Making the Web Work for Science: The Impact of e-Science and Cyberinfrastructure

CENDI/NFAIS/FLICC Workshop

December 8, 2008

Chris Greer
Director, US National Coordination Office
Networking and Information Technology Research and Development Program

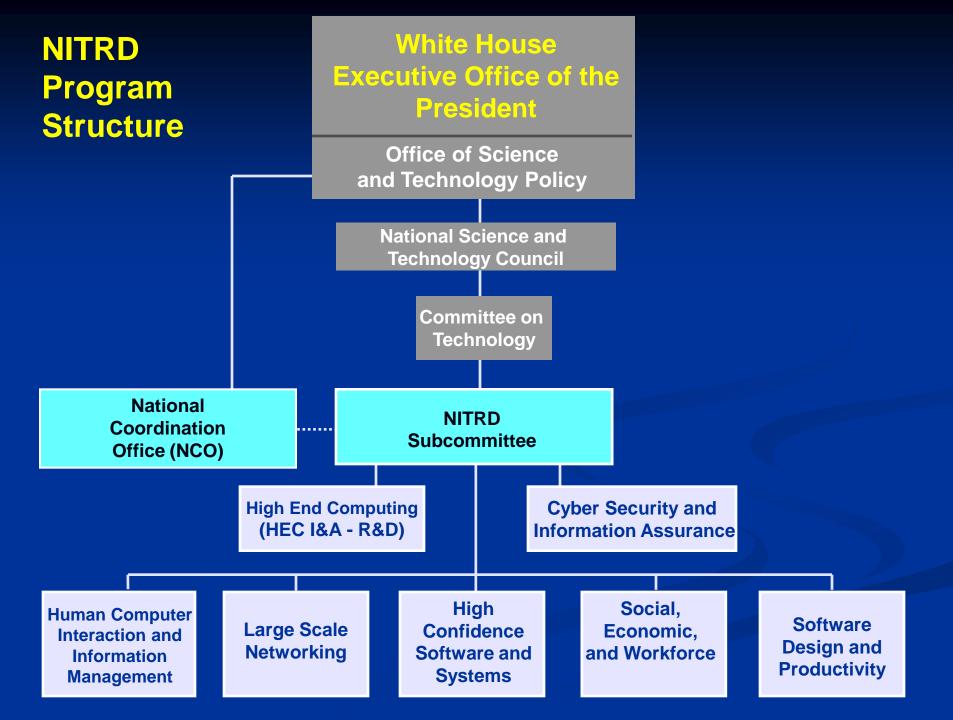
Acronyms:

NITRD

Networking and Information Technology Research and Development Program

NCO

National Coordination Office







DARPA Defense Advanced Research Projects Agency



DOE/NNSA Department of Energy - National Nuclear Security Agency



DOE/SC Department of Energy - Mathematical, Information, and Computational Science Division



EPA Environmental Protection Agency



NARA National Archives and Records Administration



NASA National Aeronautics and Space Administration



NIH National Institutes of Health



NIST National Institute of Standards and Technology



NOAA National Oceanic and Atmospheric Administration



NSA National Security Agency



NSF National Science Foundation

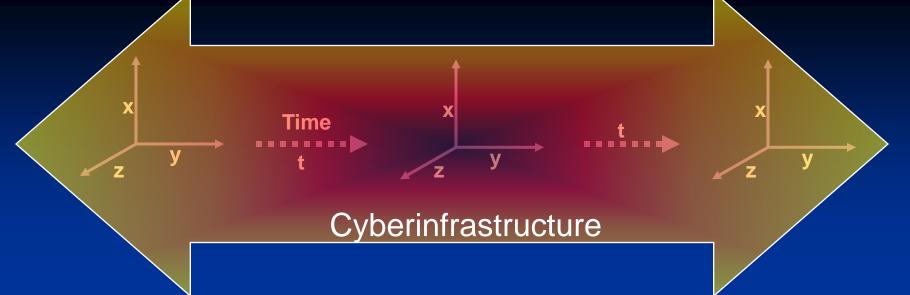


OSD and **DoD** Service research organizations, Office of the Deputy, Under Secretary of Defense (Science and Technology)

Science is global and thrives in a world that is not limited to 4-dimensions



Image: Andrew J. Hanson www.cs.indiana.edu/~hanson/





Computational capacity and capability

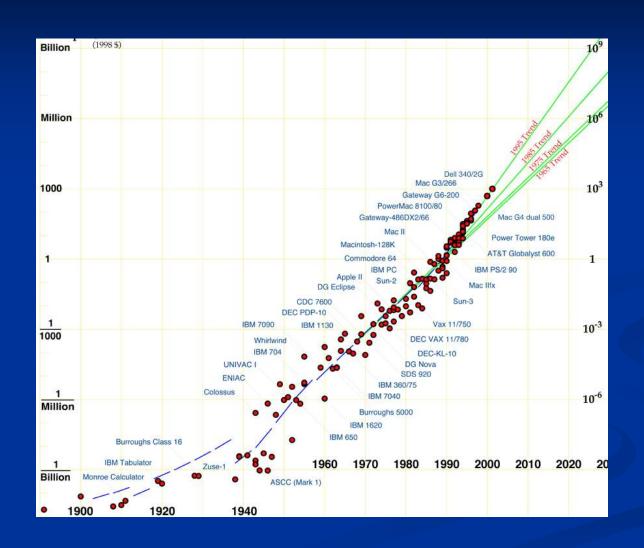


Information for innovation and discovery



Connectivity for access and interaction

Increasing Computational Capacity and Capability

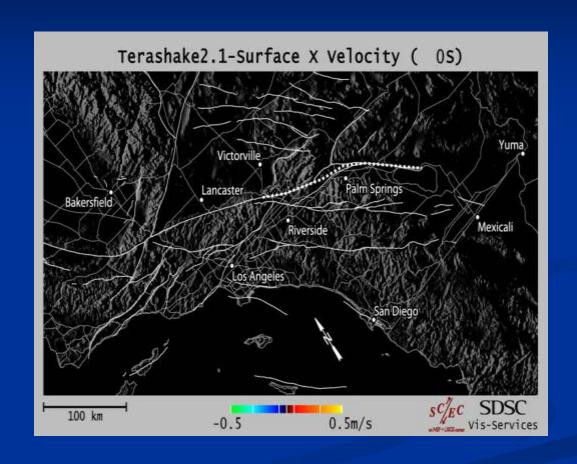


ENIAC, dedicated in 1946, was one of the first fully-functional digital computers, using 17,000 vacuum tubes for up to 5,000 addition operations per second. Today's petascale machines are designed to sustain more than one quadrillion (1,000,000,000,000,000) operations per second. A calculation these machines could complete in a week would take a machine operating at ENIAC speeds several billion years.

TeraShake Simulation

Simulation of 7.7 earthquake on lower San Andreas Fault

- Physics-based dynamic source model simulation of mesh of 1.8 billion cubes with spatial resolution of 200 m
- Builds on 10 years of data and models from the Southern California Earthquake Center
- Simulated first 3 minutes of a magnitude 7.7 earthquake, 22,728 time steps of 0.011 second each
- Simulation generates 45+ TB data



Behind the Scenes – Enabling Infrastructure for TeraShake

■ Computers and Systems

- 80,000 hours on 240 processors of DataStar
- 256 GB memory p690 used for testing, p655s used for production run, TG used for porting
- 30 TB Global Parallel file GPFS
- Run-time 100 MB/s data transfer from GPFS to SAM-QFS
- 27,000 hours post-processing for high resolution rendering

People

- 20+ people involved in information technology support
- 20+ people involved in geoscience modeling and simulation

Data Storage

- 47 TB archival tape storage on Sun StorEdge SAM-QFS
- 47 TB backup on High
 Performance Storage system
- SRB Collection with 1,000,000 files



Source: Fran Berman, Director, SDSC



......





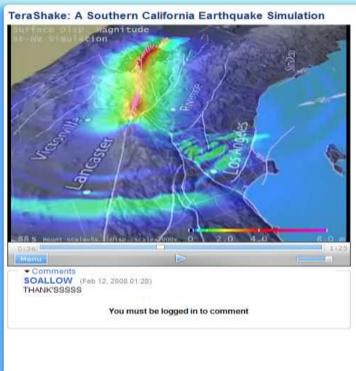


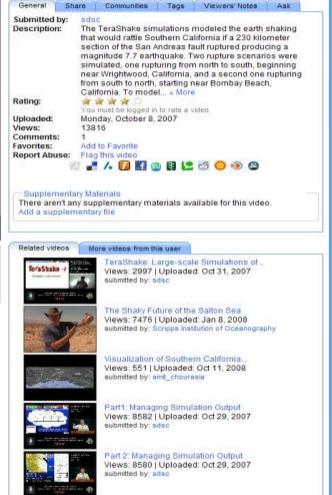


my communities

communities list

discussions list





		-	**************************************	MOTA	1	N	ARG A	11 15.440	60.933	40.699 1.0	00 44.16	N
SEQADV				ATOM	2	CA	ARG A	11 15.197	61.381	39.289 1.0	00 42.96	С
			GMCCCCCC Fronty overlable and no	MOTA	3	С	ARG A	11 15.706	60.349	38.292 1.0	00 42.40	С
SEQADV			A Critical Note of a Celiular Membrane Traffic Protein in	MOTA	4	0	ARG A	11 16.675	59.633	38.561 1.0	00 42.25	0
SEQADV			Poliovirus RNA Replication	ATOM	5		ARG A				00 43.62	С
SEQADV			and the second s	ATOM			ARG A			38.289 1.0		С
SEQADV			Tribute recent of sorig and profess comme, needed comment of means, despect, comment of comment, 41,44,4000 of comment, 41,44,4000 of comment, 41,44,400 o	ATOM	7		ARG A			37.971 1.0		C
SEQADV			Bodinie Stuttelie. Odb. SPiso feits. Neco	ATOM			ARG A			39.166 1.0		N
SEQADV				ATOM	10		ARG A				00 45.13 nn 46 91	С
SEQADV				ATOM ATOM			ARG A ARG A			38.598 1.0 40.571 1.0		N N
SEQADV			replication is abilitized by brivhelds A 6FA, implicating some componential of the cellular secretary pathway in visus growth. Formation of classification works induced by appropriate of visit proteins was not inhibited by SFA, but they were	ATOM			LEU A				00 45.04	N
SEQADV			operatory of white principly to develope the control of virtue expectation, for interest expectation, white the control of control of the con	ATOM			LEU A			36.080 1.0		C
SEQADV			in CRFT that allows provide of policytics in the presence of RFA, increation between CRFT and visal protein 3A determined the contract of CRFT and visal protein 3A determined the contract of CRFT and visal protein 3A determined.	ATOM			LEU A			35.251 1.0		c
SEQADV				ATOM			LEU A			34.519 1.0		ō
SEQADV			Althor has been relieved at 1984 as the Section State State of Section	MOTA			LEU A			35.179 1.0		c
SEQADV			Resident Advancy (A. 2000, Rockyted Children (E. 2000, Published Sourcement (E. 2000)	ATOM	17		LEU A			33.941 1.0		С
SEQADV				ATOM	18	CD1	LEU A	12 14.795	56.773	34.381 1.0	00 37.07	С
SEQADV			Contract of the state of the second part and the second contract and the secon	MOTA			LEU A			32.929 1.0		С
SEQADV			* Carell observation (continue Springer Letter for Manager Life Springer). The Springer, The Springer	ATOM			PRO A			35.344 1.0		N
SEQADV			Parent And the Control Manager Control of the Control of Control o	MOTA	21		PRO A				00 38.54	C
SEQADV			Coll. Sent reason management of release animales.	ATOM	22	С	PRO A	13 18.523	59.598	33.005 1.0	υυ 37.53	C
SEQRES	1 A		All brown profiles report BNA or purification for growters in proceedings with report or region and the property of the contraction of the contrac	ATOM ATOM						USFF is Required for Volume in	National	0 C
SEQRES	2 A		replication complesses our transference in Indianal to force content advantages. Months are sufficiently than transmit to bood compless recognition of preference in preference in preference in the force of the second state of the compless recognition of	ATOM				777				C
SEQRES	3 A		The state of the s	ATOM		3	A		В			c
SEQRES	4 A		The monitories with constrainty powers replicating SNN form using (7). They also legalled with the electronistic field pullentum confidence of the pullentum of collection	ATOM			4000	1	200	- 1		N
SEQRES		16:	turnedas [1]. the remain of the [2] signer that the COFF-dependent turnedas [1]. The remain of the [2] signer that the COFF-dependent turnedas [2].	ATOM			100		200			C
SEQRES	6 A		of send, non-emitting of position around RNA vision of the fact includes a send of send replacation completes. However, Sologel et al. box of send, non-emitting of position around RNA vision of the include a send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes. However, Sologel et al. box of send replacation completes.	MOTA		-		T SHOULD THE SHAP OF BEING	1	# \$100 A CASTAN # \$100 A CASTAN # \$40 A CASTAN		c
SEQRES	7 A		more discovering the adjustment of a clock 1900. The second of the country of the land	MOTA		1.	Amm	Fig.	#HHHHH	* BAT do not The		0
SEQRES	8 A		to the color of th	MOTA		1	- dilli	THE T		9		С
SEQRES	9 A			MOTA		1	11111	100	1111111	V-3.		0
SEQRES	10 A		makes of the proper with surface enteriors for excepted. Instantial, unifor Execute of CORE, and with the	ATOM								N
SEQRES	11 A		process, recovery for packaging of progress NNA for dispersable completing on committee in MFA (12.1.14), suggesting that these for evaluation are former for the Paris Fig. 10. Information on the processes to our falls account for all orbits a pathons.	ATOM		Nger.	B. Waricks Format	d in the presence of BFA do not support to the pursue of ST and 7 a give BFA as industri	siral BNA replication, A. I.	Polis Arolla epitos-Will our sett. In Mill Inschiller and Repolitic	Bestell in	C
SEQRES	12 A		without patients made in right and reason recognisation of regulated by publishers for its registration. Bifs is a fample should be considered as the contract of regulation of the contract o	ATOM				the guarantees AS and 7 a girls M.A. as included as a finite series of the product of the produc		in MSA speciment, and the Street buffer. The control sample & The same in	teler publishy legand married	C
SEQRES	13 A		chains of lightly suscepted which of homogeneous size which GTPain, ArX, which are key regulators of the offshir accross-	ATOM ATOM		allely fire to	171 bornel per 19	true constrong a pendicu in the polymeter p HCTs, glob				0 C
SEQRES	1 B		myKacken complete have been also associated with all of mucholik coducing to GTP, and open conformational changes. SER	ATOM		the bear	racione tally ober	other steal process and a likely involved or real products that may algorithms in	Salabition of the self-	False according pathway was mag or replication of the other in an area	general to lead how.	C
SEQRES	2 B		QQ First Fallingers emorphisps fragerousy 9 fearerise 2008 follower 6 more 11 of 000316	ATOM		\$15,505	for nations of the Our province that	as intractions with sollabor promise as downed that the amount of AH-GTP	fast stay believed to be if Although the TA-2 see	disponentie for explication in self- mention, was Figure to verticals	Brodher. de reduce	0
SEQRES	3 B		. M TYR	ATOM		results (moral)	GEF winds in the	durante of february (17); the at least of a set yells in our picklines. Marsons	publications's attitus to be to true to interfere regulation	brists official conversion (25%, it must worky mility propagations of the year	e reparted inputs of the	0
SEQRES	4 B			ATOM		Featiles descript	est all demonstrated plan with remotera	d Bal corporate of COPI costs, whose were in directly dependent on CRP1-	enhair, Or the other has between two CVBs,	and, when the entroperating map, , the remains was like parting on	epation total	N
SEQRES	5 B			MOTA		in the board	d Art activation, an	or quantum with epitegion complexed of picorespiracities to evaluate to BEN of Decoglish with aboutled the COSS	of web lighted.	of the apparent amountain terms		Ċ
SEQRES	6 B			ATOM			or a a fear fact	of Donogloba with about fired the COSP over spectral for growth of another brought is Green [18]. Three data do	replication, GMT north	E-dependent registers pike in y enably participates in the form on ERGOS oranges, and y	traction of	С
SEQRES	7 B			MOTA		THE RES	spice the resion the	bromplets C rives [16]. These data do not GRF1 activity is teledisted in telecond and here. Service show that GRF1 in	mplication completes t	the ERGE symmen, and y from an experience strain congruently and teachs. The	rience shar	0
SEQRES	8 B			MOTA		19456	aly the pollowing a holly arrive SMF1 o	ephiator and that only engineers of can receive policytes replication to the	they provide he must close structure. In pul-	hat he remodeling that unstale districted colle, however,	Agen, des L. deckgo	c
SEQRES	9 B			ATOM		Service of the last	or of RPA. Moreover obtains of GRF1-sky	or to decemblish resident of SA reads. Epitalest activates of Art is sets. These	central infection very rep cellulat conservation into a	apidip resilis in complete receptor apodis marketanen seekis, sa	nimites of a dest ER,	C
SEQRES	10 B			ATOM		Service of	from the run Ki	to mer differt to records with the Supposed Mayoney 2011 one; the	WHICK: and teles are	retters as as longer distri- sacistics palmer step cold offs [NE, Therefore informal or	conder Till.	c
SEQRES	11 B			MOTA		-distriction	others promisely of the alsonor of take	rather than intelligen of Art arthuran- er real persons, Collain pain make in-	regardly from the serviced of	ameghological religions: for time	measure of	С
SEQRES	12 B			ATOM ATOM		a sale not	dated those apparen	carrectly to progress to enterprise mits conflicting results. Our data sta-	Alt remot of GW1	porting the objects one charmains is developed the a soluble pr	produte as	N C
SEQRES	13 B			ATOM		te info	management of the	to 14.5 extract two market a result 2015, is off resolve one market state to provide with recolumns. We propose	mark hallowed raffs. Apply	is telephysiciscis from the set of quench GBF1 in policidate of a release through its present too	cells 8 no	c
FORMUL	3 1		*188 (H2 O)	TOM		this be	mention of SA will be the same of the same	with GBP1 obserts to putsing there in	such as beeding to Rabit unitarial on remodeled a	ifh and/in p115 D3.30% but et p I montionie arazum decept le	promise to	o
HELIX	1		EU A 52 GLY A 66 1 15	ATOM		When	dian ofere it fan melde planteren	colors in polestrar and likely scho- red replication. The discrete in real	pote promis SA, while it with SA and develop to	the ACC GRIFT transports and the transport of the second o	to inscring Euleriene	c
HELIX	2		HE A 118 ARG A 132 1 15	ATOM		replicati outside	man completes who me. Statistics of	with Holy much in infoliose of unfolio- cillator members highly to reduced	bifution is known to be and consequent bubbles	inher tapid degralation of rush in of the nuclear analogic that of	drut prior. of two propri	N
HELIX	3		HR A 136 ASP A 156 1 21	1 ATOM		restand	d retirate of epichine	on the melant of integral with an well as set (20,26). Thus, effective subserviors of	marker comban to last	salelle pronter to freely pron th directions [28,60]. Assolver as	eri of data	С
HELIX	4	4 LE		ATOM		control of oil	ing grooms mylton	provide a double towards for the years in whom to the yell as well as relativity a	dispersional year in public	Whenever but therefor to familie replication is not remo-	minimum of	С
HELIX	5		HE B 118 GLU B 134 1 17	11011		patron	ey desponant for the	n regarier's defense against beforeign.	received for rater	a latter president of the m	mplication	0
HELIX	6		HR B 136 HIS B 159 1 24	11 011		49	Not Pellingen m	engined arms 0	Y Brond.	ier 200 Solate 4 Sole 11	+1100.014	С
SHEET	1		GLY A 44 LYS A 46 0	ATOM			ARU A			27.733 1.0		C
SHEET	2		ALA A 17 ILE A 23 -1 N ALA A 20 O GLY A 45	ATOM	60		ARG A			26.207 1.0		C
SHEET	3		ASN A 83 ILE A 91 1 O TYR A 88 N ILE A 23	ATOM			ARG A			25.853 1.0		N
SHEET	4		SER A 71 ILE A 79 -1 N LEU A 75 O VAL A 87	ATOM	62 63		ARG A			25.819 1.0		C
SHEET	_1_	B 3	GLU A 25 MET A 26 0	MOTA	ರಿತ	MHI	ARG A	10 33.262	37.321	26.076 1.0	UU UI.05	N

CRYSTAL STRUCTURE OF PUTATIVE NUDIX HYDROLASE FAMILY MEMBER FROM CHROMOBACTERIUM VIOLACEUM

J.B.BONANNO, J.FREEMAN, K.T.BAIN, J.DO, R.ROMERO, S.WASSERMAN, J.M.SAUDER, S.K.BURLEY, S.C.ALMO,

NEW YORK SGX RESEARCH CENTER FOR STRUCTURAL GENOMICS (NYSGXRC)

27-OCT-08



Source: RCSB Protein Data Bank; www.rcsb.org

Redefining "Computer"

 "All the calculations that would ever be needed in this country could be done on three digital computers."

-- Douglas Hartree, Cambridge, 1954

•"There is no reason for an individual to have a computer in their home."

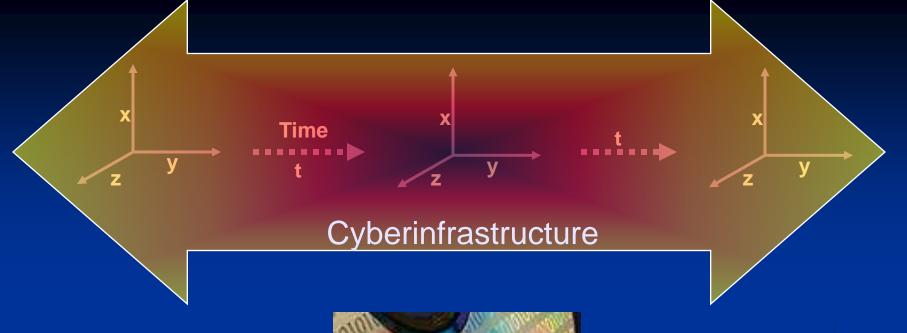
-- Ken Olsen, DEC, 1977

•"For the full year [2007], IDC said 269 million PCs were shipped worldwide"
-- International Herald Tribune, January, 2008.

•"In a sense, there are only five computers on earth."

-- Yahoo Research Chief Prabhakar Raghavan, December, 2007.

•"...some researchers at IBM believe that five computers may be four too many."
-- Nick Carr, The Guardian, February 21, 2008.



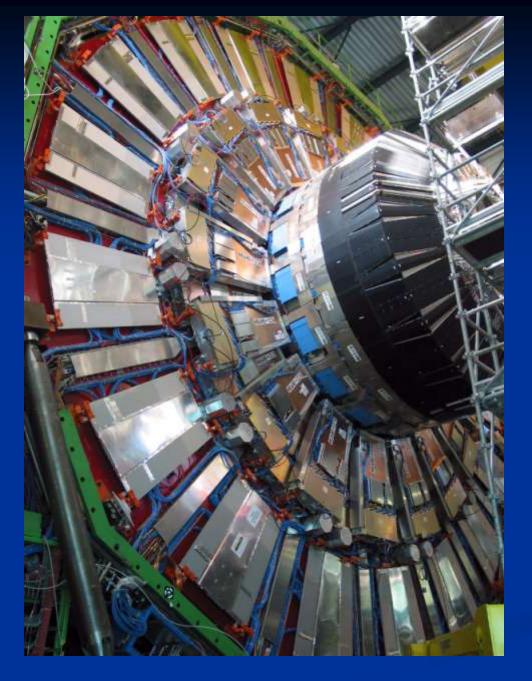


Information for innovation and discovery



"Sometime in the 2010s, if all goes well, the Large Synoptic Survey Telescope (LSST) will start to bring a vision of the heavens to Earth. Suspended between its vast mirrors will be a three billionpixel sensor array, which on a clear winter night will produce 30 terabytes of data. In less than a week this remarkable telescope will map the whole night sky And then the next week it will do the same again ... building up a database of billions of objects and millions of billions of bytes."

Nature 440:383



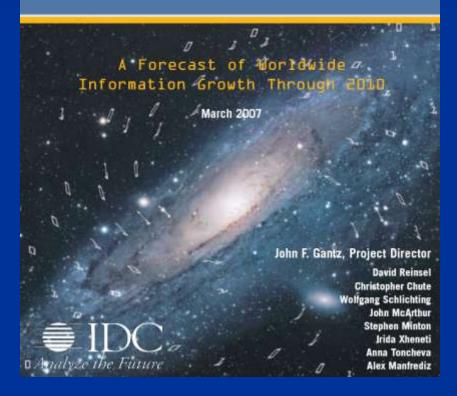
Large Hadron Collidor

Physicists will use the LHC to recreate the conditions just after the Big Bang, by colliding two beams [of hadrons] head-on at very high energy.

When LHC begins operations, it will produce roughly 15 Petabytes of data annually, which thousands of scientists around the world will access and analyse ... The mission of the LHC Computing Project (LCG) is to build and maintain a data storage and analysis infrastructure for the entire high energy physics community that will use the LHC.

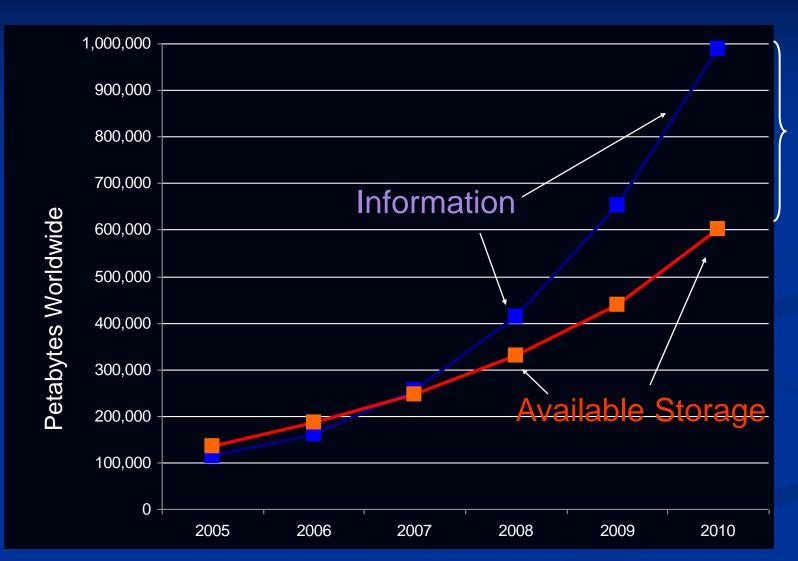
An IDC White Paper - sponsored by EMC

The Expanding Digital Universe



"In 2006, the amount of digital information created, captured, and replicated was 1,288 x 10¹⁸ bits (or 161 exabytes) ... This is about 3 million times the information in all the books ever written"

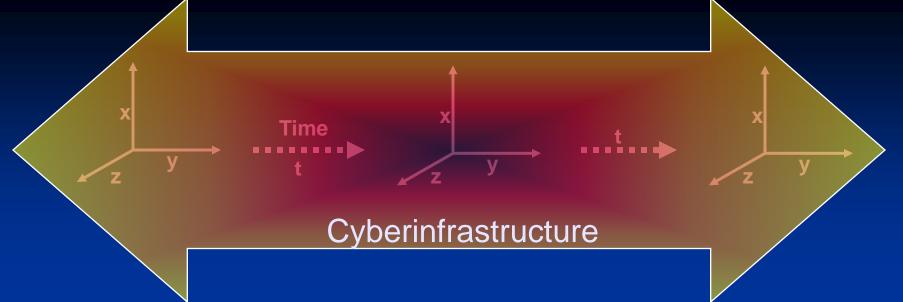
Information And Storage



Transient information or unfilled demand for storage

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

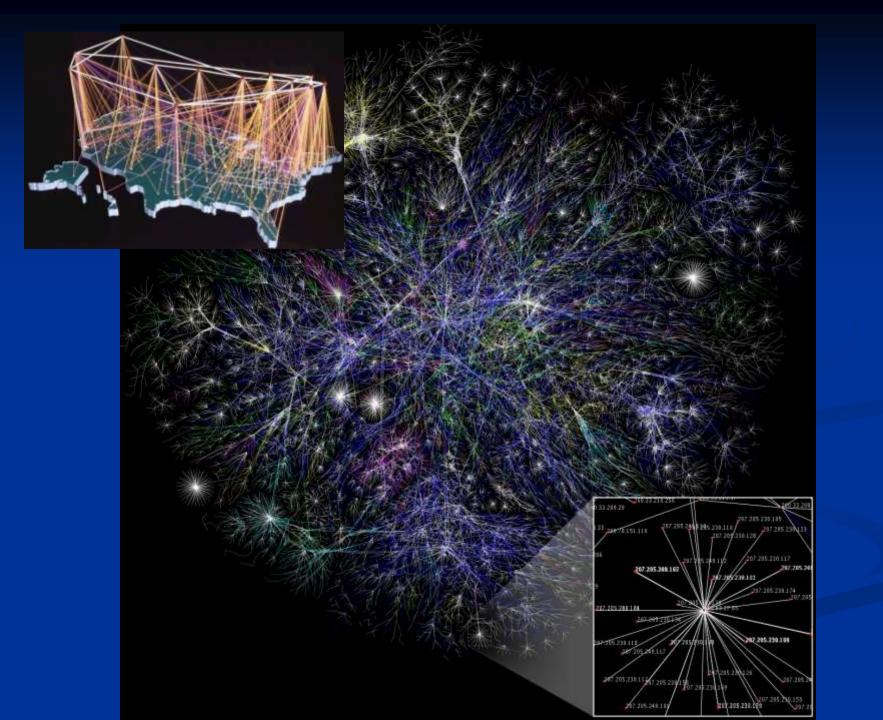
Google's founding philosophy is that we don't know why this page is better than that one: If the statistics ... say it is, that's good enough. No semantic or causal analysis is required. That's why Google can translate languages without actually "knowing" them (given equal corpus data, Google can translate Klingon into Farsi as easily as it can translate French into German).



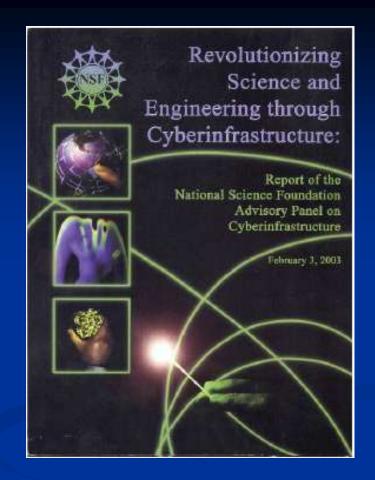


Connectivity for access and interaction

The Department of Defense's ARPANET project, launched in 1966 to explore methods for "resource sharing among computers", initially connected 4 nodes. Today's Internet links more than 1.4 billion users over more than 200,000 networks worldwide; with 14 new users added every second.



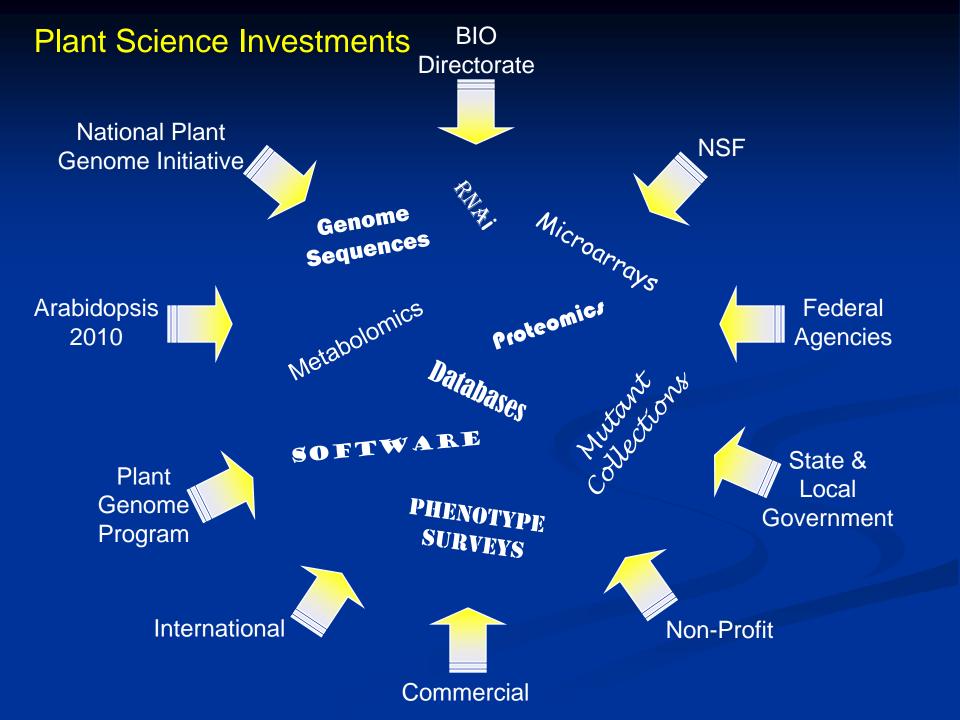
"New types of scientific organizations [that] serve individuals, teams and organizations in ways that revolutionize ...

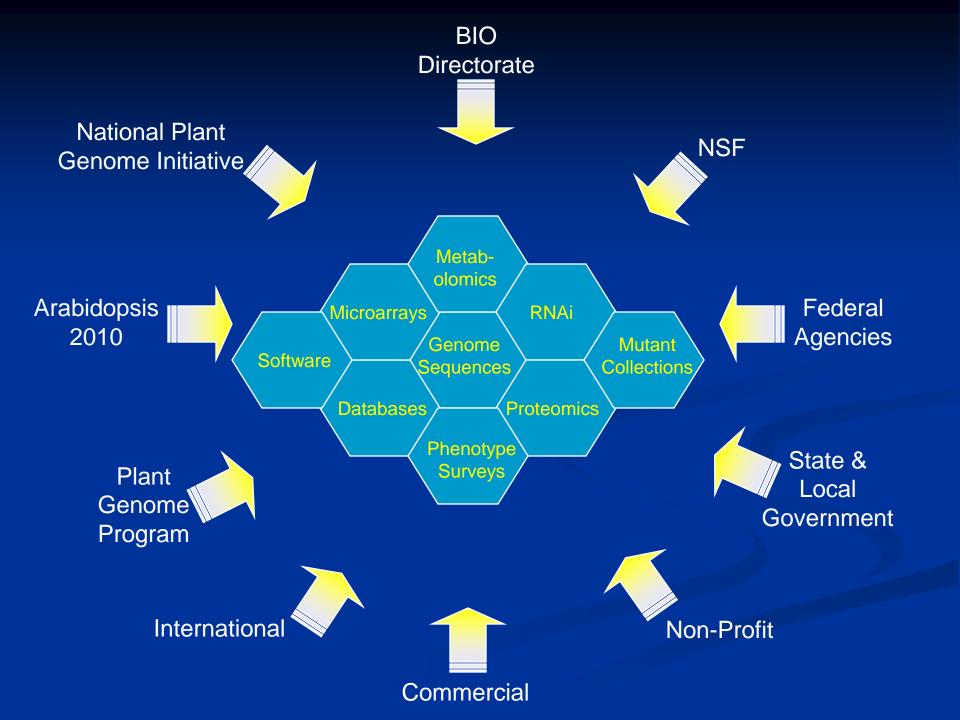


... what they can do, how they do it, and who participates."



- Richard Jorgensen, Collaborative Director, Plant Sciences/BIO5, University of Arizona
- Steve Goff, Director of Community Interactions, University of Arizona
- Rick Blevins, Director of Cyberinfrastructure Development, University of Arizona
- Martha Narro, Director of Education, Outreach and Training, University of Arizona
- Greg Andrews, Team Leader, Computing Infrastructure Team, Computer Science, University of Arizona
- Sudha Ram, Co-Team Leader, Integrated Solutions Team, Management Information Systems, University of Arizona
- Lincoln Stein, Co-Team Leader, Integrated Solutions Team, Bioinformatics, Cold Spring Harbor Laboratory
- Steve Rounsley, Team Leader, Synthesis Activities Team, BIO5, University of Arizona
- Vicki Chandler, Team Leader, Education, Outreach and Training Team, Director of BIO5,
 University of Arizona
- Susan Brown, Team Leader, Social Science Team, Management Information Systems,
 University of Arizona
- Kobus Barnard, Computer Science, University of Arizona
- Travis Huxman, Ecology and Evolutionary Biology, University of Arizona
- Dan Stanzione, Director, Fulton High Performance Computing Center, Arizona State
 University
- Ann Stapleton, Biology, University of North Carolina-Wilmington
- Matt Vaughn, Bioinformatics, Cold Spring Harbor Laboratory
- Doreen Ware, Bioinformatics, USDA/Cold Spring Harbor Laboratory





A Discovery Environment





Web Services Interface

Distributed Content Creation



(e.g. community annotation)

Community Interoperability frameworks







Web 2.0





Campus Clusters; **TeraGrid**



Machine Level Interface (API)





Software



Algorithms



Visualization

Computational tools & resources

Discovery Landscape

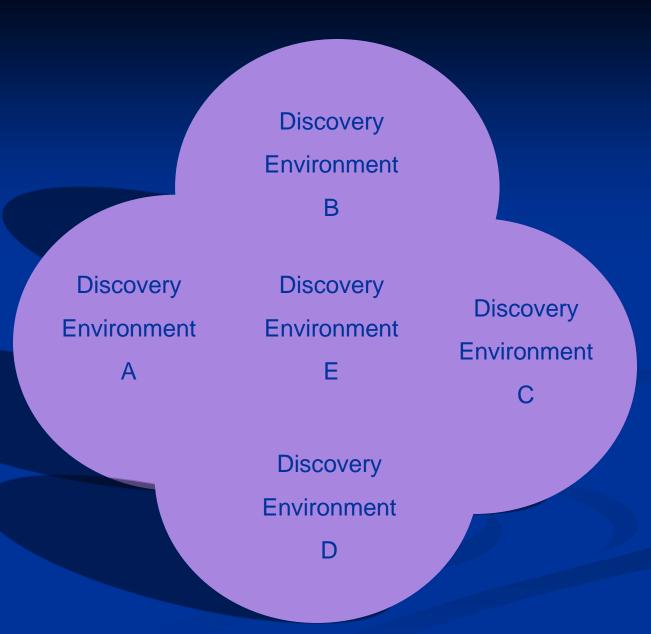
Distributed

Integrated

Open source

Evolvable

Extensible



The Future: Next Generation Networking and Information Technology

Emergent Cognition

Beyond Virtual

Pathways to Trust and Confidence

Emergent Cognition

Harness new levels of intelligence, intuition and perception that emerge as greater than the sum of the parts in the union of cyber, human, and social capabilities

Example: Enable the combined use of logic, analysis, and modeling; expert intuition and perception; and market judgment and crowd wisdom in understanding the properties and behavior of a global economy.

Beyond Virtual

Extend the envelope of capabilities by making the virtual and the physical one.

Example: Cognitive computing – Provide the ad hoc capability to access the full sensing, information, and computational capabilities of the every day objects around you to respond to an emergency or to an ordinary need.

Pathways to Trust and Confidence

Ensure the value to society of cyber systems is not limited by a lack of trust and confidence.

Example: Provide capabilities for identity management, information security, privacy protection, and tiered access to enable an electronic health records system that can be confidently used for individual patient care, public health, and biomedical research.

eLibrary - Single point of access connecting resources worldwide:

Digital and physical repositories

Computational capabilities

 Digital environments for access and interaction by people, machines, and organizations

Software tools and resources

Contact information:

greer@nitrd.gov