges

How is Research different from other groups?

- **Projects are smaller, cheaper, shorter.**
 - Project cycle shorter than budget cycle!
- **Needs and solutions are dynamic.**
 - Needs and options evolve quickly.
- **Good soon is better than perfect later.**
 - Projects are iteratively refined.
- **Validation and reliability secondary to functionality.**
- **Lots of custom development and gluing.**

RCI ensures alignment between Research & CIT.



➤ **Steering Committee**

- Strategic needs
- Budgeting
- Staffing

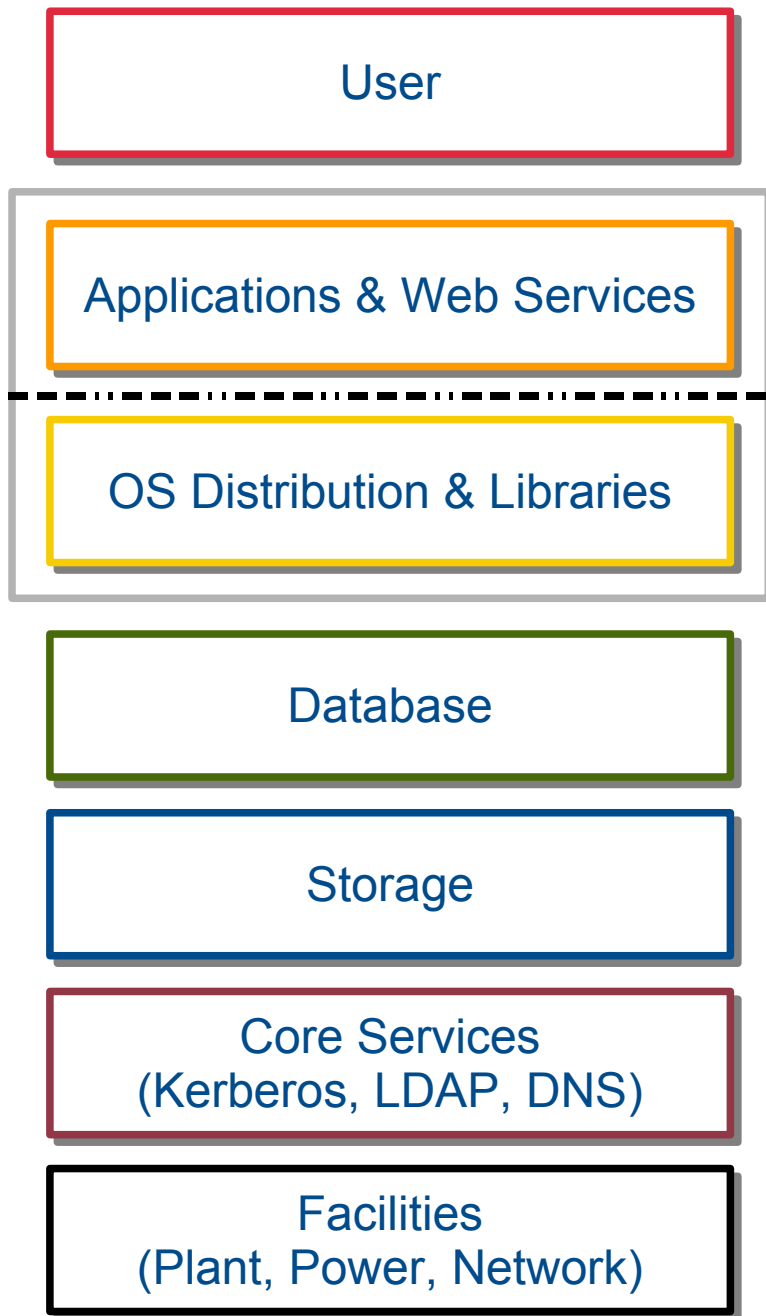
➤ **Operations**

- Communication
- Project oversight

➤ **Staffing**

- 4 – 100% (2 open)
- 6 – >½ time
- Shared legal, budget admin, CRM, other staff.

Clear roles make for good partnership.



➤ **Research owns the applications.**

- Command line and web
- Database instances and content

➤ **CIT is responsible for infrastructure.**

- Facilities and core services
- Storage, database, compute

➤ **We partner on the rest.**

- Disaster recovery planning
- ELN/e-signatures
- High-throughput screening and sequencing
- Animal facility infrastructure

Research computing is driven by needs.

➤ **Diagnostics**

- Computational discovery of markers that identify disease and enable personalized medicine.

➤ **Small Molecule Development**

- Computational chemistry design and compound screening; electronic lab notebooks.

➤ **Biomedical Imaging and Microscopy**

- Image acquisition and reconstruction to understand biological mechanism.

➤ **Structural Biology**

- Structural basis of antibody and small molecule action.

➤ **Bioinformatics**

- LIMS and scientific analysis/support.

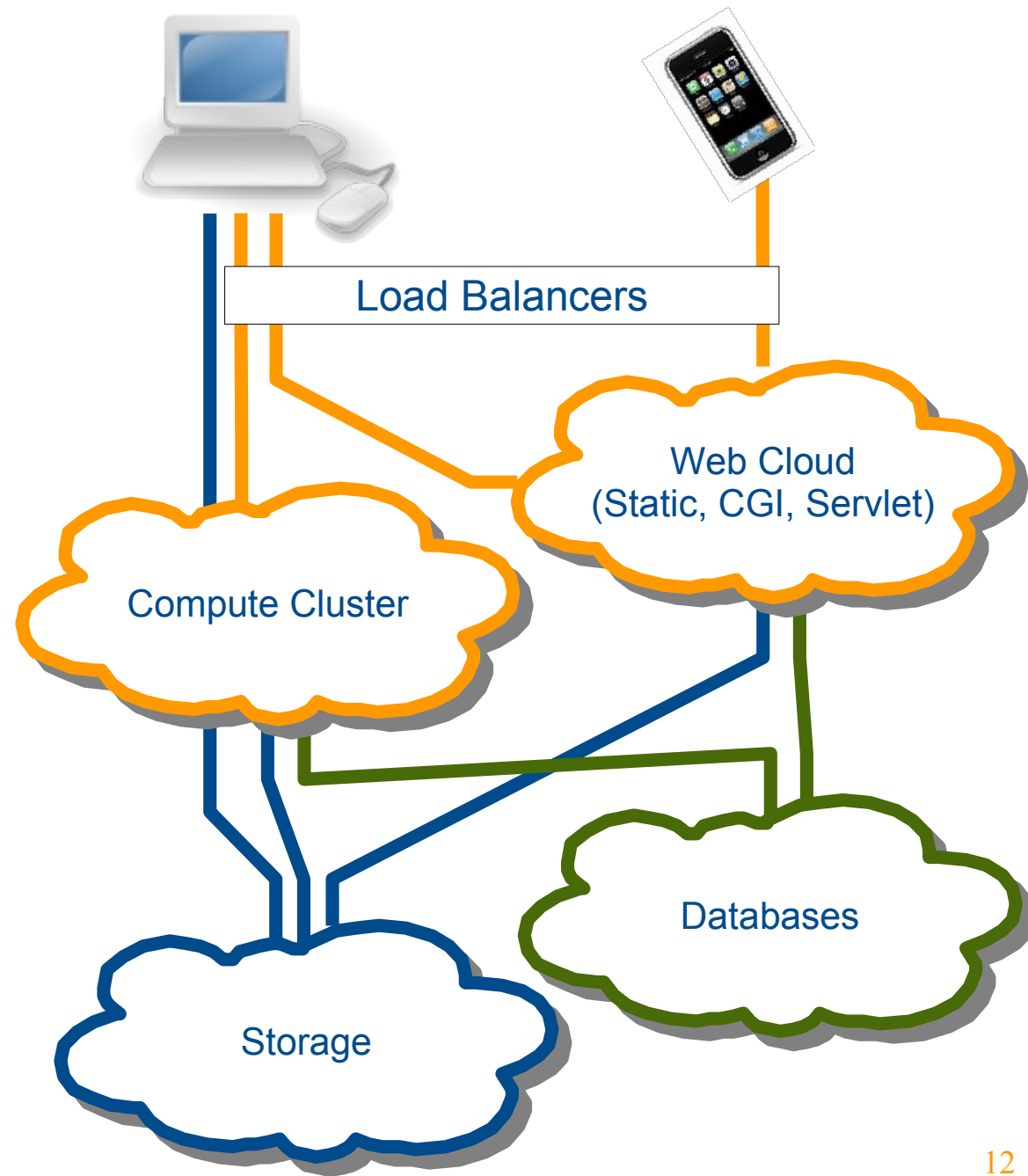
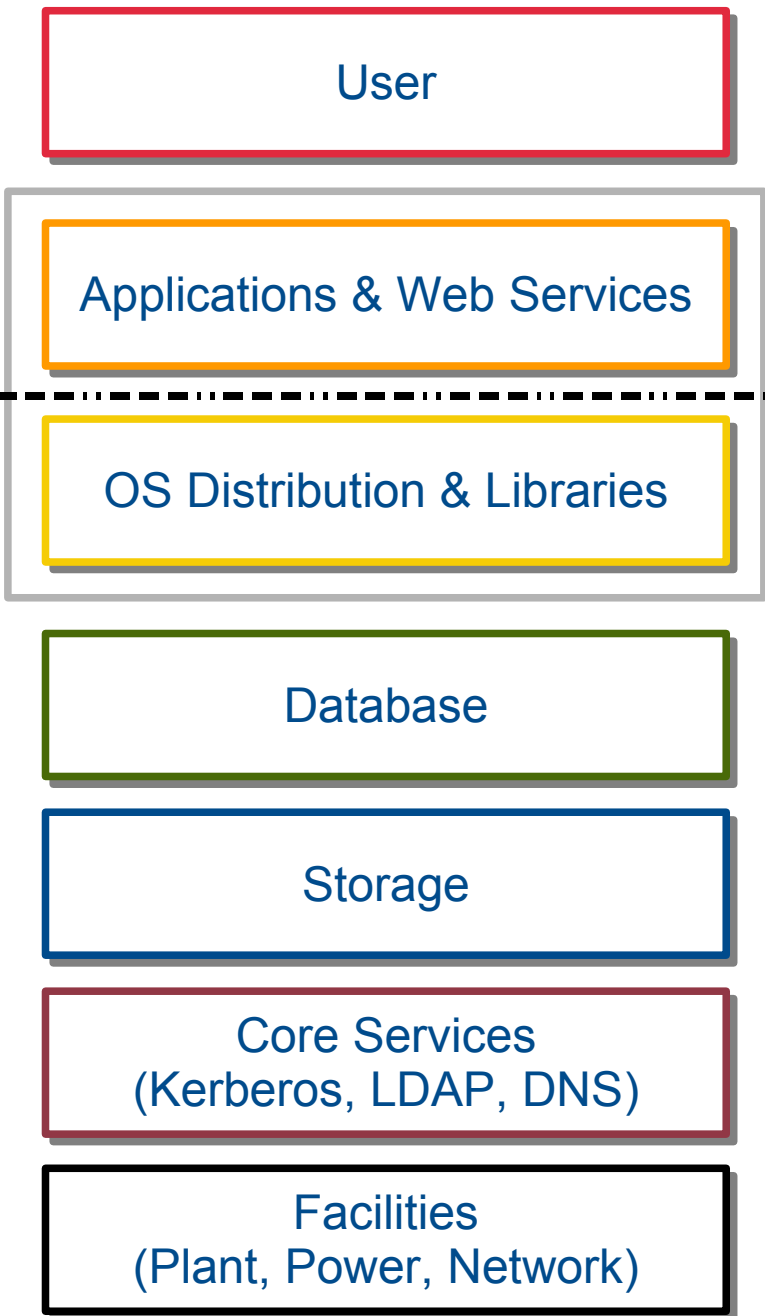
A few examples of needs.

	Expectations of CIT	2009 needs	Benefit
Users	Desktop & workstation support (Mac, Win, Linux);	Desktop application deployment	Use corporate mechanisms and unburden Research scientific personnel
Applications	Systems architecture; service installation support; operations support on major systems	Electronic lab notebook (selection, architecture, installation support)	Improve integration and searching of Research data
		Data integration and decision support tools	Increase value of existing data, improve decisions through availability of complete and timely data.
OS Distribution & Libraries	System administration	Upgrade to modern Linux distribution	Enable access to additional tools with greatly reduced effort.
Computing Hardware	Provision and operate server hardware; identify new compute hardware opportunities	Enable rare, large-scale computing needs efficiently (cloud?)	
Database (cloud)	Administer, support, and backup all databases.	Centralized user auth'n/z	Improve security policy and decrease unreliable manual efforts
		Credential-based access	Enable secure, decentralized computing
		MySQL support	Provide for multiple existing and unsupported MySQL instances.
Storage (cloud)	Provide reliable, high-performance storage; monitor and plan for growth.	Improved IO performance	
		Secure user access	
		Data triage & information life cycle management	Decrease storage costs
Core Services	Provide highly available DNS, LDAP, AD, etc.	LDAP upgrade	
		Credential-based logins	
		Group cleanup (AD-LDAP sync; consolidation; unification)	
Facilities	Maintain modern data center and physical infrastructure.		

Specialized Projects / System Planning	High-throughput sequencing architecture and support	
	Disaster recovery planning	
	Tomcat AppService planning	
	WebAuth	
	External collaboration & web hosting	

Computing Environment

Research Computing Environment



Files and databases are the foundation.

➤ **Primary NAS**

- Two active-active NetApp 6080, w/SATA & FC
- Remote mirror for disaster recovery & tiered data
- No tape
- Exceptions used as necessary

➤ **Primary SAN**

- HP EVA 8K for databases and exceptional needs

➤ **Primary Database**

- Oracle 10g on Linux (SLES10)

➤ **Gaps**

- NAS Performance
- ILM / data triage
- Virtualized storage for CIFS
- Alternative storage and database options

➤ **Shell & batch computing**

- 30 x 8-core, 64GB HP DL685 blades (Opteron)
- Altix 3700, 96 Itanium2 cores, 512GB RAM
- PBS Pro cluster scheduler
- Novell SLES10 (SLES11 coming)

➤ **Legacy:**

- Going: Tru64/alpha
- Gone: Solaris/SPARC, IRIX/mips

➤ **Gaps**

- Cluster scheduler tuning
- Missing nails for the Tru64/alpha coffin

➤ **Static and CGI**

- VMWare virtual machines running Linux, Apache

➤ **Java Application Services**

- WebLogic 9 on Solaris

➤ **F5 Load Balancer**

➤ **Gaps**

- Tomcat support
- Holistic monitoring

We got a lot right.

➤ **Security**

- Kerberos provides reliable authentication for many services, across multiple hosts, and on many platforms.
- Enables us to “push” user identity close to the data.

➤ **File system layout**

- A single logical filesystem and NAS enable data sharing in a heterogeneous environment.

➤ **Goldilocks Migrations**

- Aim for not too soon and not too late.

➤ **Cluster strategy**

- Great adoption, few complaints, no barriers.

➤ **Awesome talent**

- We've hired exceptionally talented systems architects, engineers, admins, and support staff.

From here, where?

➤ **Technology & Services**

- Organize ELN efforts.
- Implement disaster recovery hardware.
- Explore and expand hosted cloud computing
- Write documentation.

➤ **Organization & Attitude**

- Actively prioritize needs using available staff.
- Strive to use the right amount of process in the right places.
- Embrace continual change.

Planning for disasters is hard.

Research disaster recovery needs

Share Autos

File Edit View Format Insert Tools Form Help

10pt B. Abc A \$ % 123

	A	B	C	D	E	F	G	H	I	J	
1	Research Disaster Recovery Needs										
2	This is a draft. The numbers are not accurate. Do not rely on these data for planning yet.										
3				Outage Tolerance (d)		Dependencies					
4	App/Service	Group	Type	SSF Up	SSF Down	stor	comp	db	web	other	D
5	/gne/home (home directories)	RCI	Infra	7	60	x			x		
6	/gne/research/apps	RCI	Infra	7	60	x					
7	/gne/research/data	RCI	Infra	7	60	x					
8	/gne/research/web	RCI	Infra	7	60	x					
9	animal facility networking	Bioinfo	Infra	2	7	x		x	x		
10	Anydata	Bioinfo				x		x	x		
11	BMI data acquisition	BMI				x				acq computers	
12	BMI rendering	BMI				x	x			Altix	
13	Cell Bank	Bioinfo							x		
14	DNA Sequencing	Sequencing				x		x	x		
15	Genehub	Bioinfo	Tool						x		
16	GenenGenes	Bioinfo		14					x		
17	GPS	Bioinfo	Tool						x		
18	GSeqEdit	Bioinfo	Tool						x		
19	GSeqWeb	Bioinfo	Tool						x		
19	HTB Reaquest	Bioinfo	Rea Svs						x		

Research needs and Corporate needs must be balanced.

- **Modern infrastructure through rapid evolution and agile processes.**
- **Operational efficiency through standardization and consolidation.**



Acknowledgments

➤ **Steering Committee**

- Lynne Ahn
- Jeff Blaney
- Nick van Bruggen
- Chris Jones
- Cris Lewis
- Melissa Starovasnik
- Chris Wiesmann
- Zemin Zhang

➤ **RCI**

- Albion Baucom, Jim Fitzgerald, David Konerding
- +2 openings!

➤ **CIT**

- Storage: Steve Cachia, Chris Chu, Phil Seto
- A&E: Munther Megdadi, Marc Lambert
- DBA: Ben Nguyen, Jignesh Joshi
- Sysops: Kathy Rinaldi, Boman Abadan, Paul Bulanadi, Simran Hansrai, Mahwish Hamid, Michael Kennedy
- Lots more!

➤ **Elsewhere**

- Kevin Clark, Borlan Pan, Nick Skelton