Big Data, Big Network:

Big Opportunities for Collaboration between Mexico and the United States

10-11 Octubre 2013 - Ensenada, México

BigData, Big Network

Areas of focus for the workshop included but were not limited to science gateways, bioinformatics, oceanography, seismography, astronomy, nano-materials sciences, digital media, and scientific instruments. **All areas of research that involve current or potential cross-border collaboration between Mexico and the United States were welcome** if they are or can be enabled by the 10-Gigabit network connectivity between Mexico and the US.

List of Abstracts for BDBN Workshop

* * * * *

Focus Area: Science Gateways

Submitter: Nancy Wilkins-Diehr, UCSD

Summary: This presentation will report on a 2-year study entitled “Opening Science Gateways to Future Success” that evaluated the characteristics of successful gateways and analyzed sustainability issues. That work served as a foundation for the Science Gateway Institute proposal.

2) Title: HUBzero: An Open Source Platform for Building Science Gateways **32**

Submitter: Michael McLennan PURDUE University

Summary: HUBzero is an open-source software platform used to create science gateways or “hubs” for scientific collaboration, research, and education. It has a unique combination of capabilities that support science and engineering. A little like YouTube.com, HUBzero allows people to upload content and “publish” to a wide audience, but instead of being restricted to short video clips, it handles datasets, analysis tools, and other kinds of scientific content.

3) Title: nanoHUB: A Science Gateway with global usage and impact **41**

Submitter: Gerhard Klimeck, PURDUE University

Summary: This Presentation will provide an overview of these processes and their impact as they are supported on nanoHUB.org today. Annually, nanoHUB provides a library of 3,000+ learning resources to 260,000+ users worldwide. Its 270+ simulation tools, free from the limitations of running software locally, are used in the cloud by over 12,000 annually. Its impact is demonstrated by 1030+ citations to nanoHUB in the scientific literature with over 11,000 secondary citations, yielding an h-index of 50, and by a median time from publication of a research simulation program to classroom use of less than 6 months. Cumulatively, over 19,000 students in over 1,000 formal classes in over 185 institutions have used nanoHUB simulations.

4) Title: Towards a long-term e-Science infrastructure in Latin America **80**

Submitter: Jesús Cruz Guzmán, UNAM

Summary: The collaborative experience in Latin America had built an e-Infrastructure, but the fast emergency of new technologies and the difficult in using for many users present a big challenge for their support. Science Gateway is a way of provide access to many services, one effort in LA is realized by the SCALAC (Servicios de Computo Avanzado Para Latino América y el Caribe) community, organized inside CLARA Net, and other similar services for Mexican Institution for Research and Education. The principal aim is to build advance computing services based on standard technologies, with flexibility adapted to the community requirements. The science gateway is the bridge between the scientific applications and e-Infrastructure resources devoted to support the Virtual Research Communities. Here the services architecture is presented and the status of the services supporting the work of the Virtual Research Communities.

5) Title: A Standard-based Science Gateway Framework to Seamlessly Access HPC, Grid and Cloud Resources Distributed Worldwide **100**

Submitter: Roberto Barbera, INFN Italy

Summary: Proposes to present the Catania Science-Gateway Framework and the results of the program outlined by the CHAIN-REDS project to demonstrate standard-based interoperability amongst high performance computing, grid and cloud resources distributed in Latin America, in the US and the rest of the world and based on several different middleware stacks.

* * * * *

Focus Area: Digital Media

6) Title: Big Data, Big Networks, Big Displays -- the Next WAVE **129**

Submitter: Tom DeFanti, UCSD

Summary: Video wall collaborations between CICESE and UCSD have been ongoing for several years. At UCSD, both 2D and stereo 3D walls are in use, displaying big data up to 64 megapixels resolution. Specific effort has been invested to optimizing high-speed local and distance file serving and collaboration using multiple 10Gb/s and 40Gb/s networks and software tuned to synchronized image sharing (SAGE), extremely high-resolution static and streaming image viewing (MediaCommons), and immersive virtual reality experiences (CalVR), as well as to driving surround sound and focused SoundBender audio speaker arrays. Recent results adapting flash memory big data technology championed by the San Diego Supercomputer Center to "FIONA" PCs driving 2D/3D big data displays locally with 40Gb/s network interfaces to impedance match multiple 10Gb/s connections, along with their applications will be presented, with focus on omics and archaeology, as two examples with great cross-border potential. The latest big displays at UCSD, the WAVE and WAVElet, and use of emerging UHDTV (4K) panels will also be described.

Submitter: Laurin Herr, Cine GRID

Summary: This session will introduce CineGrid, a non-profit membership organization that has nurtured a community of members from many countries over the past 9 years who collaborate to promote research and exchange of ideas in this area. This session could also include a panel discussion with participants from Mexico and the USA (to be decided) who would explore potential future projects between the two countries using a 10Gbps research network link.

* * * * *

Focus Area: Oceanography

Submitter: Luis Farfán, Ismael Villanueva, Julian Delgado, CICESE

Summary: We studied ten tropical cyclones that, during the period 1996-2010, made landfall in northwestern Mexico from the Eastern Pacific Ocean. Our goal is to present a synthesis of the predicted tracks, wind fields and rainfall patterns around the landfall area. The model simulations start in advance of the actual landfall and results are also used to make an analysis of the interaction between the incident cyclone and coastal topography. Datasets issued by the National Hurricane Center are used to compare tracks and intensity with respect to observations as well as the official forecasts issued on a real-time basis.

9) Title: Transport and dispersion of hydrocarbons in estuaries of the Gulf of Mexico: an example of a proposed data-intensive application ….. **229**

Submitter: Arnoldo Valle-Levinson, UFL

Summary: The Gulf of Mexico Research Initiative was established, after the Deepwater Horizon oil spill in April 2010, to improve understanding of hydrocarbon transport in the Gulf of Mexico. In response to an upcoming call for proposals, a group of researchers from U.S. and Mexico plans to create a Research Consortium. This Consortium will concentrate in the study of hydrocarbon transport in estuarine systems around every Gulf State: Florida, Alabama, Mississippi, Louisiana, Texas, Tamaulipas, Veracruz, Tabasco, Campeche, and Yucatan. Participant researchers will not only come from these Gulf States, but also from other regions in the US and Mexico. Investigations will include field measurements and numerical simulations. Field measurements will consist of surface drifter deployments, dye dispersal studies, and shipboard measurement of environmental variables. Measurements and simulations will produce many Terabytes of data that will require effective storage, transfer, and access for a seamless international collaboration. The participation of CICESE and XSEDE in this Consortium will ensure such a seamless and effective international collaboration. This contribution will present the idea of the Consortium to seek feedback and possible strengthening of the collaborative effort.

10) Title: AMLIGHT, Simulation Datasets, and Global Data Sharing …… **242**

Submitter: Jean-Bernard Minster SCRIPPS, John J. Helly, UCSD, Steven Day, San Diego State University, Raul Castro Escamilla, CICESE, Philip Maechling, Thomas H. Jordan, Southern California Earthquake Center, Amit Chourasia, San Diego Supercomputer Center

Summary: The advent of continental scale and intercontinental high-bandwidth networks opens up new options for the architecture of ICS U World Data System. No longer must large datasets be duplicated at many nodes and, in the concept of a system of systems, server nodes and client nodes can in fact be vastly different in terms of their scope, usage, and commitments, and may not necessarily maintain vast data holdings. At the same time, consideration must also be given to failure-modes and fault tolerance of increasingly critical network infrastructure. AMLIGHT should be a very practical testbed to explore the usefulness and applications of such concepts. We propose to initiate a broadly based discussion of these topics in conjunction with other stakeholders worldwide, especially in Latin America.

* * * * *

Focus Area: Nano-materials Science

11) Title: 1H-MoS2 nanoparticles grown on graphene and 1H-BN monolayers **260**

Submitter: Donald Galvan, G. Alonso, S. Fuentes, UNAM

Summary: A theoretical study has been performed on a sandwich made by MoS_2 nanoparticles grown on graphene and BN monolayers. The calculations reported in this study have been carried out by means of the tight binding method within the extended Hückel framework using YAeHMOP (Yet Another Extended Hückel Molecular Orbital Programs) computer package. The analysis performed includes Energy Bands, Total and Projected Density of States and Mulliken Population Analysis.

* * * * *

Focus Area: Bioinformatics

12) Title: Comparative Human Microbiome Analysis **310**

Submitter: Larry Smarr, Calit2, Weizhong Li, UCSD

Summary: We are carrying out very deep metagenomic sequencing of human gut microbiomes from healthy subjects and from people with the autoimmune Inflammatory Bowel Disease. We compare one subject with IBD to metagenomic datasets downloaded from the NIH Human Microbiome Project repository, including 35 healthy subjects and 20 with IBD. We also analyze the changes in this one subject over multiple times, including comparing before and after drug therapy. The dataset of Illumina short reads for one person is ~10GB. The total comparison dataset contains ~0.5 trillion DNA bases. These Big Data had to be moved across the network to the San Diego Supercomputer Center where over 200,000 cpuhours were consumed in the analysis and then back to Calit2 where a 64 megapixel wall was used for visual analysis. This approach could be extended for cross-border comparisons of human gut microbiomes to examine differences in food intake and various disease states.

13) Title: Protein-protein interactions on viral self-assembly analysis **323**

Submitter: Mauricio Carrillo-Tripp, LANGEBIO, Vijay Reddy, SCRIPPS

Summary: Our current collaboration involves the generation, and eventual transfer, of Big Data related to the atomic structure of viral capsids, the protein shell that protects and transports the viral genome from one host to the other. The main goal is to identify and understand the molecular mechanisms involved in the building of the these particles, which go through an spontaneous self-assembly process inside the cell. The data analysis needed in our research collaboration, which is not currently using the cross-border 10 Gigabit network connection, could greatly benefit from its use.

* * * * *

Focus Area: Seismography

14) Title: Current state of permanent GPS network in northern Baja California, Mexico: Challenges and opportunities for next decade **343**

Submitter: Javier Gonzalez-Garcia, Alejandro Gonzalez-Ortega, CICESE

Summary: Currently, the permanent GPS network in northern Baja California consist of 19 stations, which produce ~40Mb/day of raw data with sampling rate at 15 seconds. As more GPS stations are planned to be installed during next couple of years and the sampling rate would be change to 1-10 hz for real time applications, it will become imperative to have appropriate data flow. With this, the data flow will be increased by ~10^3 more data than today, with low latency and real time communications would made possible Alert for big earthquakes giving at least 10 sec. to make important decisions before the arrival of the first seismic waves. With no doubt the topographic/survey community and engineering in general and navigation in particular all this and more will be benefit with this living revolution.

Submitter: Octavio Lazaro-Mancilla, UABC

Summary: As researcher of Laboratorio de Sismología y Geofísica Aplicada de la UABC I have collaborated with researchers of Seismolab of CALTECH. Seismological collaboration include the deployment of temporary seismometers in field after April 4, 2010 El Mayor-Cucapah earthquake. For storage of data we have used flash cards that after were sent by mail to the Seismolab for analysis. On the other side, I have collaborated with researchers of CICESE, CALTECH, Virginia Tech and USGS in an active–source seismic imaging experiment in the Mexicali and Imperial Valleys in 2011. Nowadays CICESE, CALTECH researchers and I are working in a urban fault mapping project in the City of Mexicali using a variety of geophysical data and aerial images. In this project it is possible to do studies of LIDAR in urban areas in Mexicali and repeat them if possible after a certain time for detection of differences that can indicate subsidence and / or fault movement. For education and research I think that better facilities are needed for real time transmissions of large data sets such as LIDAR and satellite images including seismology data.

Submitter: Joann Stock, Robert Clayton, CALTECH

Summary: Caltech Seismological Laboratory personnel have both formal and informal collaborations with researchers at CICESE, UNAM (Instituto de Geofisica), UABC-Mexicali, UABC-Ensenada, and University of Sonora. Seismological collaborations are extensive and varied and would benefit from improved speed of transmission between UNAM and Caltech, which may prove to be particularly crucial in the event of a major earthquake. For deployment of temporary or permanent seismometers, the ability to directly connect instruments to the internet and transmit data in real time is a strong advantage and is key to newer instrument designs. However data transmission was compromised in the Mexicali Valley area by the 2010 earthquake, which caused interruption of internet and wireless telephone services as well as infrastructural damage. For the educational and research collaborations, better facilities are needed for video conferencing including the ability for simultaneous presentation of large 3D data sets such as LiDAR point clouds, DEMs, aerial and satellite imagery, and high resolution bathymetry. Ideally we would like to jointly examine and work with these in real time, involving investigators on both sides of the border, for both one-on-one research collaborations and thesis committee meetings. A high-speed network is essential for future earth-science collaborations between Mexican and American scientists.

Submitter: Victor Wong-Ortega, CICESE

Summary: This project pretend to set up new instrumentation to measure seismic activity in the earthquake prone zone of Mexicali Valley, Baja California, México, to the Northwest Seismic Network of México (RESNOM) of CICESE and to the National Center for Prevention of Disasters (CENAPRED). Real-time data provided from these devices installed in the northwest of Mexico will be used to protect lives and property in north Baja California region. USGS, CENAPPRED and CICESE will collaborate to execute this project.

Submitter: Enrique Pacheco Cabrera, AEM

Summary: Mexico presents natural and anthropogenic disasters of different categories and magnitudes, such as earthquakes, floods, forest fires, volcanic eruptions, emergencies in oil fields, etc. The Mexican Space Agency (AEM) proposes to develop a national system for storage and data processing, Geomatics, spatial and astrophysicists (SNAP-DGEA) seek to integrate, develop and consolidate infrastructure, storage, processing and distribution of such data, which represents BIG DATA and therefore an obligated need for BIG STORAGE and BIG NETWORKS for the distribution systems as well as capabilities of computation of very high performance for processing and visualization.

* * * * *

Various Focus Areas

19) Title: Integrating distributed LIDAR data repositories into the OpenTopography service infrastructure **430**

Submitter: Chaitan Baru, Viswanath Nandigam, UCSD

Summary: Operating from the San Diego Supercomputer Center, the OpenTopography facility provides online access to high-resolution LIDAR topography data and tools by hosts Lidar datasets contributed by data acquisition projects and serving as a metadata hub for remote Lidar datasets. OpenTopography's multi-tiered software architecture—consisting of an infrastructure tier, services tier and applications tier—supports scalability and extensibility.

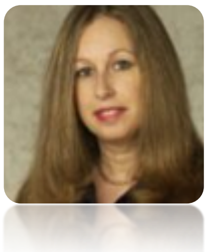
For example, CICESE in Ensenada, Mexico could host and operate data repositories with local data management procedures and data distribution timelines, while making these data available to a much larger user community via OpenTopography. Processing of these data could be performed at the remote location, or by services hosted by OpenTopography. This is technically accomplished by extending the OpenTopography service layer over the different data providers hosting the datasets, which requires packaging and distributing a customized instance of the Opal toolkit that would wrap the remote software application into OPAL-based web services and then integrating these services into the OpenTopography framework. The availability of a highspeed network links between CICESE in Ensenada and SDSC in San Diego would also enable caching of datasets on OpenTopography's storage servers, which can improve performance and make the system more robust.

20) Title: The Present and Future Development of Observatorio Astronomico Nacional in San Pedro Martir (OAN-SPM) **448**

Submitter: Michael Richer, William Lee, UNAM

Summary: The site of OAN-SPM, operated by Instituto de Astronomia-UNAM, is among the best in the world for optical-infrared astronomy. Several medium and large scale projects in various stages of construction and planning will generate large amounts of data, requiring stable and broad connections for download and upload. Further, the raw and processed data will need to be analyzed and transmitted to sites depending on the particular collaboration partners. The ReIonization and Transients InfraRed project (operating), the Trans-Neptunian Automated Occultation Survey-II (under construction), and the San Pedro Martir Telescope (in planning) with the US among partners, all represent an internationalization of the

Observatory, and a true bi-national astrophysics laboratory which will require, and greatly benefit from, a stable, high speed and broadband connection. We will describe the needs of these projects and their potential to advance cross-border collaborations in science, technology and the development of human resources.



Submitters: Robert Grossman UChicago, Heidi Alvarez FIU

Summary: The Open Science Data Cloud (OSDC) is a six petabyte science cloud for researchers to manage, analyze and share their data and to get easy access to data from other scientists. The OSDC is operated by the not-for-profit Open Cloud Consortium (OCC), which operates cloud computing infrastructure for the research community. The OCC includes academic institutions, companies, and government agencies. The OSDC has international partnerships with scientists in the United Kingdom, Brazil, Canada,The Netherlands, Japan and China, and is always interested in expanding its international partnerships. The OSDC plans to begin interoperating with science clouds in other countries beginning in 2014.

Submitters: Andrei Techernykh, Jose Lozano Rizk, CICESE

Summary: organizations with terabytes of digital content. The scheduling of jobs on multiprocessors is generally well understood and has been studied for decades. Many research results exist for different variations of the scheduling problem. However, the big data communication-aware scheduling problems have rarely been addressed.

In this talk, we will discuss a model for cloud computing applications taking into account a variety of communication resources used in real systems. This communication-aware model, called CA-DAG, allows making separate resource allocation decisions, assigning processors to handle computing jobs, and network resources for data transmissions. We will discuss the benefits, weaknesses, and performance characteristics of such a model and resource allocation strategies in presence of uncertainty.

The Science Gateway Institute

Opportunities for cross-border collaboration

Nancy Wilkins-Diehr
San Diego Supercomputer Center
wilkinsn@sdsc.edu

Science Gateways

Gateways or Web portals:
A natural result of the impact of the internet on worldwide communication and information retrieval

Only ~20 years since the release of Mosaic!

- Implications on the conduct of science are still evolving
  - 1980’s, Early gateways, National Center for Biotechnology Information BLAST server, search results sent by email, still a working portal today
  - 1989 World Wide Web developed at CERN
  - 1992 Mosaic web browser developed
  - 1995 “International Protein Data Bank Enhanced by Computer Browser”
  - 2004 TeraGrid project director Rick Stevens recognized growth in scientific portal development and proposed the Science Gateway Program
  - Today, Web 3.0 and programmatic exchange of data between web pages
- Simultaneous explosion of digital information
  - Growing analysis needs in many, many scientific areas
  - Sensors, telescopes, satellites, digital images, video, genome sequencers
  - #1 machine on Top500 today over 10,000x more powerful than *all combined* entries on the first list in 1993

Science Gateways

Did you know?

- NSF supercomputers can be accessed through user-designed web interfaces as well as from the command line
- In 2012 40% of XSEDE users came through gateways

2.1. Categories of Research Activities

Within this broad scope, TeraGrid projects support various categories of research consortiums and services. Most often, TeraGrid projects can be classified in one of the following categories:

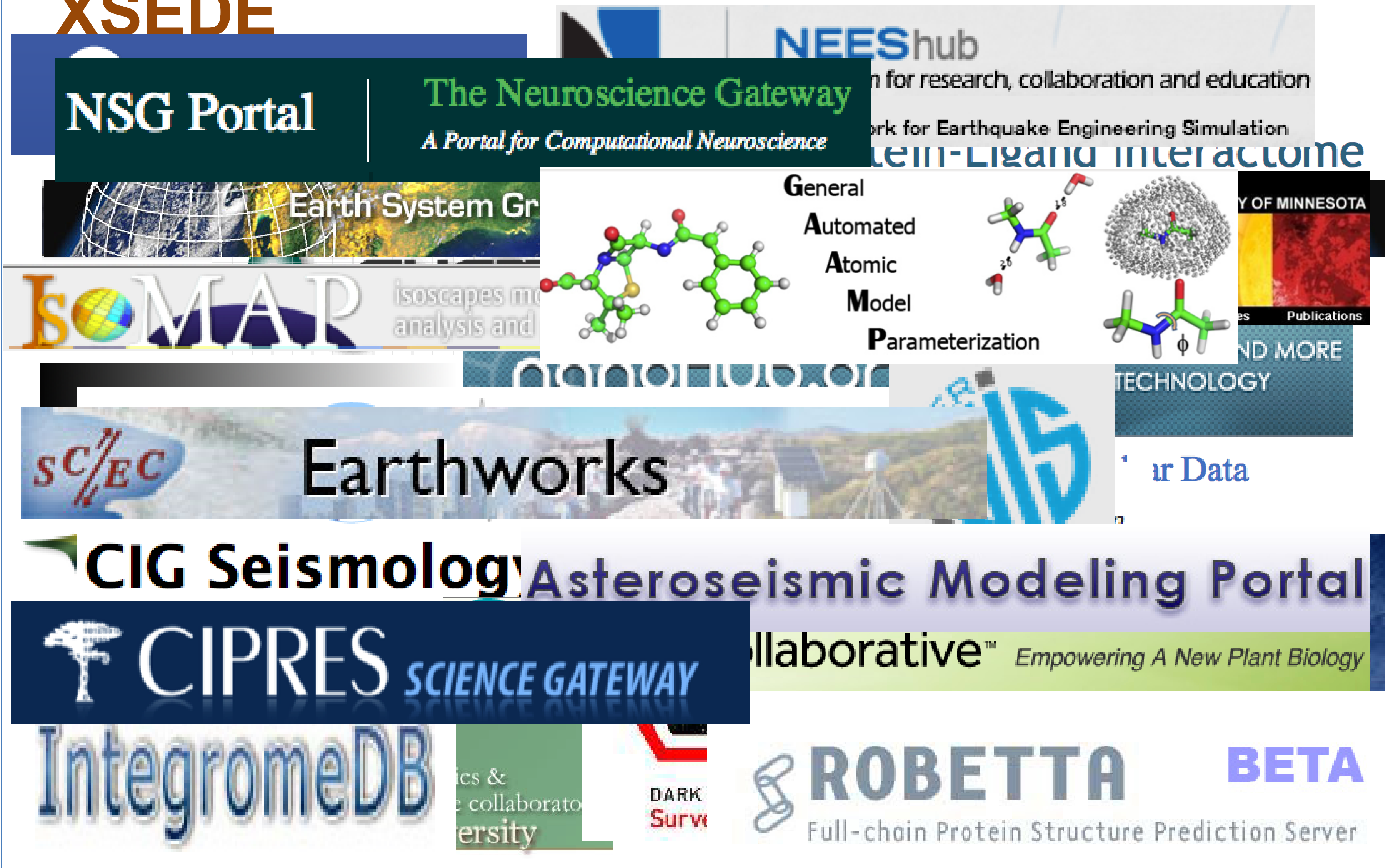
1. Single Principal Investigator (PI): Most projects support a single PI and possibly a small research group working closely and co-authoring papers.
2. Multi-PI, Large Research Collaborations: Projects of this type are characterized by a single PI representing a large group of collaborating co-PIs who are working on subprojects within the overall collaboration. A single request is submitted, and a single project is allocated. The management of the allocated resources is left to the discretion of the principals on the request.
3. Large-scale Consortiums: Projects of this type are intended to support large-scale, funded projects that work together as a consortium. Often in these cases, a mechanism already exists for allocating community or project resources (e.g., an instrument such as a telescope or detector), and that mechanism will also be used to make allocations from the time granted to the community project to the individual investigators. Requests for this type of project typically describe the internal processes for managing access of individual investigators within the consortium.
4. Gateways or Community Services: Projects of this type provide services to a large community of users who are typically not directly collaborating with the project PI. An example of such a project would be an application portal service providing access to software and computer time to a community of biology researchers via a web-based interface. Requests to provide such a service must describe the details of the services provided, the methods used, the expected consumption of resources, and mechanisms for monitoring the users and usage of the service. Statistics of community usage should be reported quarterly and in renewal requests for resources, progress reports, and end-of-project reports. TeraGrid provides How to Write a Winning Gateway Proposal to assist in writing these requests.

2.2. PI Eligibility

The principal investigator (PI) of an allocated project is the person responsible for the accuracy of the resource request and the management of the ensuing allocation. This person must be a researcher or educator at a U.S. academic or non-profit research institution. A PI may not be a high school or undergraduate student; a qualified advisor, for example, a high school teacher or faculty member, must serve in this capacity. In most cases, a graduate student may not be a PI; however, see Section 8.1 for an exception for NSF Graduate Research Fellows and Honorable Mention awardees. A post-doctoral researcher is eligible to be a PI.

A U.S.-based scientist, engineer, or educator who has a joint appointment with a university or non-profit research institution

Today, there are over 25 gateways using XSEDE



Cyberinfrastructure for Phylogenetic Research (CIPRES)

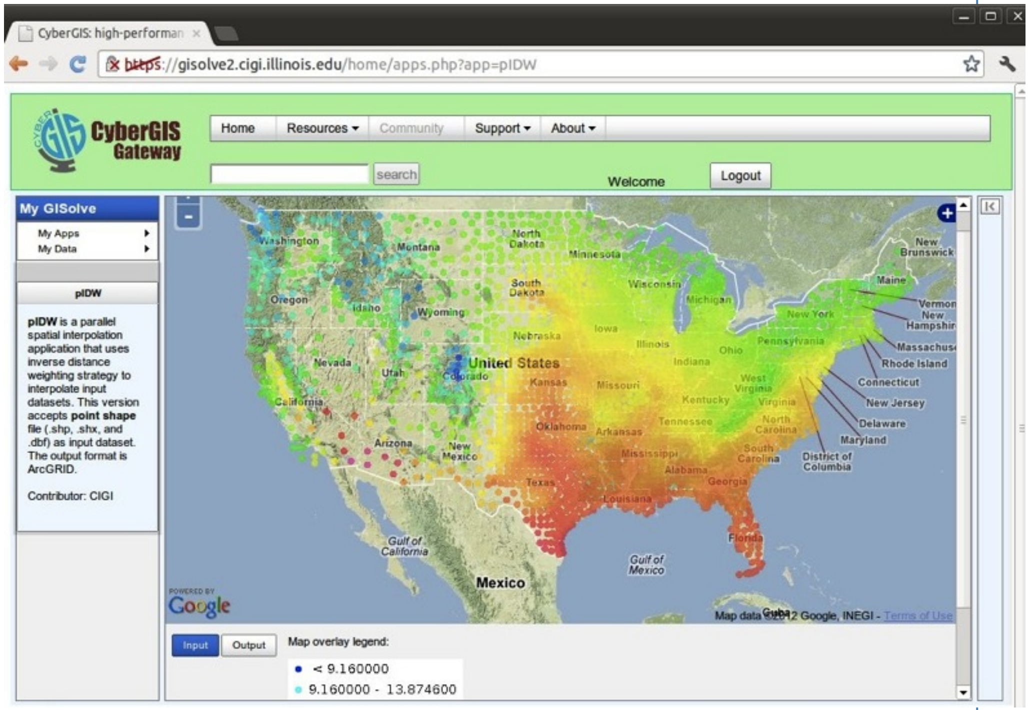
- Most popular science gateway in TeraGrid
  - ~25% of all XSEDE users
- In use on 6 continents
- Cited in major journals (Cell, Nature, PNAS)
- Used at major research institutions (Stanford, Harvard, Yale)
- Used by 57 researchers for curriculum delivery
- Used in 80% of EPSCoR states
- Recently used by a 15-year-old high school student who won the Massachusetts state science fair with no support from gateway staff

Science Gateways

CyberGIS

Software Integration for Sustained Geospatial Innovation

- Application of high-end cyberinfrastructure to GIS
- Influence on multiple domains
- Improved decision support
- Spatial joins, layers of multiple datasets at different resolutions
- Goal is core set of composable, interoperable, manageable, and reusable software elements
- Collaborative geospatial problem solving environment



Science Gateways

Analytical Ultracentrifugation: Emerging computational tool for the study of proteins

UltraScan 9.0

The Center for Analytical Ultracentrifugation of Macromolecular Assemblies, UT Health Sciences, Borries Demeler, PI

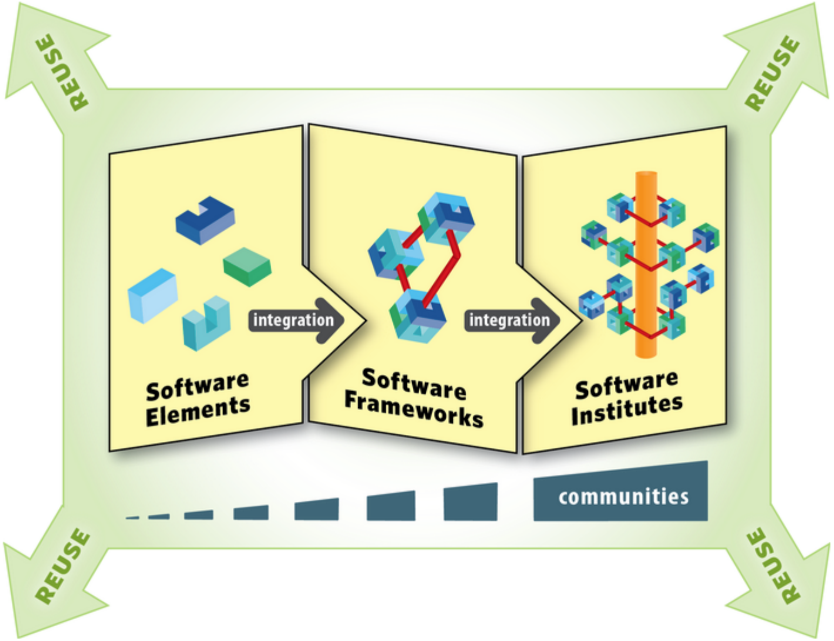
- Samples from researchers all over the world
  - Some (Germany, Australia) have their own ultracentrifuges and use only the analysis capabilities, others send samples to UT to spin
- Spin the samples at high speeds, learn about macromolecule properties
- Monte Carlo simulations
- Observations are electronically digitized and stored for further mathematical analysis

Science Gateways

NSF software vision implemented in 2010

Software Infrastructure for Sustained Innovation (SI2) program

- Scientific Software Elements (SSE)
  - Small groups create software that advances one or more area
- Scientific Software Integration (SSI)
  - Larger interdisciplinary teams, software frameworks
- Scientific Software Innovation Institutes (S2I2)



http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817

Science Gateways

Institutes: Long term hubs of excellence

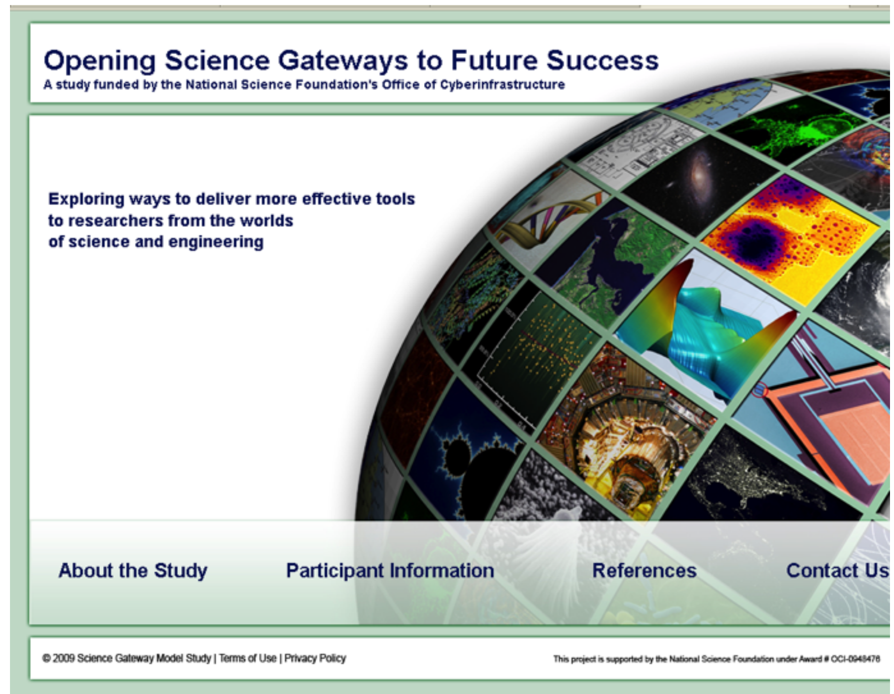
- Serve a research community of substantial size and disciplinary breadth
- Expertise, processes, architectures, resources and implementation mechanisms to transform research practices and productivity
- Support, outreach, workforce development, proactive approach to diversity
- Pathways to community involvement

http://www.nsf.gov/pubs/2011/nsf11589/nsf11589.htm

Science Gateways

Simultaneous NSF study identifies limitations to short-lived science portals or gateways

- Characteristics of short funding cycles
  - Build exciting prototypes with input from scientists
  - Work with early adopters to extend capabilities
  - Tools are publicized, more scientists interested
  - Funding ends
  - Scientists who invested their time to use new tools are disillusioned
    - Less likely to try something new again
  - Start again on new short-term project
- Need to break this cycle and fund for long-term success



Science Gateway Institute conceptualization award in 2012

Science Gateways

After reading stacks of reports, we wanted to capture our findings in a memorable way

- Youtube!
- http://www.youtube.com/watch?v=4ziEt0LRxEA

SCIENCE GATEWAYS .ORG

Are you building websites that serve your science discipline?

Do you wish you could *connect with and learn from others* who are doing the same thing?

We are building an institute to serve you—and others like you—with resources, services, experts, and ideas for creating and sustaining science gateways. Sign up to join the conversation : *http://sciencegateways.org/volunteer/*

science gateway /sī′ əns gāt′ wā′/ *n.*

1. an online community space for science and engineering research and education.
2. a Web-based resource for accessing data, software, computing services, and equipment specific to the needs of a science or engineering discipline.

Science Gateways

Millions of dollars are spent on gateways, but developers face several challenges:

- **They often work in isolation** even though development can be quite similar across domain areas.
- **They bridge cyberinfrastructure**—locally, campus-wide, nationally, and sometimes internationally.
- **They need foundational building blocks** so they can focus on higher-level, grand-challenge functionality.
- **They struggle to secure sustainable funding** because gateways span the worlds of research and infrastructure.

Through Summer 2014, we will be engaging interested members of the science and engineering community in a planning process for a **Science Gateway Institute** (SGW-I). The goal of the institute would be to provide coordinating activities across the National Science Foundation, offering several services and resources to support the gateway development community:

- An **incubator service** offering consultation and documentation about business planning and software development.
- An **extended support team** to build gateways and share their expertise.
- A **forum** to connect members of the development community.
- A **modular, layered framework** that supports community contributions and allows developers to choose components.
- **Workforce development** to help train the next generation for careers in this cross-disciplinary area.

Sharing expertise about technologies and strategies would allow developers to concentrate on the novel, challenging, and cutting-edge development needed by their specific user communities. First, we want to hear from you. If you would like to participate in this planning process or stay informed about our progress, please contact us.

http://sciencegateways.org/volunteer/

Science Gateways

Incubator Service

Assist with the entire lifecycle of a gateway:

- Business plan development and review
- Development environment, consulting, documentation and software recommendations
- Software repositories
- Software engineering facilities
- Software assessment services
  - like Open Source Software Advisory Service, Apache assessment service, Software Sustainability Institute (UK)
- Build-and-test facilities
- Hosting service
- Offering gateways expertise in the following areas:
  - Usability assessment
  - Licensing
  - Sustainability
  - Project management
  - Security

Science Gateways

Gateway-building Support

- Institute staff assigned to a project for months, up to a year
  - Assist with gateway development or implementation of advanced features
    - Workflows, fault tolerance, sensor feeds, HPC simulations
  - Teach research teams what it takes to build, enhance, operate, and maintain gateways after support ends
  - Peer-reviewed request process open to all

Gateway Forum

- Gathering place for scientific web developers across NSF directorates, agencies, and international boundaries
- Social forums, white papers, blogs, testimonials and user stories
- Annual conference
- Broad and engaging symposium series
- Gateway training program
  - Synchronous and asynchronous, video tutorials
  - Best practices, case studies
- Showcase of successful projects
- Environment that enables continuous community feedback

Science Gateways

Gateway Framework

- Modular, layered approach
  - Supports community contributions
  - Grocery store approach allows developers to pick and choose the components they need
- Tiered architecture
  1. Value-added services
     - Publication channel for delivering content to a wider audience
     - Information repositories for good design practices
     - Information/code samples for best practices in user-interface and user-experience design
  2. Core web framework which includes hosted site creation and content management
  3. Platform API to provide a cohesive set of RESTful web services upon which the previous two layers rely
  4. Systems layer where the hardware and low-level middleware reside
     - Clouds and cloud services, HPC systems, grid middleware, data warehouses, databases, instrumentation, and distributed data stores

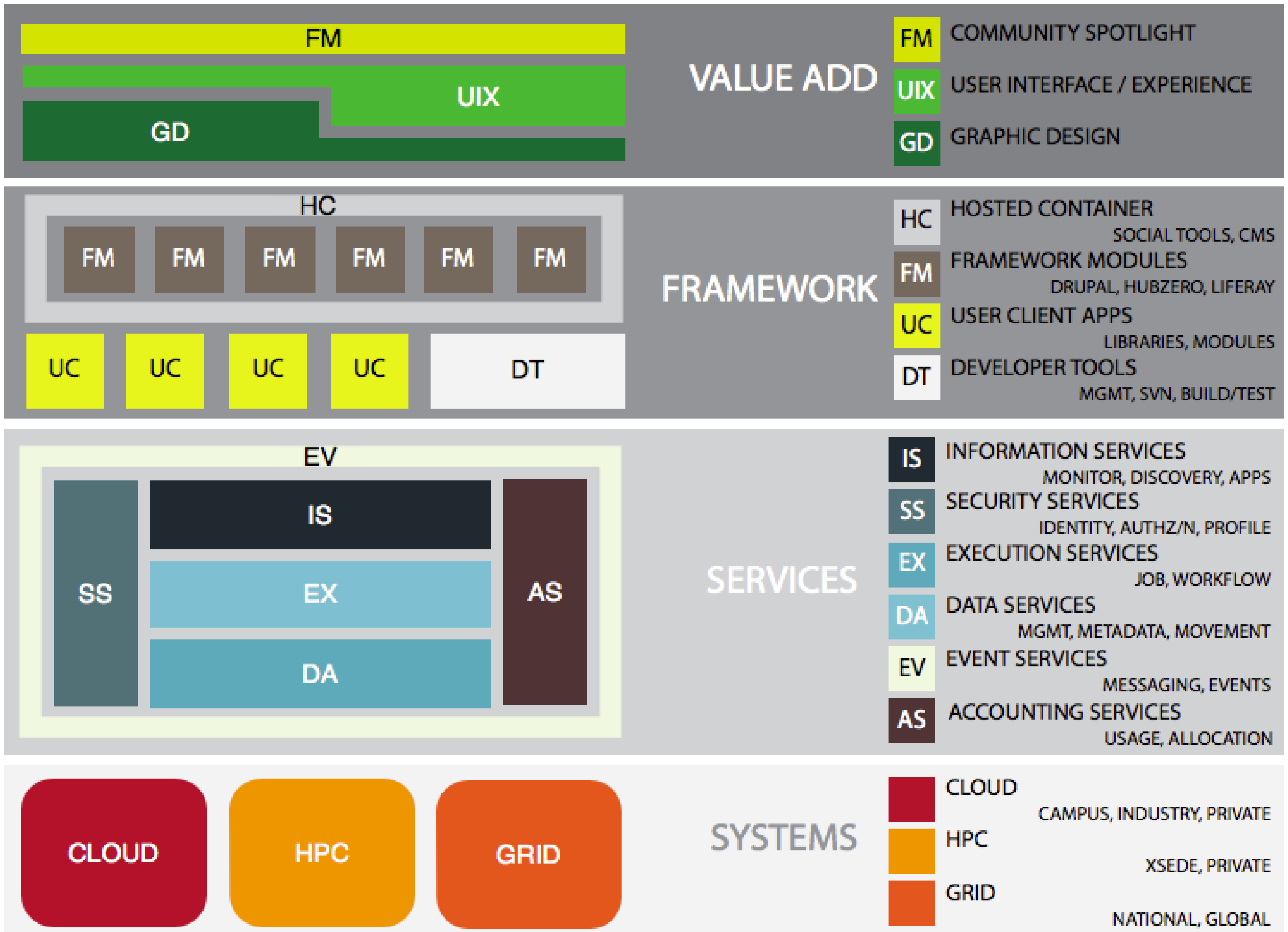


Figure 1. High-level architecture of software offerings and value-added services provided by the institute.

Workforce Development

- Terrific opportunities for students and IT professionals
  - Much science gateway development currently done by campus IT
- Gateway building training
  - Web development is a natural interest area for students
    - Very visual, see results of programming instantly
  - Builds cross-disciplinary communication skills
    - Talk to scientists, construct a gateway that meets their needs
  - Utilize existing programming opportunities such as Google Summer of Code
- Opportunities to proactively address diversity

Science Gateways

Conclusion

- Gateways can provide terrific mechanism to make use of high-speed networks and serve a large user community
- Very interested in a collaborative organization that involves many gateway perspectives
- More info at http://www.sciencegateways.org
- Thank you!

The HUBzero Platform for Scientific Collaboration

Michael McLennan
Director, HUBzero® Platform for Scientific Collaboration
Purdue University

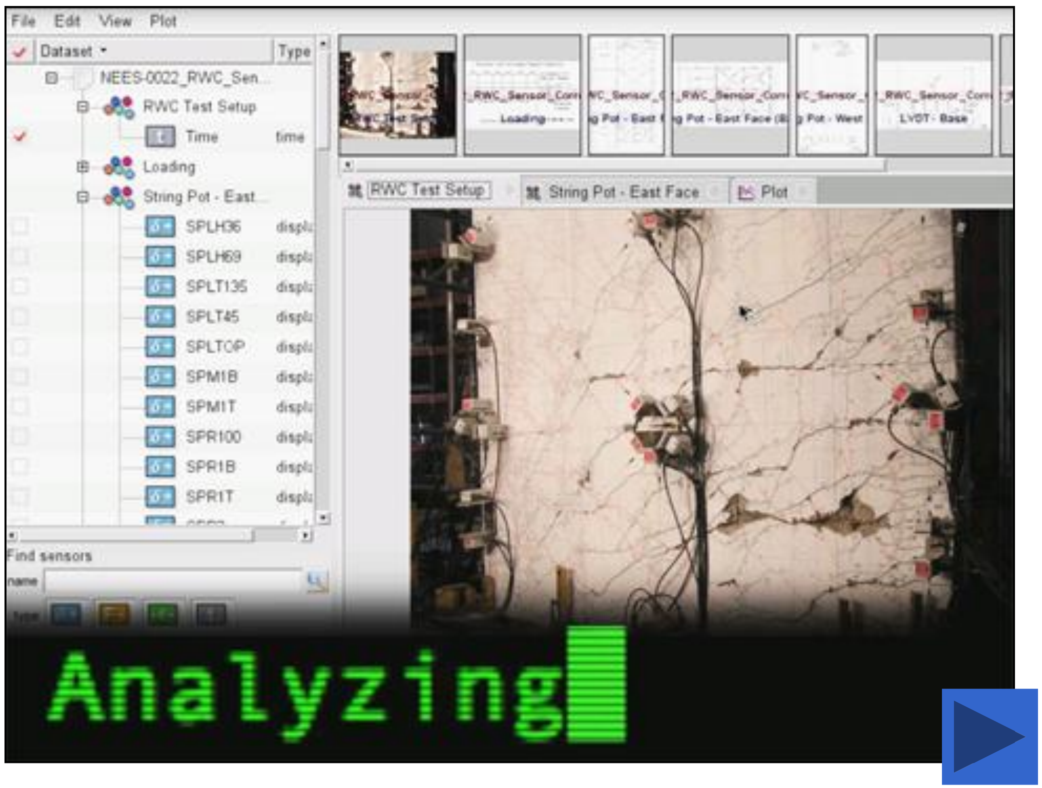
What is HUBzero?

Open source software platform
used for building

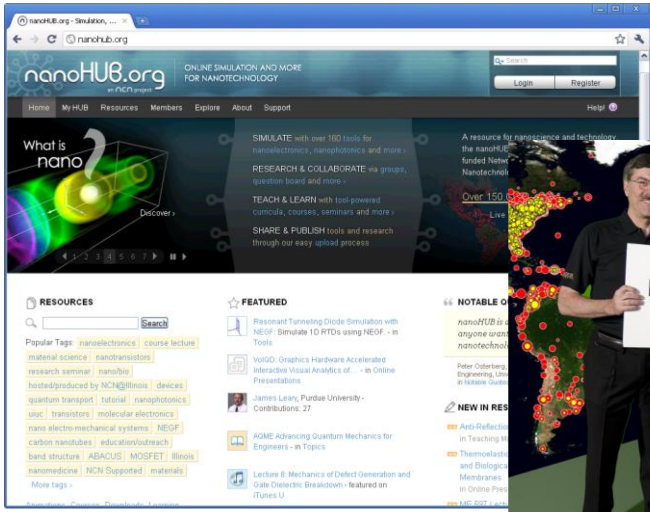
“Science Gateways”
“Collaboratories”
“Hubs”

Our Mission

Build Collaboratories for Research, Education, and Collaboration



Started with nanoHUB.org



Network for Computational Nanotechnology

Established in 2002

Awards
EEC-0228390
EEC-0634750
OCI-0438246
OCI-0721680

540,063 Visitors

258,791 Users

276 Simulation Tools

4,312 Resources

1,032 Citations

14,000 Students at 185 Insitutions

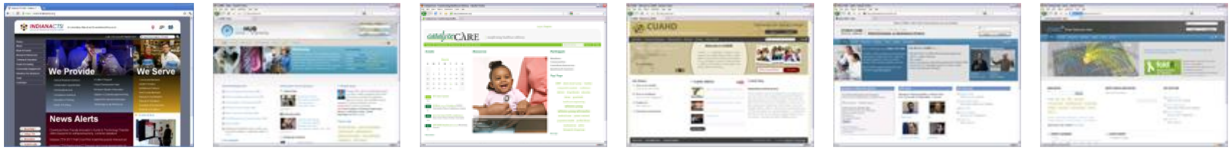
50+ Hubs for many disciplines

hubzero

Works for different scientific communities

Engineering

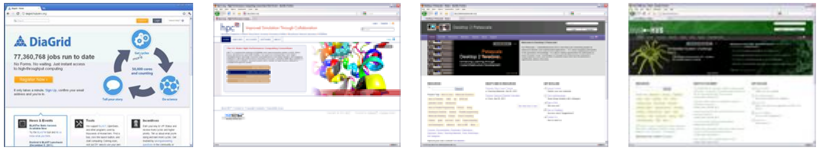
Healthcare



Environment

Education

Computing



50+ Hubs for many disciplines

~1,000,000 visitors total

visitors	users	
554,223	266,292	nanoHUB.org
249,187	69,232	nees.org
58,690	28,972	GlobalHUB.org
36,660	23,143	pharmaHUB.org
35,807	12,380	vhub.org
44,478	3,539	HABRIcentral.org
17,037	3,923	cceHUB.org
16,211	5,523	ciHUB.org
13,592	4,278	StemEdHub.org
12,198	2,185	iemhub.org
14,184	2,129	C3Bio.org
7,501	1,272	cleerHUB.org

Global community

38

Purdue University 27

Indiana University · The University of Wisconsin Madison · Clemson University · Washington State University · University of Connecticut 1881 · SUNY The State University of New York

hubzero Foundation, LLC

- Non-profit organization
- Independent owner of HUBzero code
- Promotes dissemination and outreach
- Sponsors HUBbub Conference
- Coordinates software contributions

Get Involved!

39

1

Download and have at it!

open source

2

Join the Foundation

hubzero PRINCIPAL

PARTICIPANT IN THE HUBZERO FOUNDATION

3

Consulting and Hosted Solutions

More information

Michael McLennan
Director, HUBzero® Project
mmclennan@purdue.edu

http://hubzero.org/pressroom

- Simulation/modeling tools
- Data management
- Collaboration and social networking
- Analytics to measure impact

A Science Gateway with Global Usage and Impact
nanoHUB.org

Gerhard Klimeck, Purdue University, gekco@purdue.edu

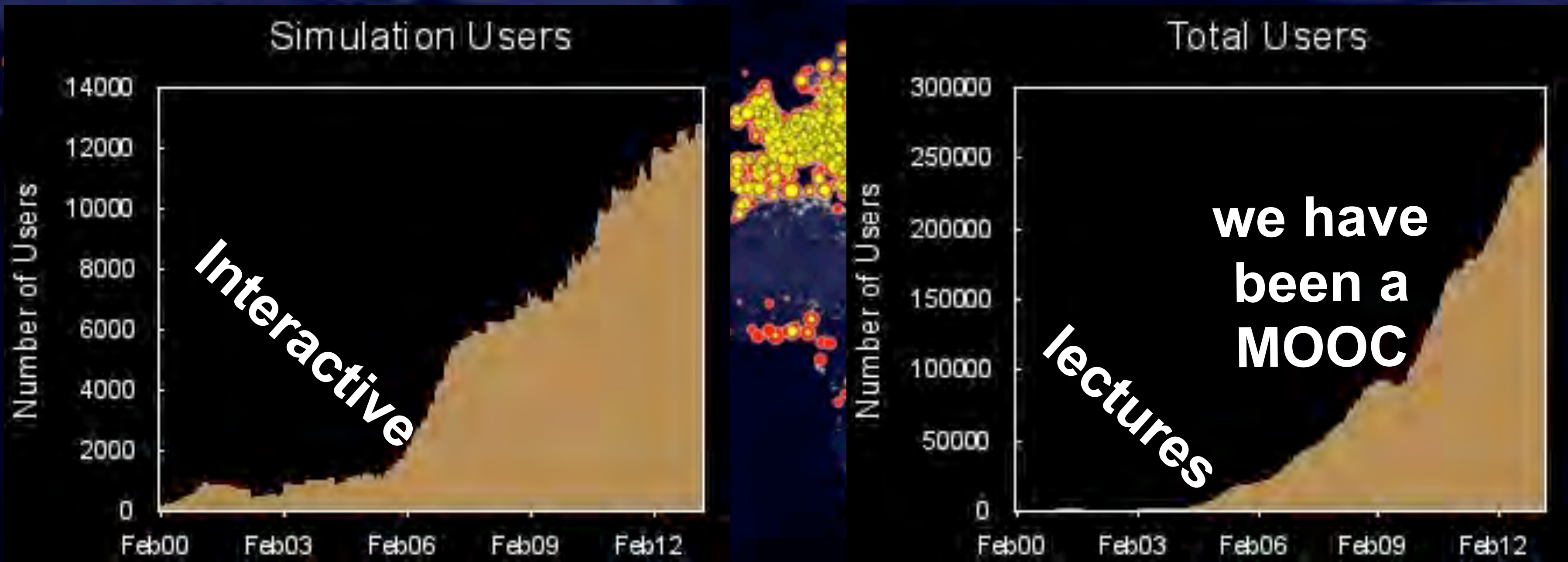
PURDUE UNIVERSITY

Mythbusting Scientific Knowledge Transfer with nanoHUB.org

Gerhard Klimeck, Purdue University, gekco@purdue.edu

Documenting with Real Data
Research<=>Education, Collaborative, Global Impact

Gerhard Klimeck, Purdue University, gekco@purdue.edu

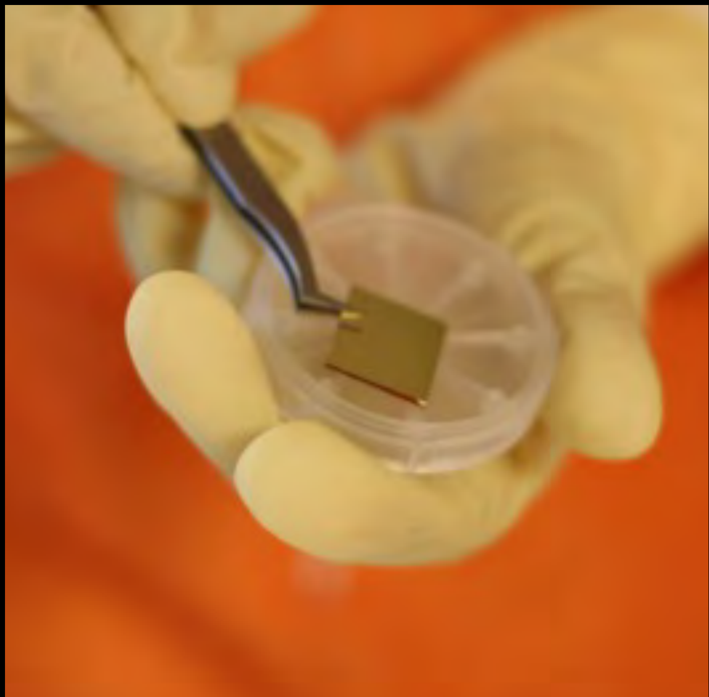


Over 13,000 / 280,000 Users Annually

nanoHUB usage

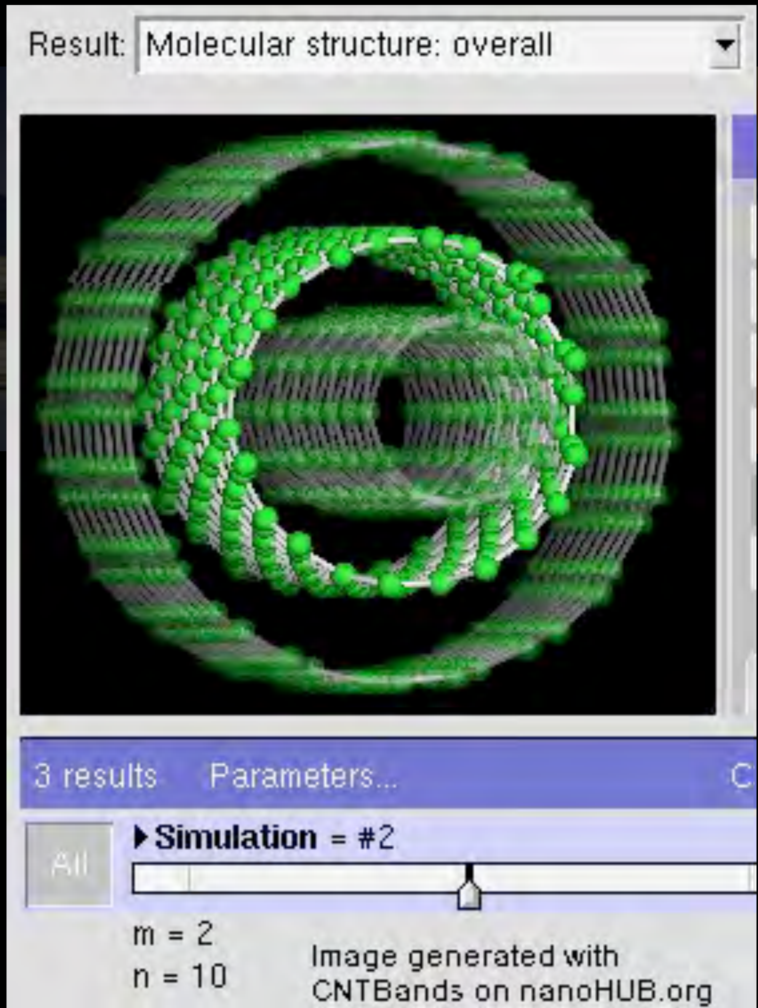
PURDUE UNIVERSITY

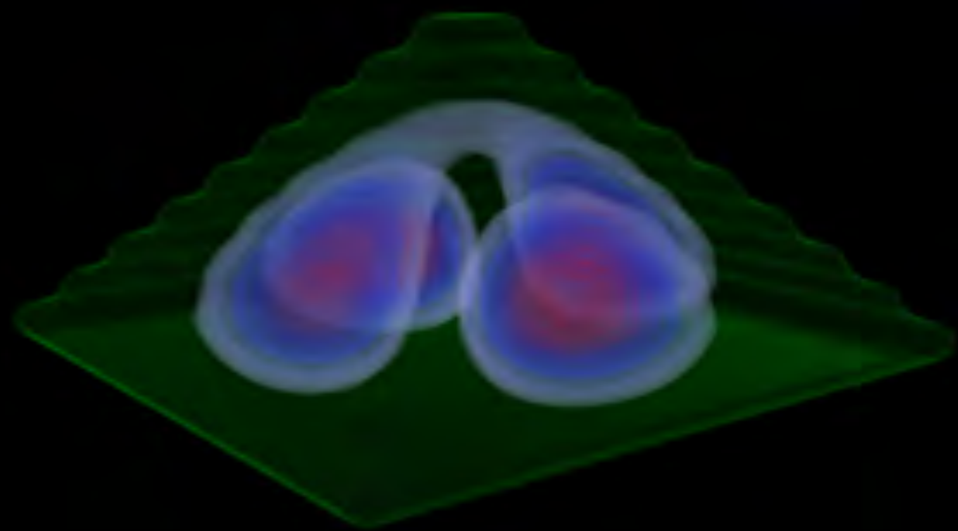
Nanotechnology



Extensive Facilities

Nano Models



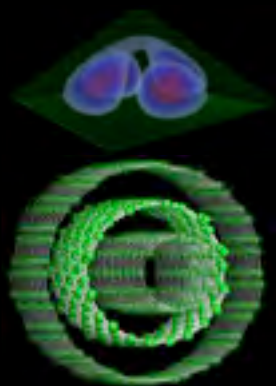


Quantum Dots

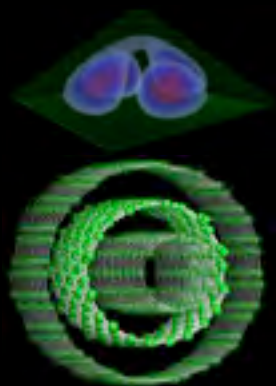
Artificial Atoms

Carbon Nanotubes

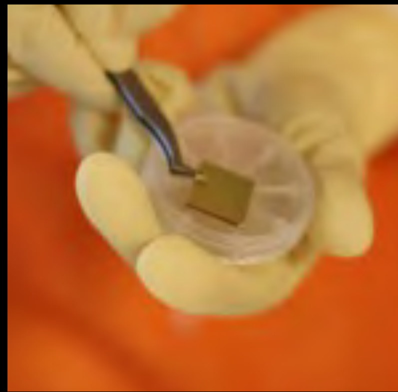
Computational Nano

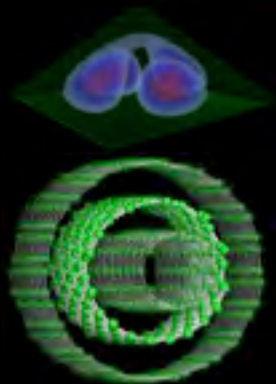


Computational Nano

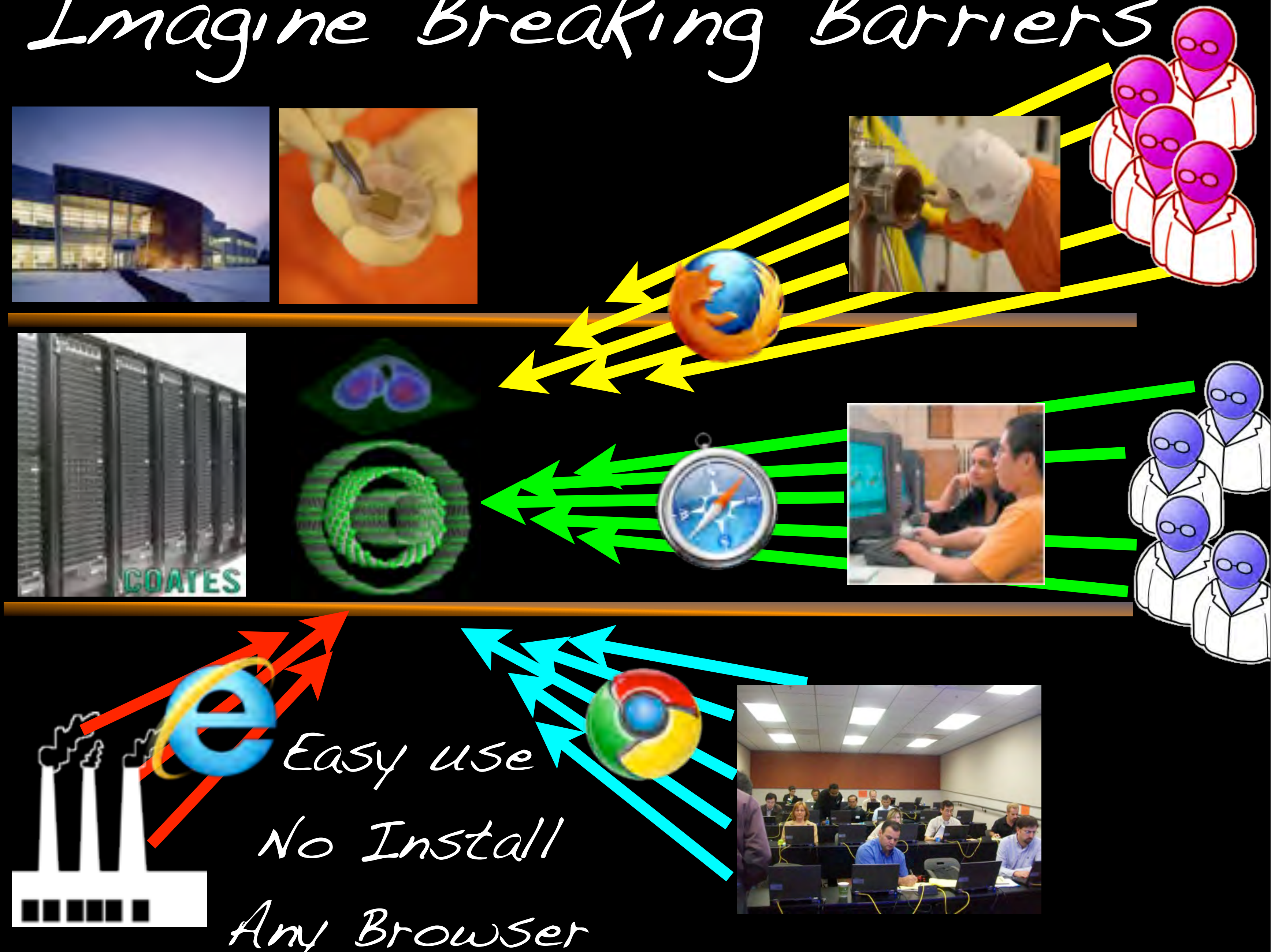


Different Worlds




Imagine Breaking Barriers




51
COATES
Easy use
No Install
Any Browser

Why is this so hard?

?

Most research codes

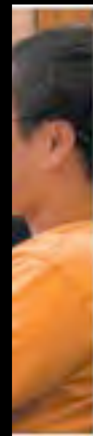
are written by one user

for one user

```
Structure
{
    Material
    {
        name                  = GaAs
        tag                   = substrate
        crystal_structure     = simplecubic
        atoms                 = (GaAs)
        Lattice:a_lattice     = 0.565
        regions               = (1)
        Bands:TB:s:param_set                              = nanohub
        Bands:TB:s:nanohub:E_S_GaAs                       = 12.1307935176
        Bands:TB:s:nanohub:V_S_S_Sig              as_H    = -20
        Bands:TB:s:nanohub:p                      ential  = 125
    }
    Domain
    {
        name                       = structure1
        type                       = pseudomorphic
        base_material              = substrate
        dimension                  = (18.0,19.0,9.0)
        periodic                   = (false, false, false)
        crystal_direction1         = (1,0,0)
        crystal_direction2         = (0,1,0)
        crystal_direction3         = (0,0,1)
        space_orientation_dir1     =  (1,0,0)
        space_orientation_dir2     =  (0,1,0)
        regions                    = (1)
        geometry_description       = simple_shapes
    }
```

?

User Hostile



❶ Input

About this tool
Questions?

Number of States: 7

Surface passivation: ☑ yes

Device Structure | Light Source

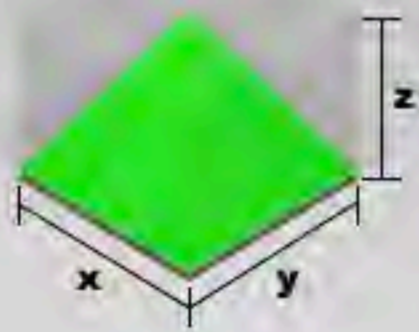
Geometry: Pyramid

?

X dimensions: 10nm

Y dimensions: 10.5nm

Z dimensions: 5nm



Effective Mass: 0.067

Discretization: 0.565nm

Energy gap: 1.43eV

Why is this so hard?

Accessible (no installation)

Most research codes are written by one user

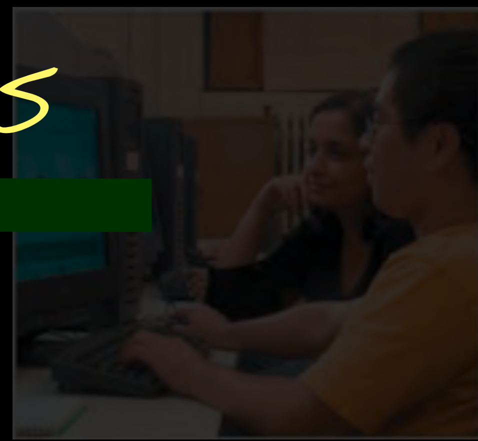
HUBzero

Developer Friendly for one user

Rappture

User Friendly

It has been very hard!



Emerged Myths

Accessible (no installation)

HUBzero

Developer Friendly

Rappture

User Friendly

Emerged Myths

User Friendly

Cannot use research codes for education

Must write own code to do research

Experimentalists cannot use research codes

Accessible (no installation)

NO End-to-end Science Cloud Possible

Developer Friendly

Building User Interfaces too Difficult

Must rewrite code for web deployment

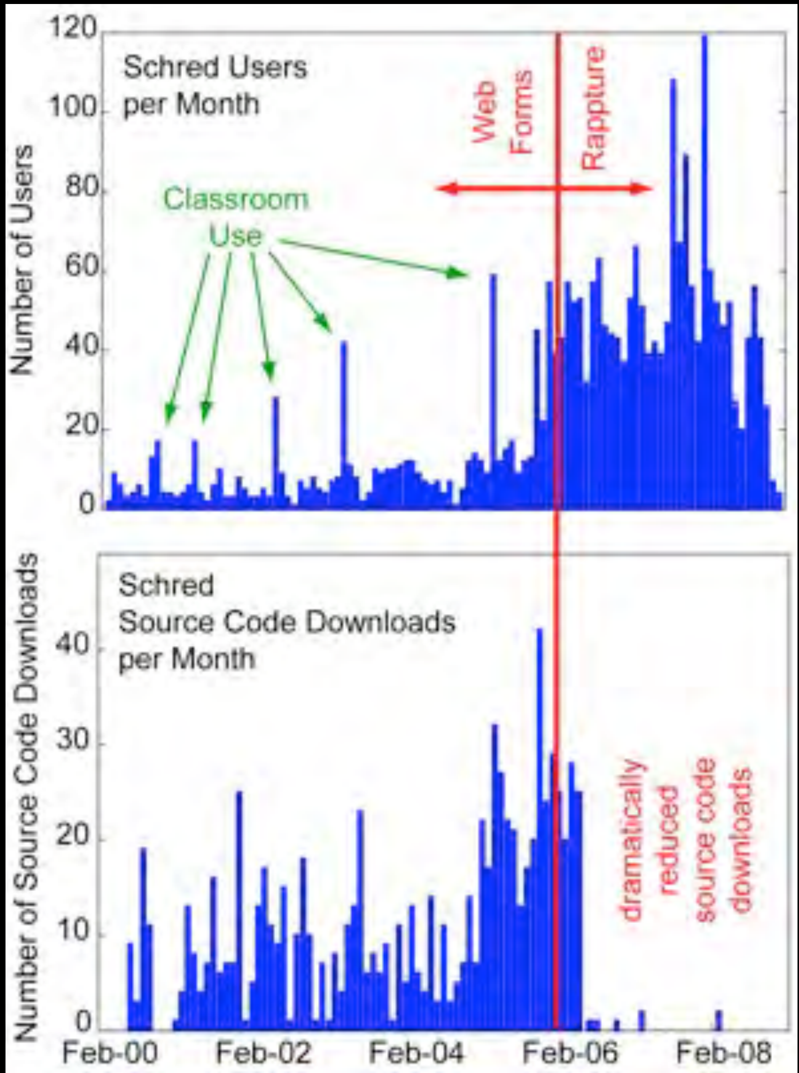
There is no incentive to share codes

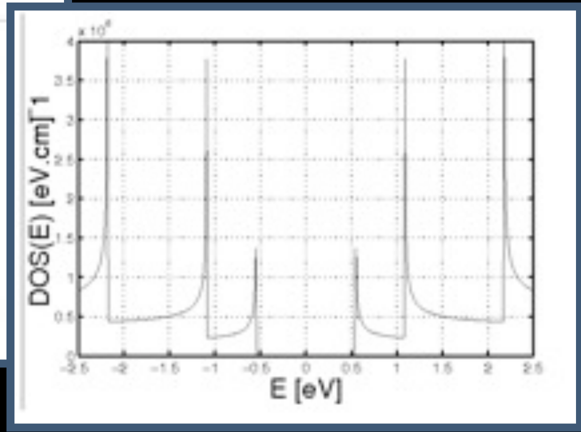
Customers

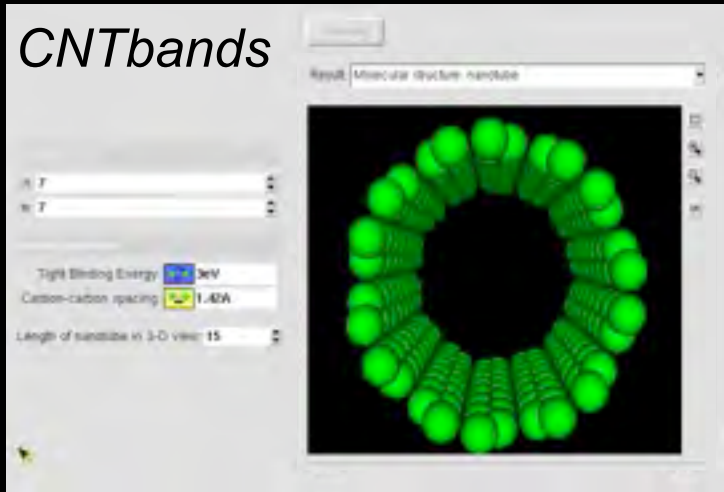
Market

Suppliers

Importance of a good GUI

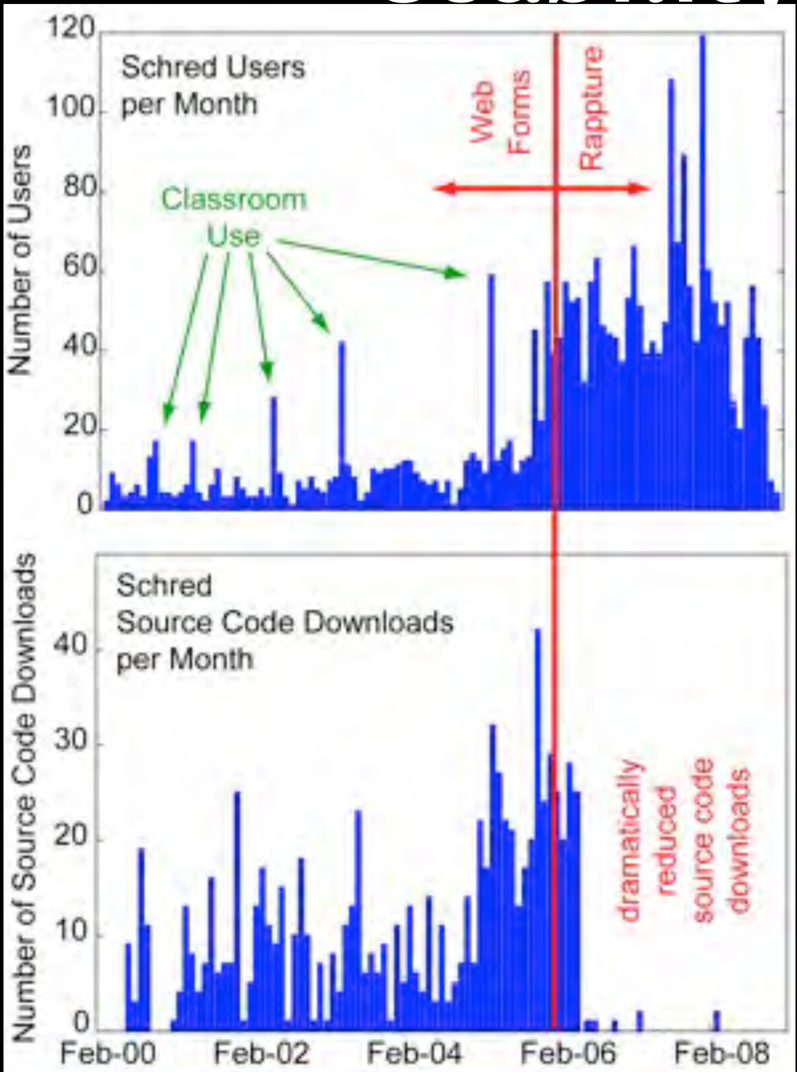






Same behavior across all similar converted tools

Balancing Usability and Capability



nanoHUB

iPhone / iPad

Emerged Myths

User Friendly

Cannot use research codes for education

Must write own code to do research

Experimentalists cannot use research codes

Accessible (no installation)

NO End-to-end Science Cloud Possible

Developer Friendly

Building User Interfaces too Difficult

Must rewrite code for web deployment

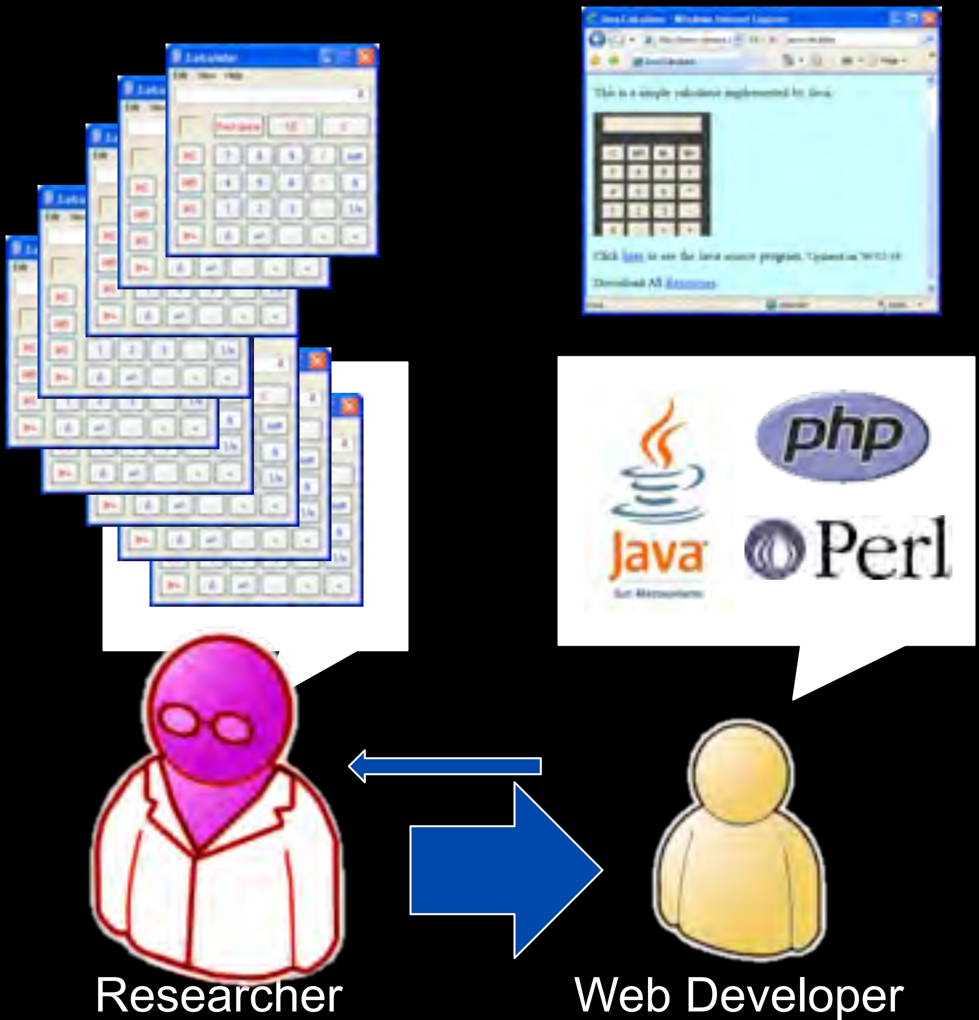
There is no incentive to share codes

Customers

Market

Suppliers

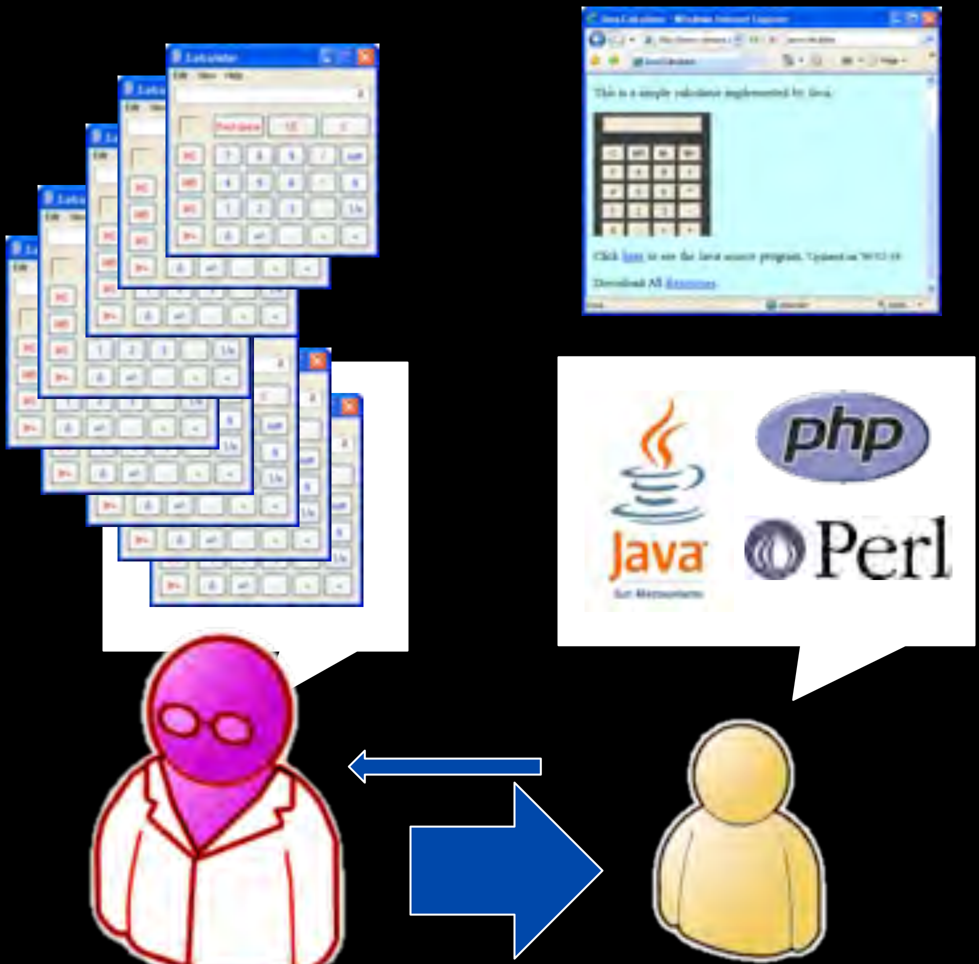
Usual Science Gateway Process



- **175 tools / 4 years: => $88M**
- $500k/tool
- NO new research!
- Not validated by researcher (disowned)
- Researcher has much better version
- Code rewrite takes 2-3 years

Many Proposals read alike

Usual Science Gateway Process



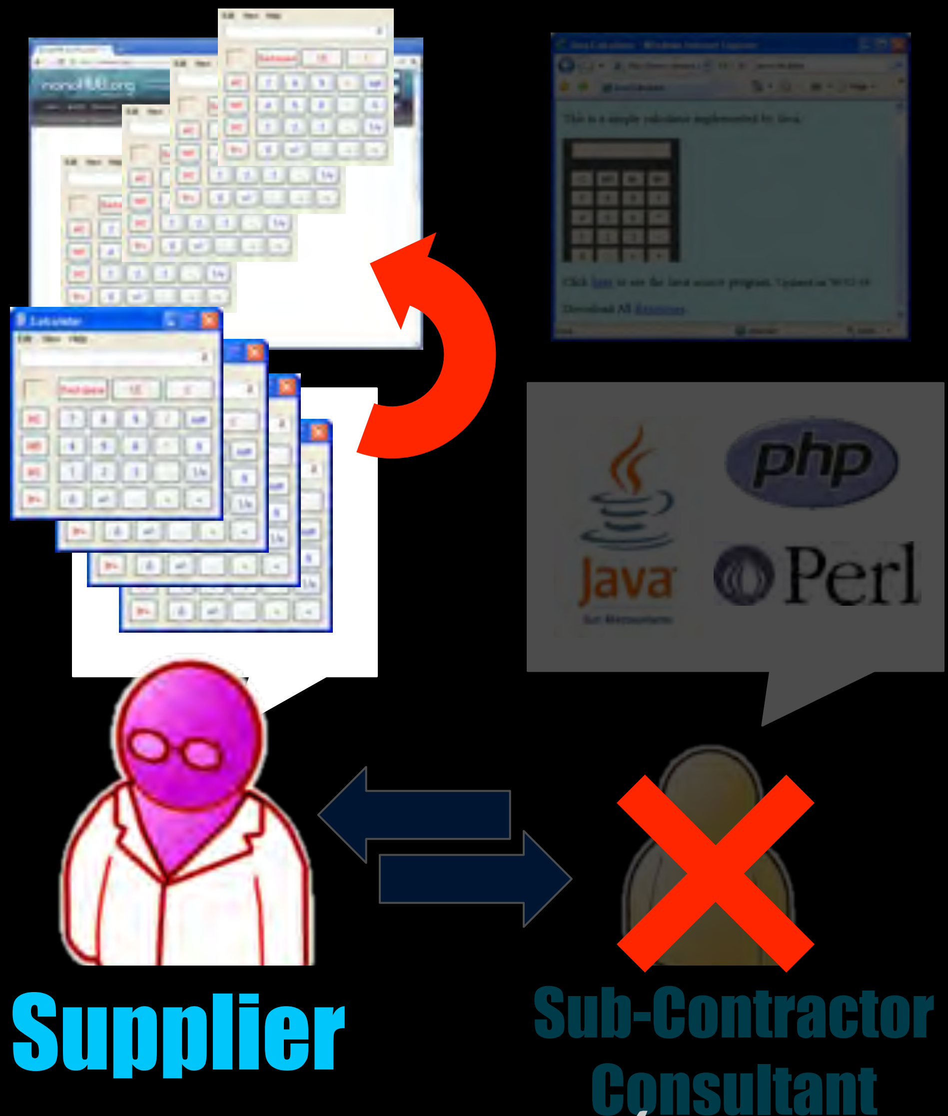
- ~~175 tools / 4 years: => $88M~~
- $500k/tool

Customers / Users

- Scale back expectations
- Not research codes
- Toy applications
- Not deep research
- Maybe for education?

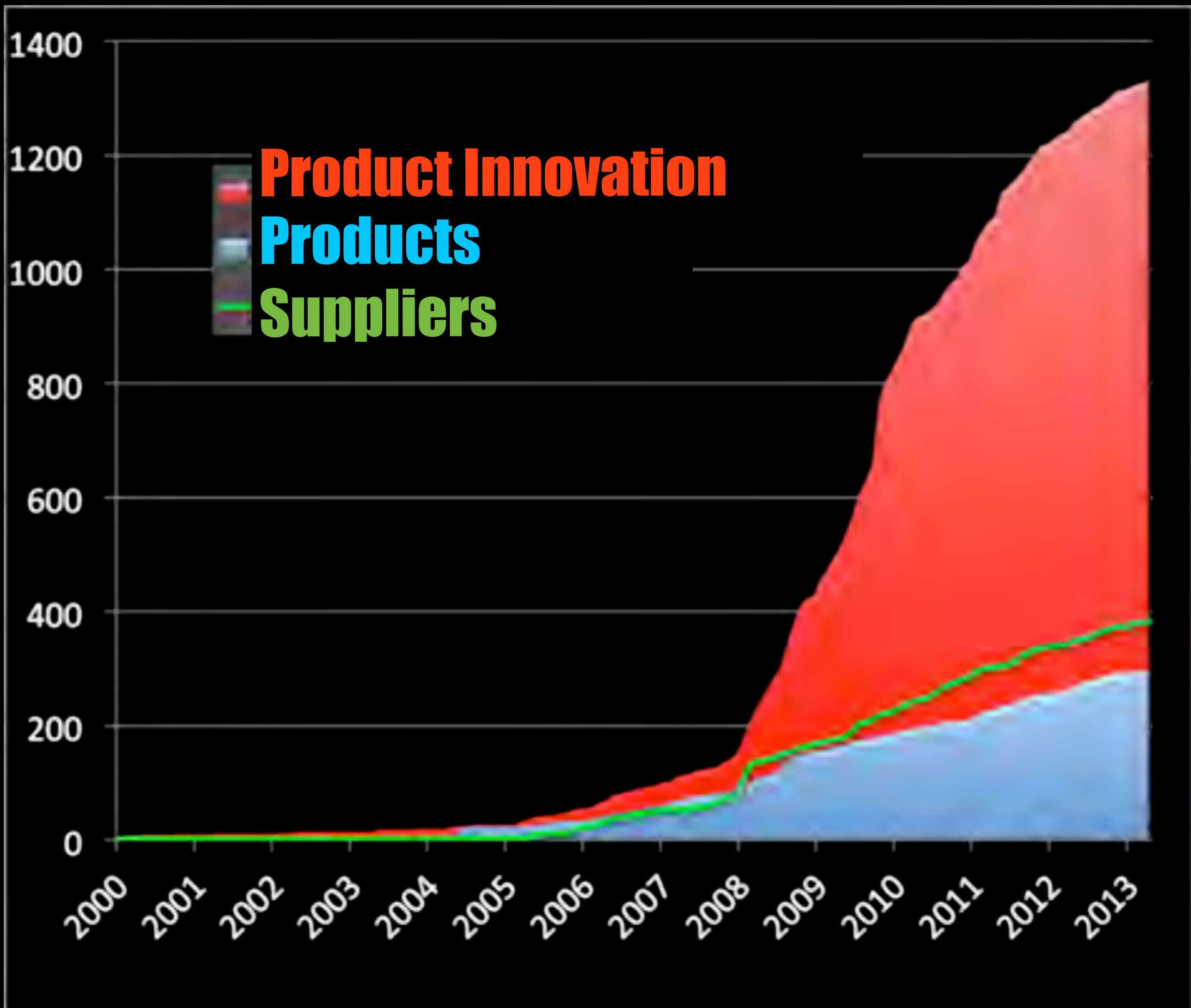
Generating a Bad Reputation

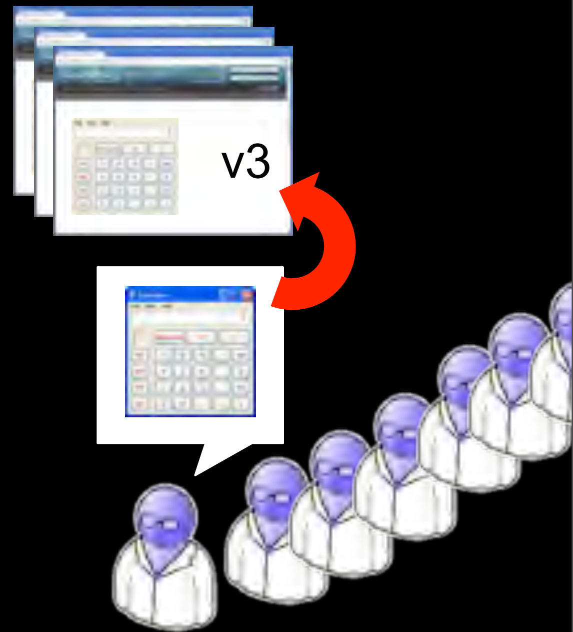
nanoHUB Process



- **175 tools / 4 years without $88M**
- Eliminate bottlenecks
  - No Middleman
  - No Rewrite
  - Retain ownership
- Rapid Deployment: 2-3 years → 1-2 weeks
- **Rappture** toolkit
- **HUBzero** Ecosystem

nanoHUB is different





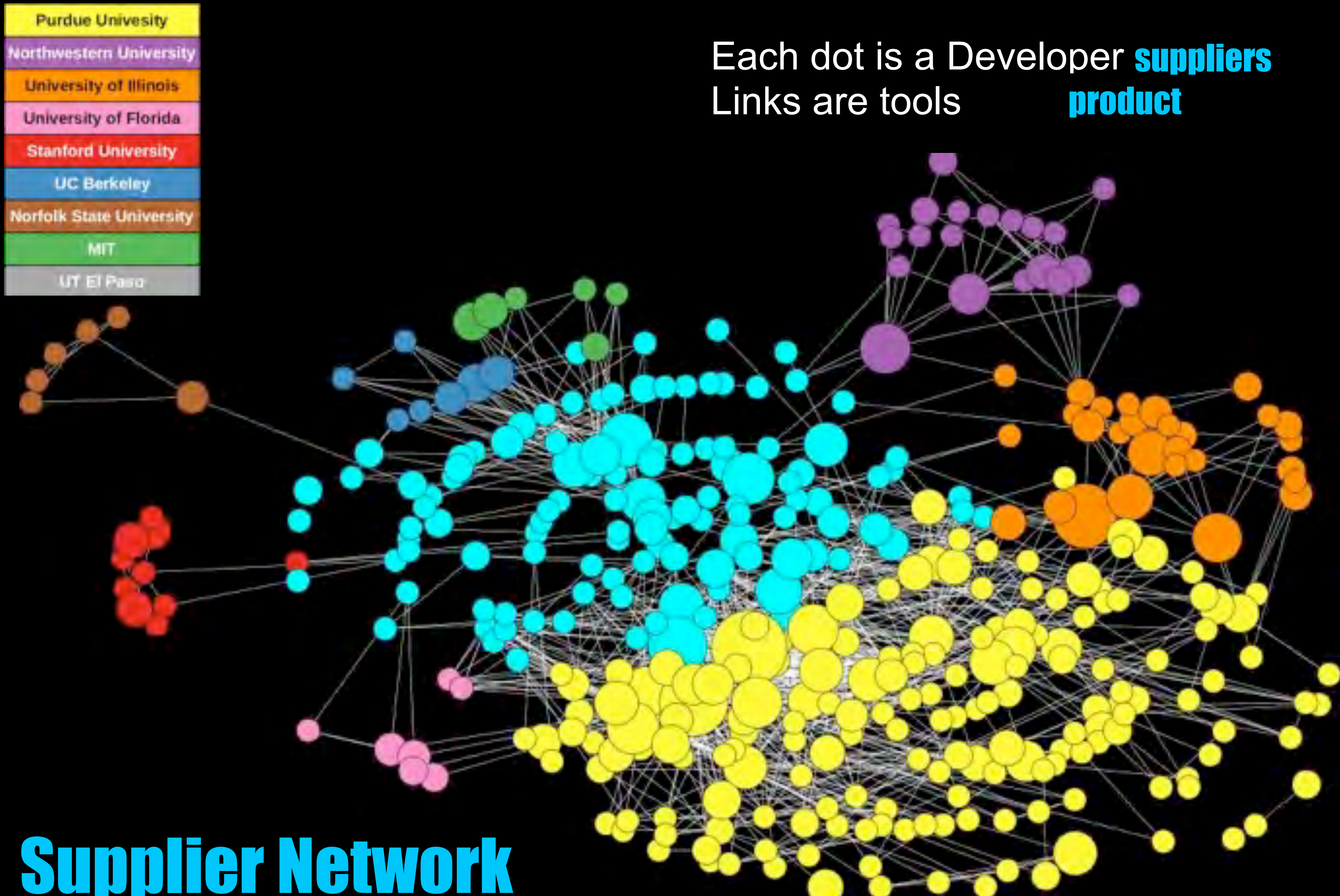
Continual Engagement

380+ Developers NOT PAID by NCN

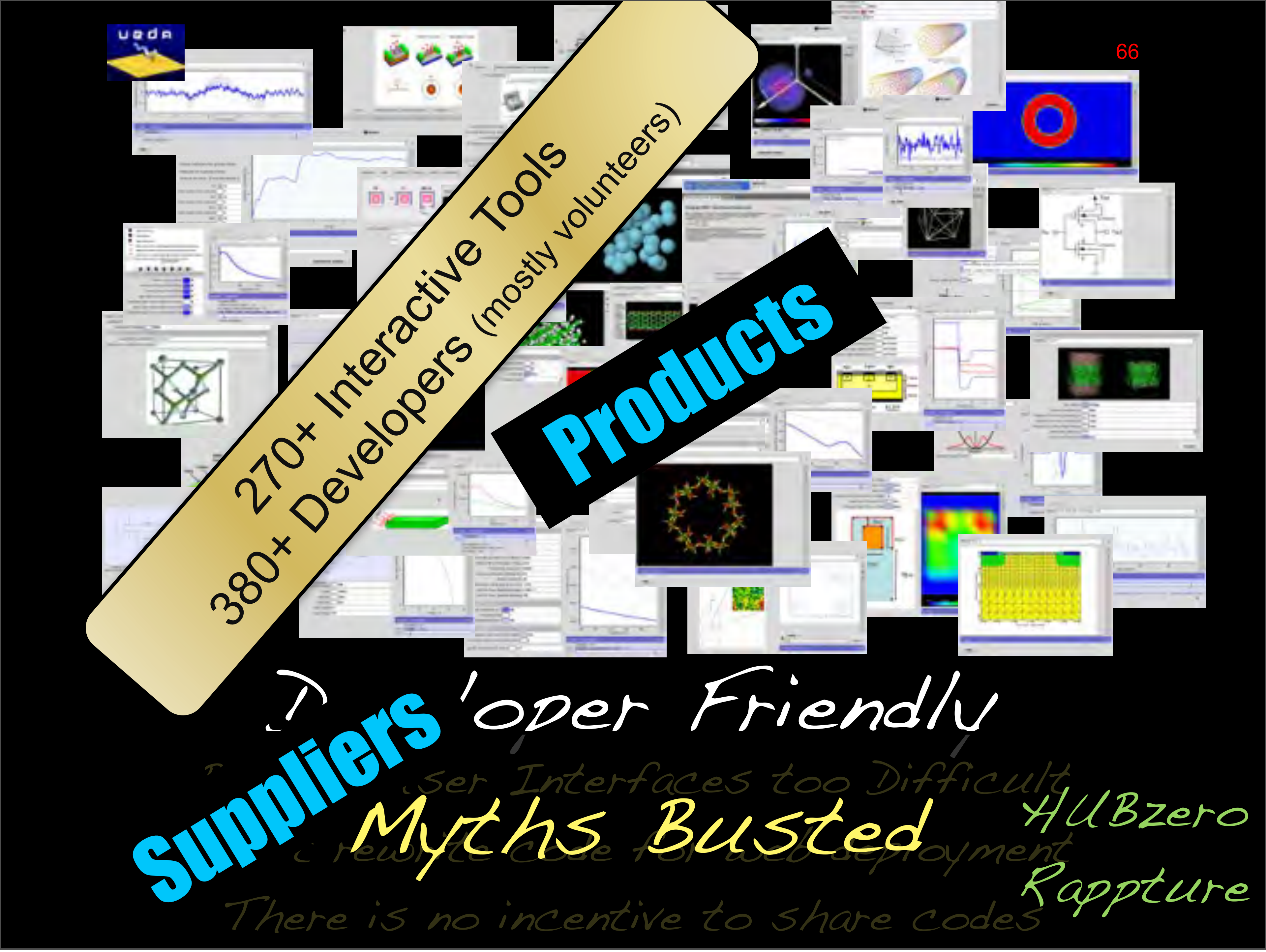
nanoHUB can prove it

Developer Collaboration Network

Each dot is a Developer **suppliers**
Links are tools **product**



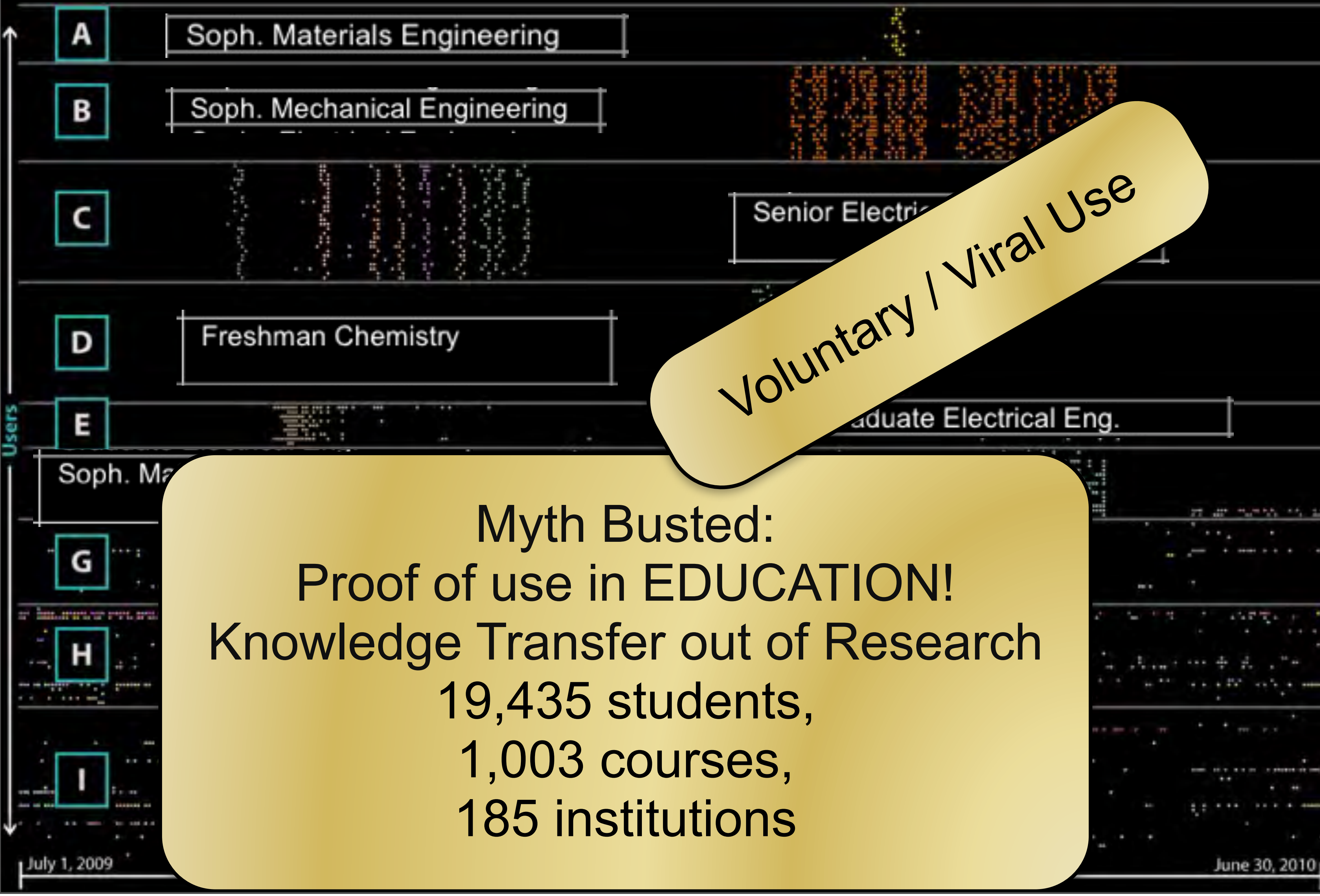
Supplier Network


66
270+ Interactive Tools
380+ Developers (mostly volunteers)
Products
Suppliers
Myths Busted
There is no incentive to share codes
HUBzero
Rappture

Emerged Myths

Cannot use research codes for education
Accessible (no installation)
Must write own code to do research
NO End-to-end Science Cloud Possible
Experimentalists cannot use research codes

Formal Education vs. Research

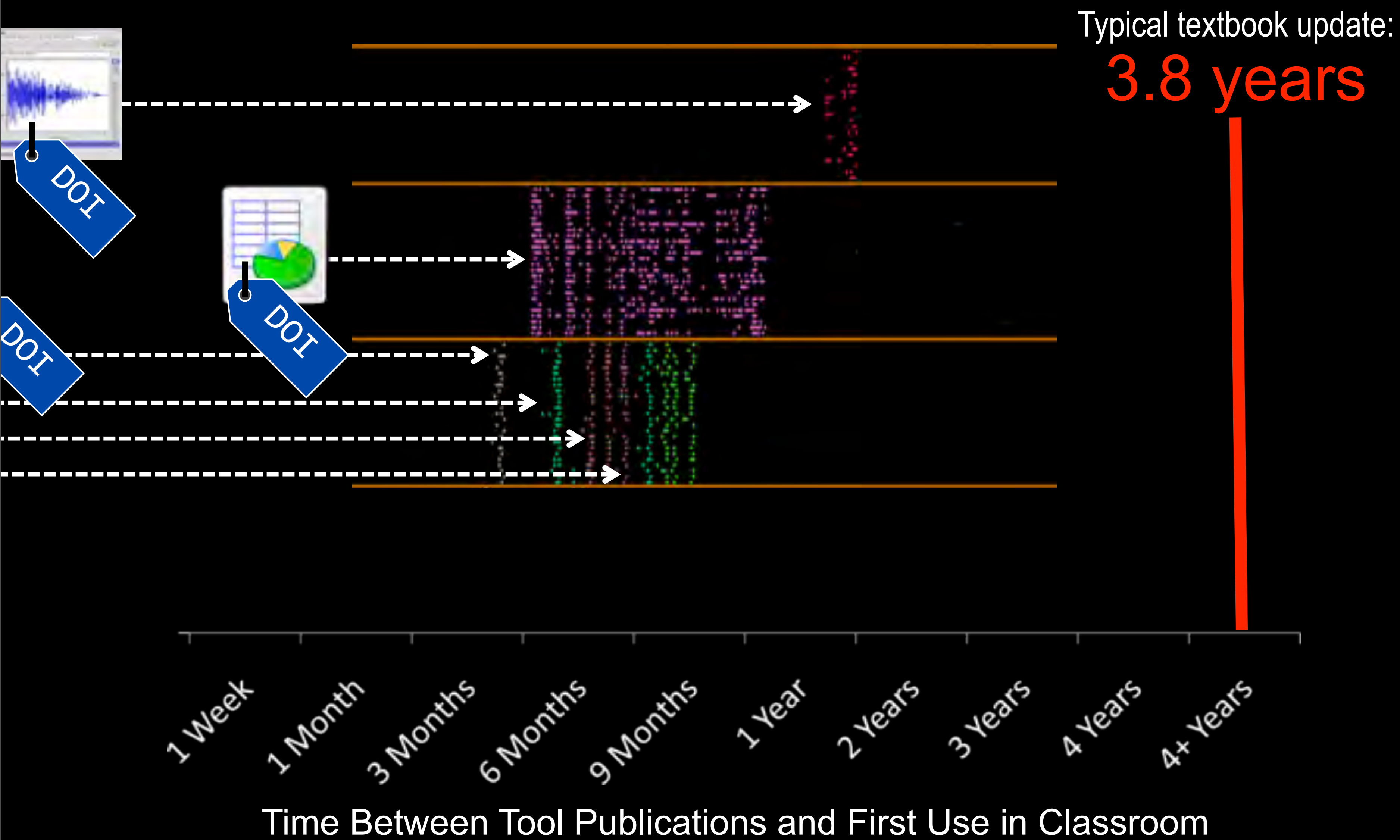


Voluntary / Viral Use

Myth Busted:
Proof of use in EDUCATION!
Knowledge Transfer out of Research
19,435 students,
1,003 courses,
185 institutions

Time to First Adoption

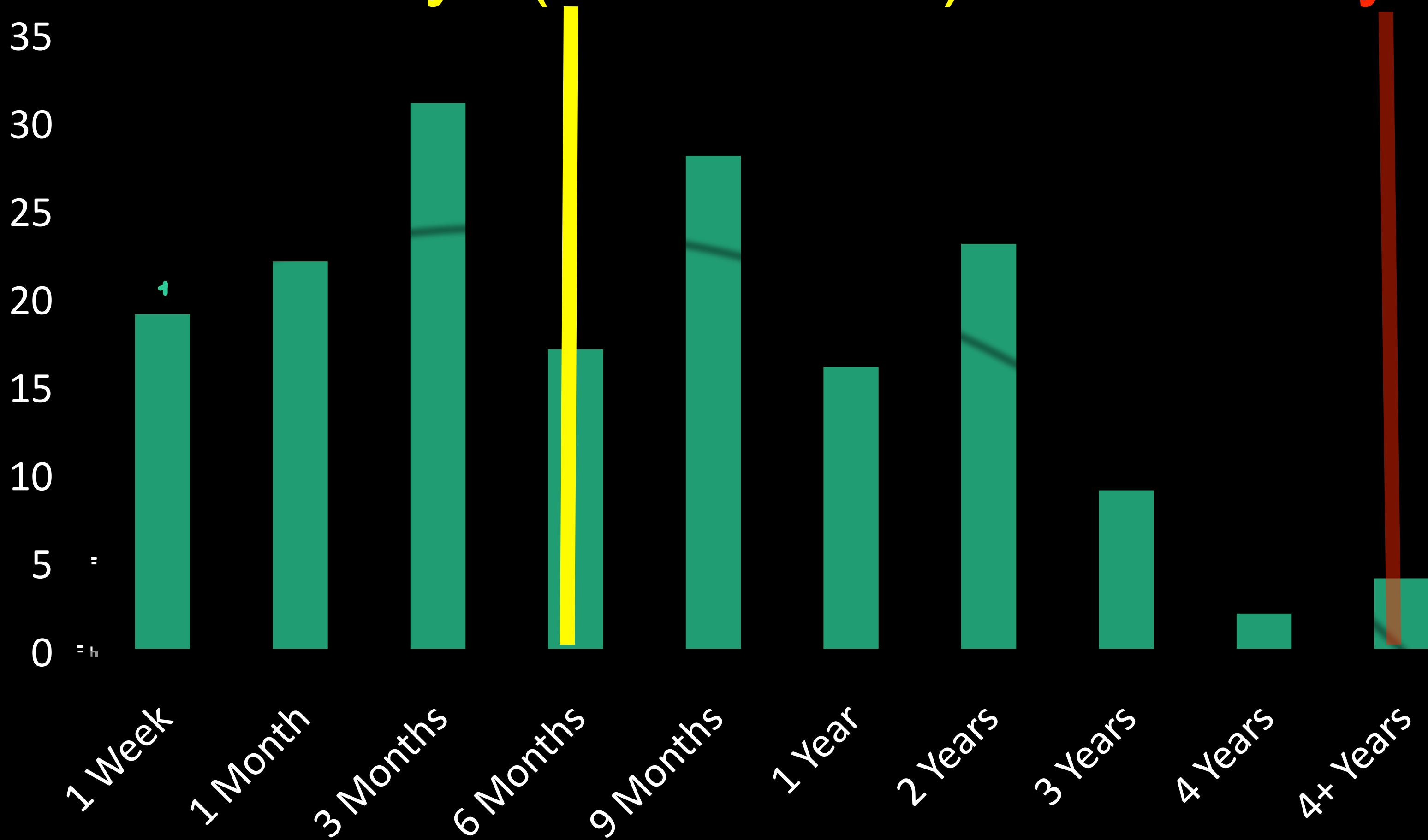
Time-To-Market



Rapid Adoption of Research
Time-To-Market

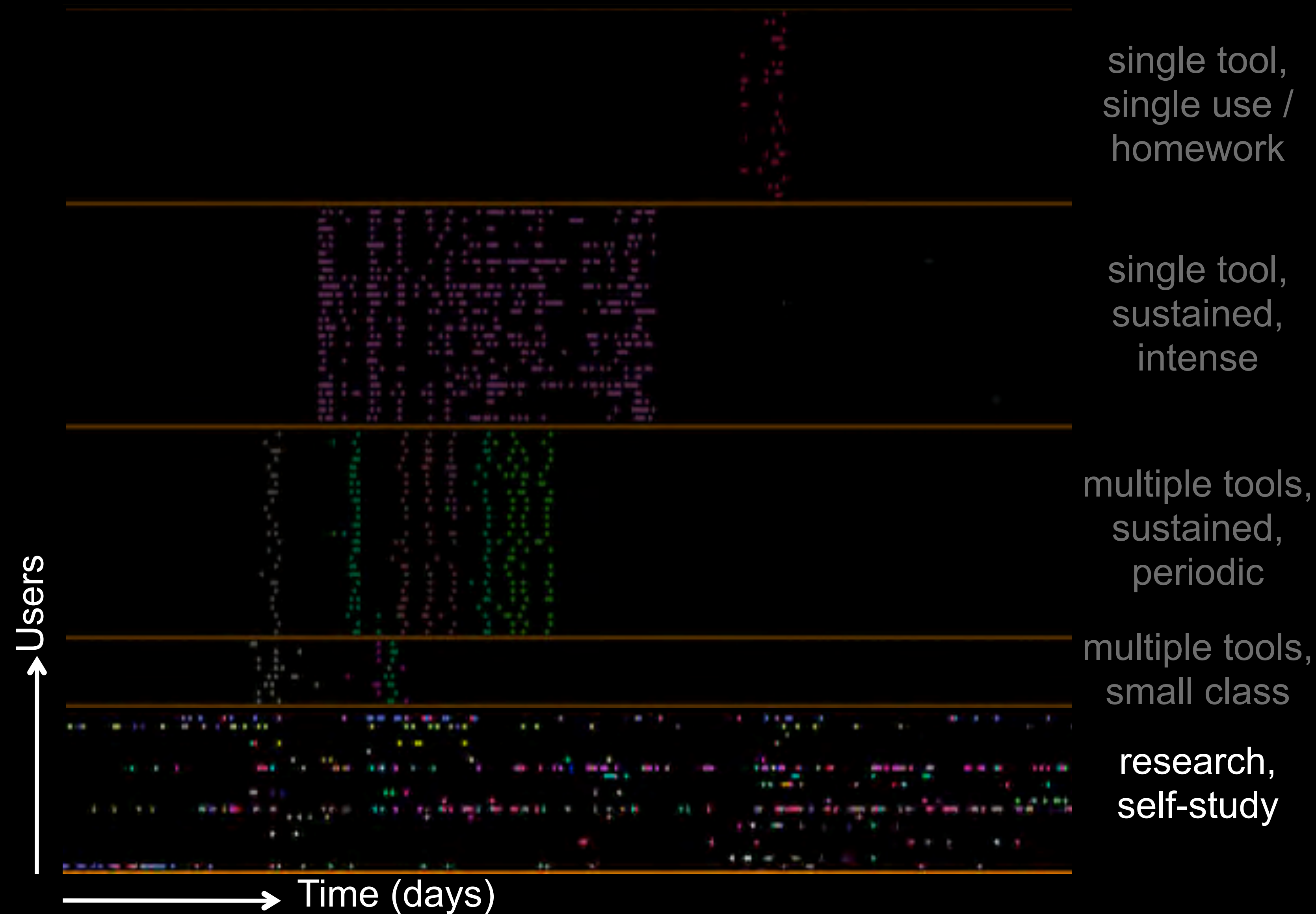
Median adoption time:
174 days (5.7 months)

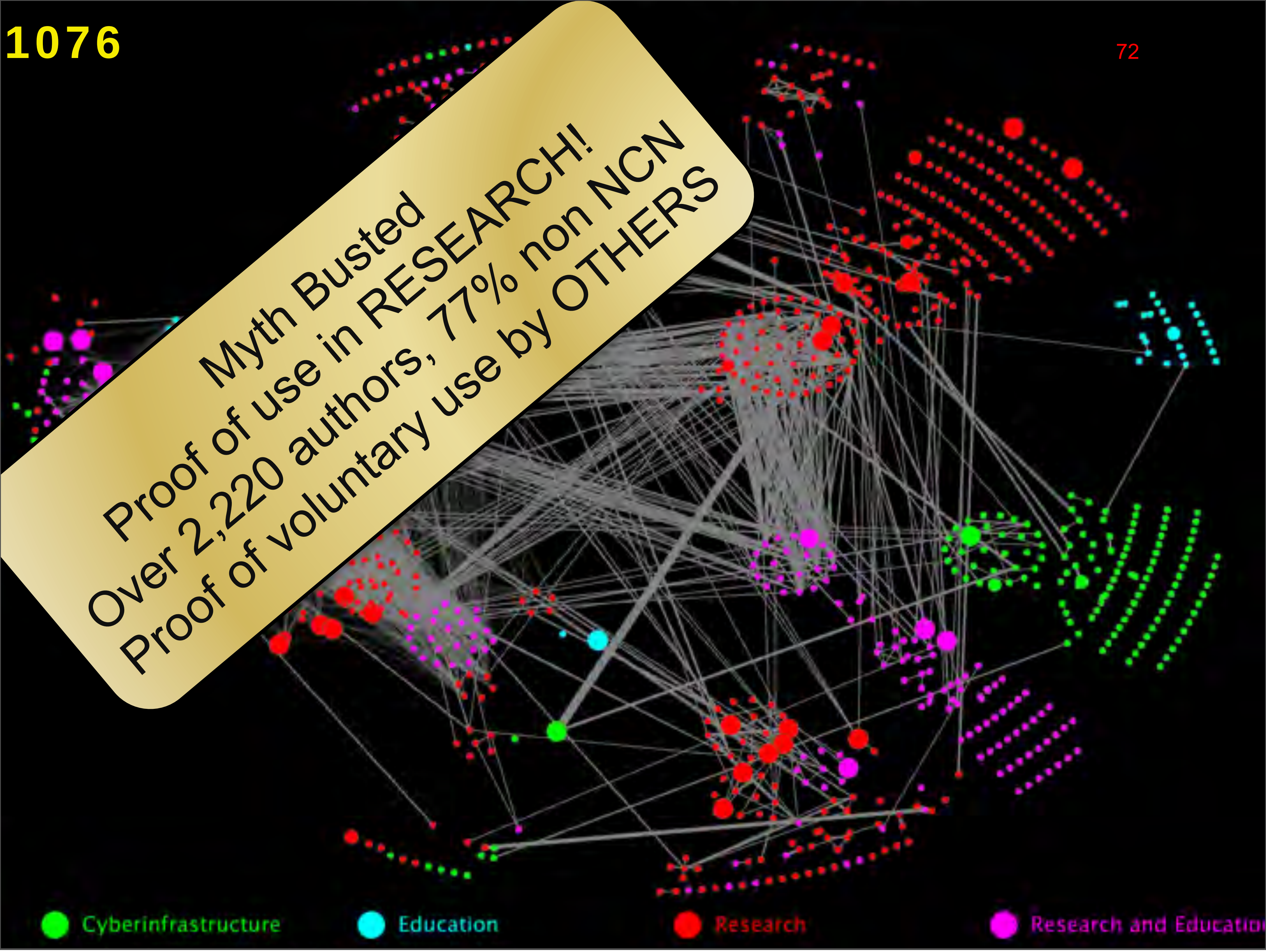
Typical textbook update:
3.8 years

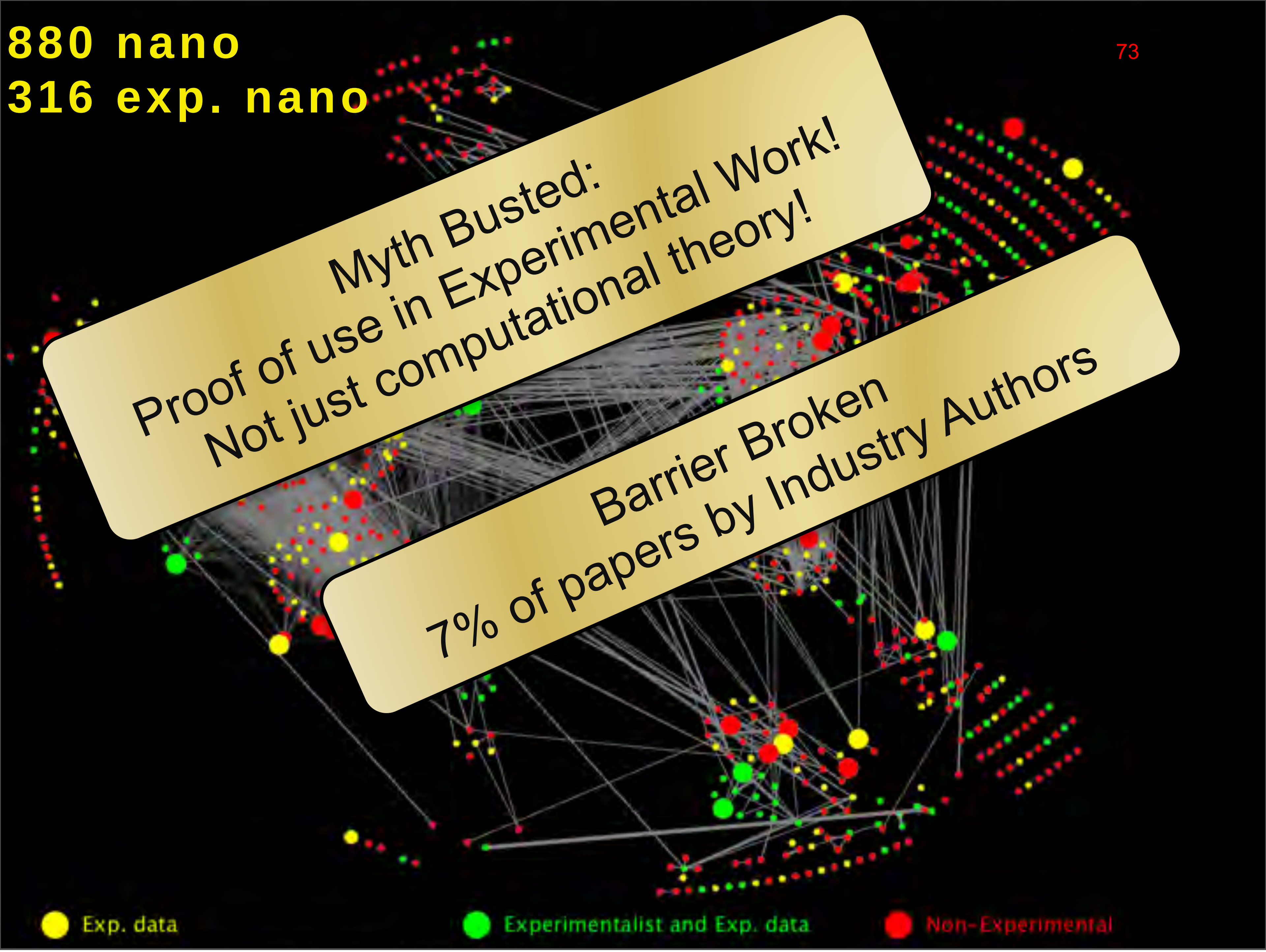


Time Between Tool Publications and First Use in Classroom

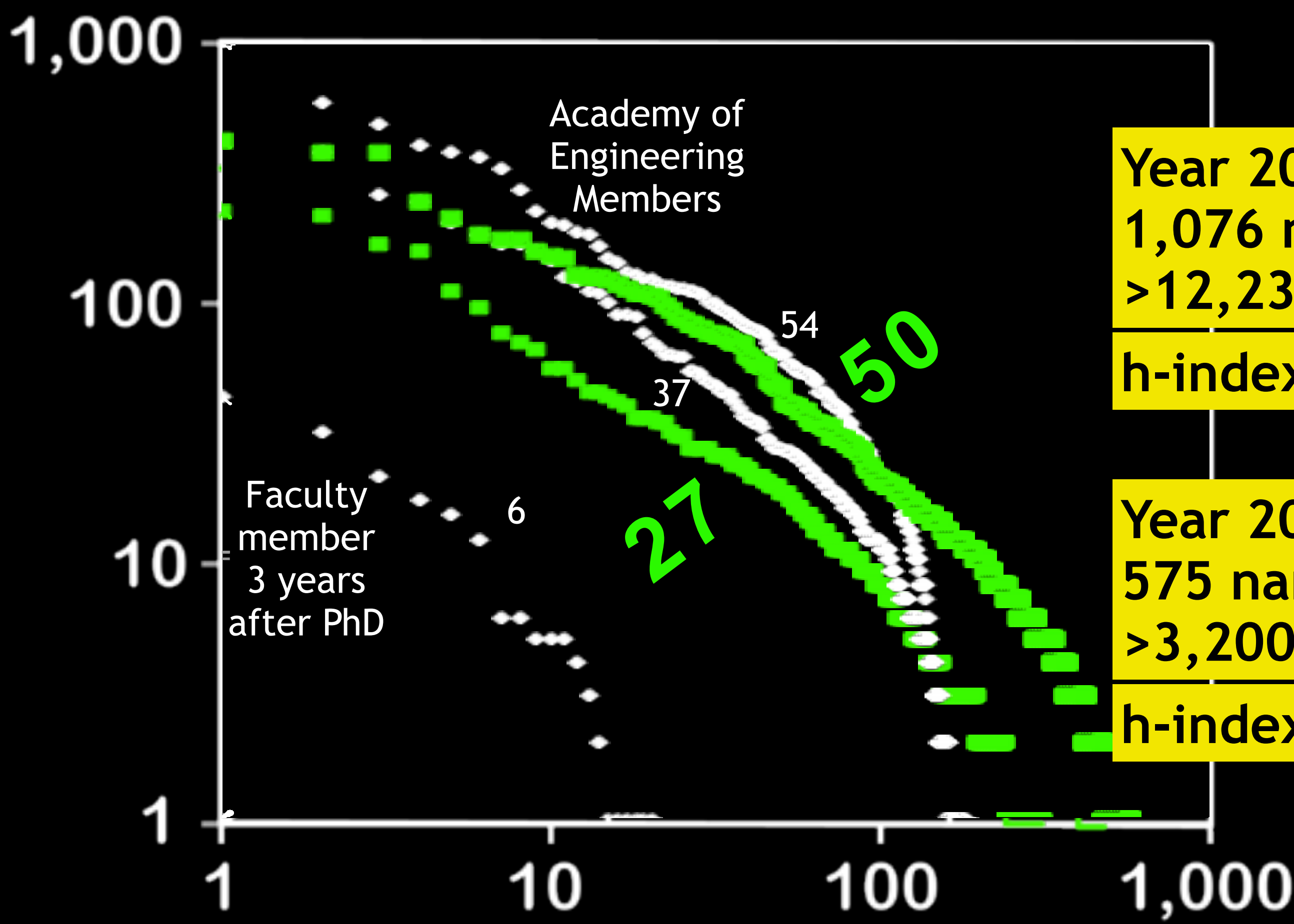
User Behavior Analysis
=> Is Research Possible?




Myth Busted
Proof of use in RESEARCH!
Over 2,220 authors, 77% non NCN
Proof of voluntary use by OTHERS
Cyberinfrastructure
Education
Research
Research and Educatio


880 nano
316 exp. nano
73
Myth Busted:
Proof of use in Experimental Work!
Not just computational theory!
Barrier Broken
7% of papers by Industry Authors
Exp. data
Experimentalist and Exp. data
Non-Experimental

h-index: Research Quality Indicator

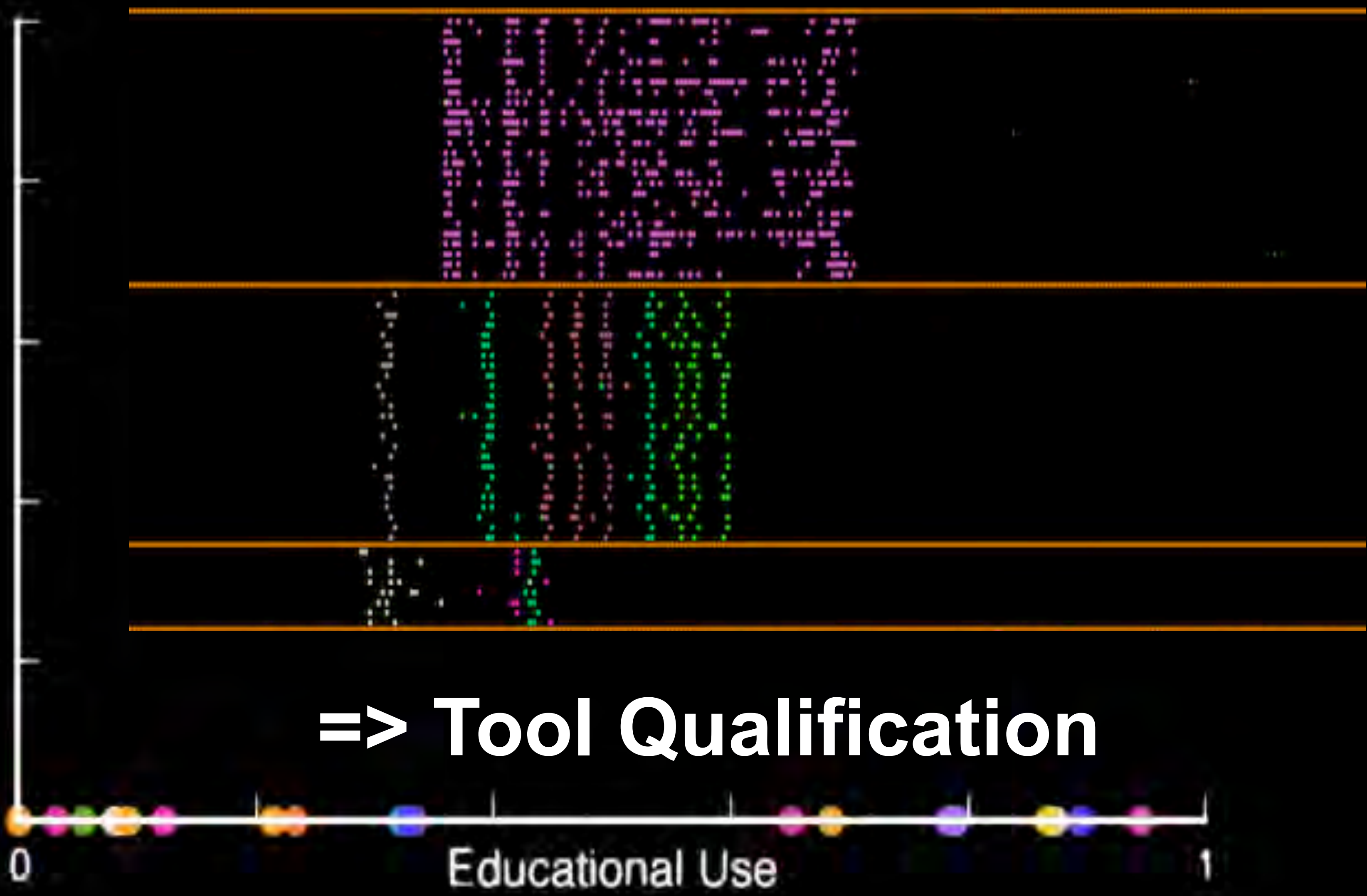


Year 2013
1,076 nanoHUB citations
>12,236 secondary citations
h-index: 51

Year 2009
575 nanoHUB citations
>3,200 secondary citations
h-index: 27

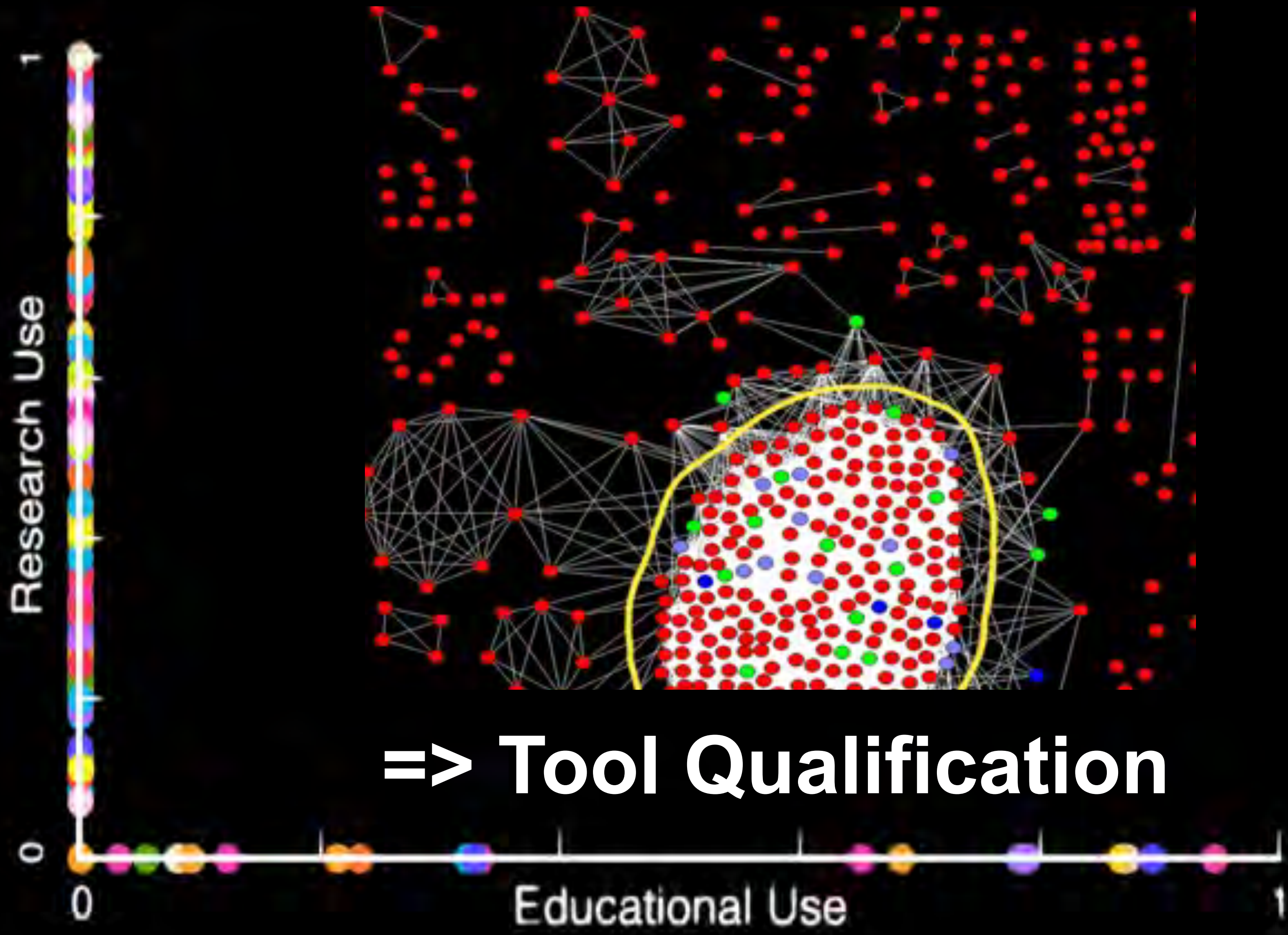
Is it good research?

Usage Patterns

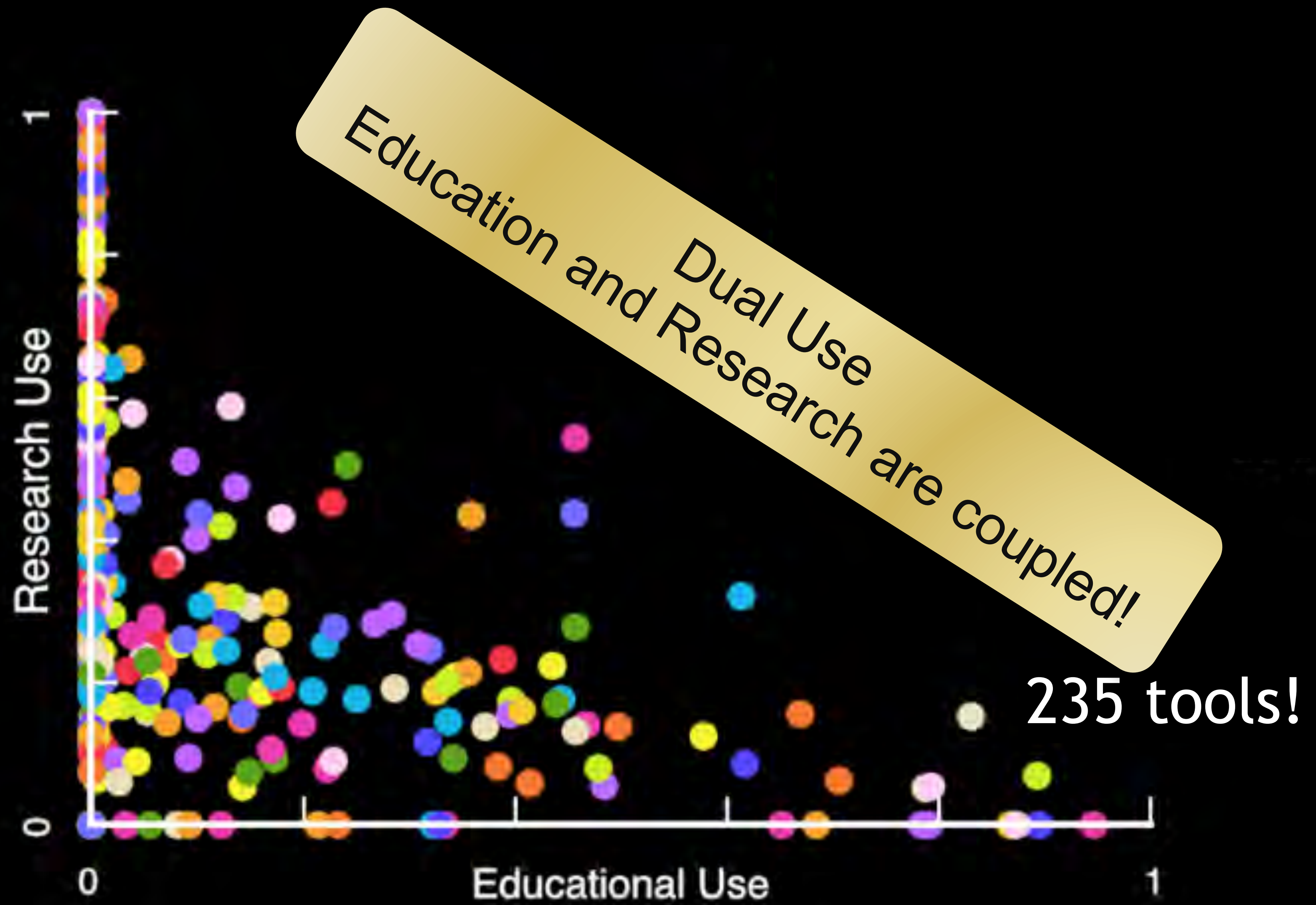


=> Tool Qualification

Literature Citations



=> Tool Qualification


Dual Use
Education and Research are coupled!
235 tools!
Research Use
Educational Use
0
1
0
1

Emerged Myths

User Friendly

Cannot use research codes for education

Must write own code to do research

Experimentalists cannot use research codes

Accessible (no installation)

NO End-to-end Science Cloud Possible

Developer Friendly

Building User Interfaces too Difficult

Must rewrite code for web deployment

There is no incentive to share codes

Customers

Market

Suppliers

Big Data Challenges

User Friendly

Think about a new class of Users!

=> End Users in Application Sciences

Accessible (no installation)

Free and no special access privileges

Developer Friendly

Think about YOUR grad students

and their incentives!

Customers

Market

Suppliers

Universidad Nacional Autónoma de México
Facultad de Estudios Cuautitlán

Towards a long-term e-Science infrastruture in Latin America

BIG DATA BIG NETWORK
Big Opportunities for Collaboration between Mexico and California
10-11 Oct 2013 • Ensenada, Mexico

Jesús Cruz Guzmán
CUDI/CLARA-TT / UNAM
cruz@unam.mx

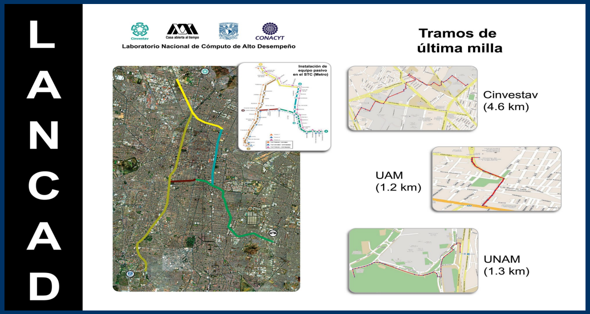
Science Gateway

cudi

RedCLARA
+ Red + Ciencia

Background

- Grid projects participation
- Laboratorio Nacional de Grids de Supercómputo
- Laboratorio Nacional de Cómputo de Alto Desempeño



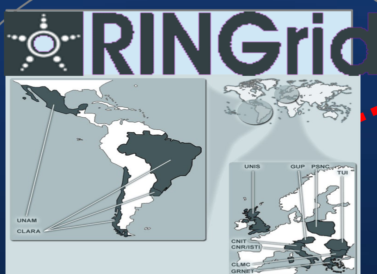
JRU-MX, Instituciones integrantes

- Centro de Investigación Científica y de Educación Superior de Ensenada,
- Corporación Universitaria para el Desarrollo de Internet, A.C.,
- Instituto Tecnológico de Veracruz,
- Instituto Tecnológico y de Estudios Superiores de Monterrey,
- Instituto Politécnico Nacional,
- Universidad Politécnica del Estado de Morelos*
- Universidad Autónoma del Estado de Morelos,
- Universidad de Sonora,
- Universidad Michoacana de San Nicolás de Hidalgo,
- Universidad Nacional Autónoma de México
- Universidad Autónoma de Aguascalientes*
- Universidad Juarez Autónoma de Tabasco*

* en proceso de incorporación

Grid service evolution

83



(testbed)

1/2006-12/2007

-2

(production infra.)

4/2008-3/2010

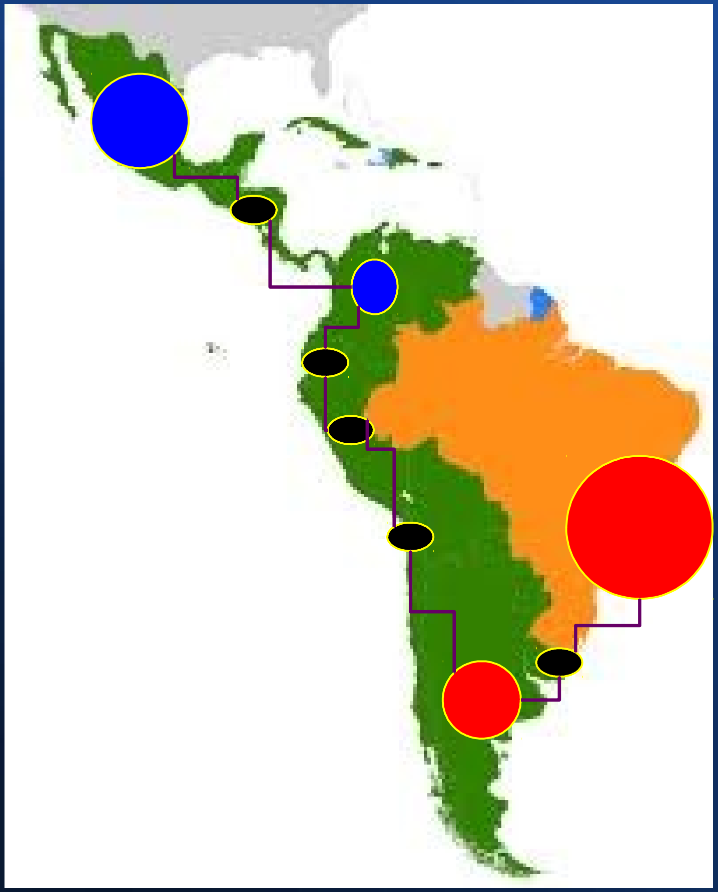
gisela

(sustainable service)

9/2010 – 8/2012

Formación de Recursos Humanos

SCALAC



HPC National Sistems

HPC National Comunities

Resource Centers

NREN + CLARA Facility

Work Plan

- Service Development and Administration

Services in production

- *Network Services*
- *Grid*

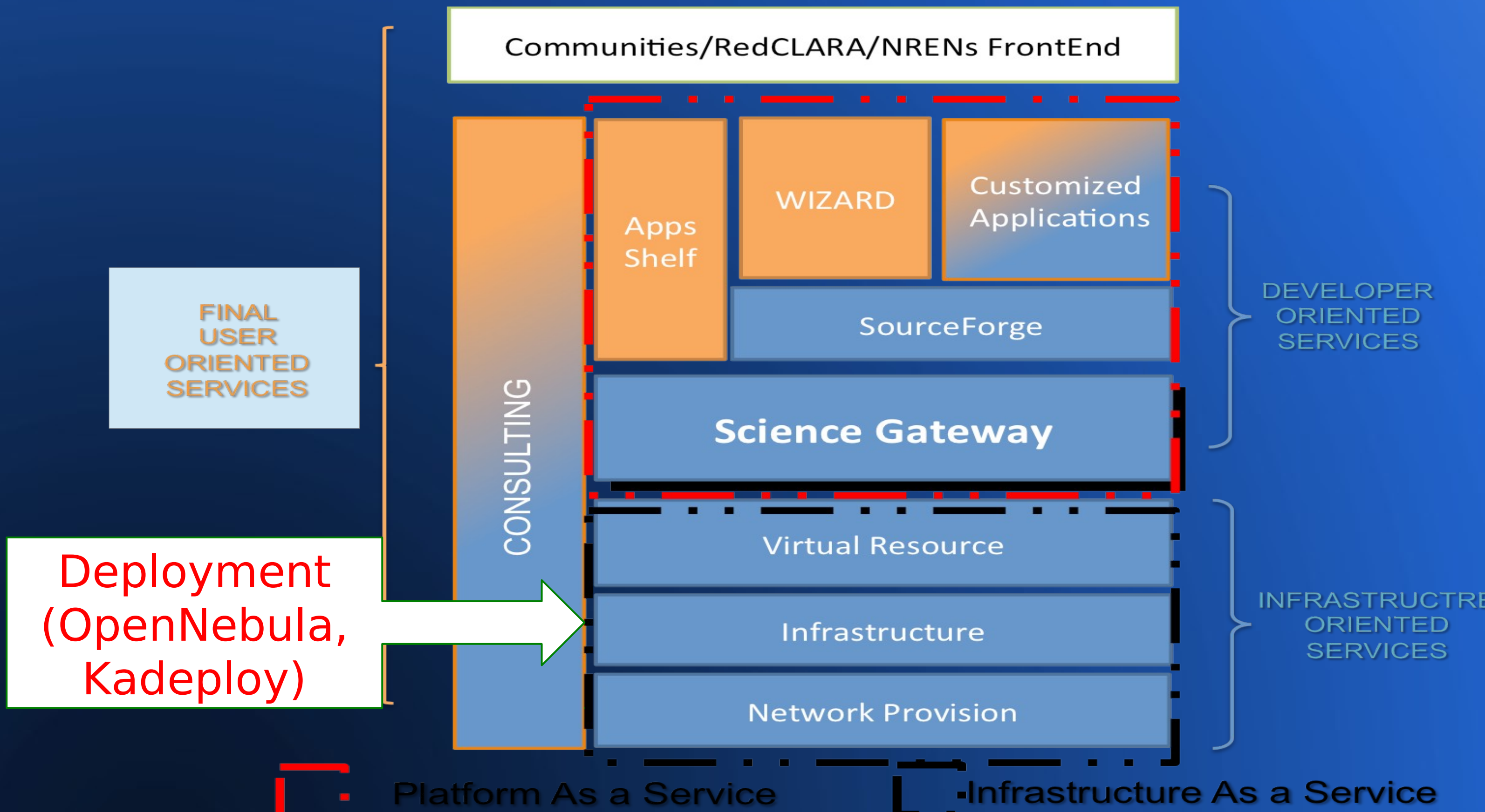
New services incorporation

- *Cloud*
- *HPC*
- *RIS*

New Services required

- *New applications*
- *Meta applications*
- *Technology for movil devices*

Service Architecture



E-Infrastructure International Support

Figure 22: Service tasks at the Latin American level

Grid infrastructure inter operable con EGI

- Operational services

Universidad Nacional Autónoma de México, Universidad de los Andes de Colombia y Centro Brasileiro de Investigaciones Físicas.

GISELA VO: prod.vo.eu-eela.eu

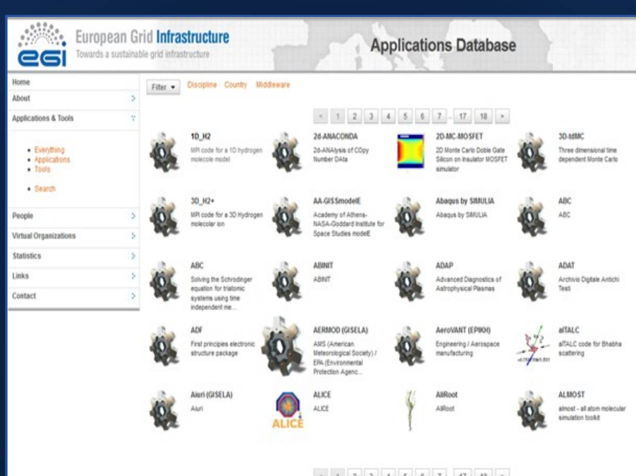
- Central services support: VOMS, MyProxy, BDII-TOP, WMS-LB, LFC, AMGA, UI, CREAM-CE, SE
- grid-unam@googlegroups.com

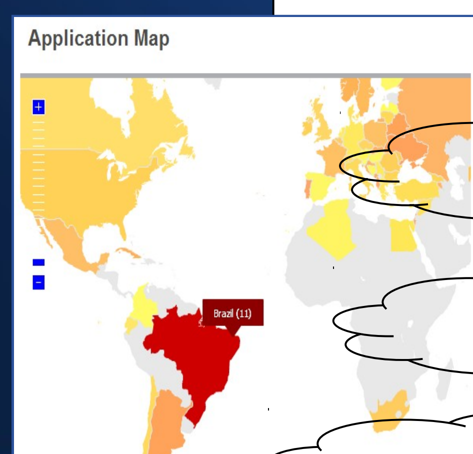
Science Gateway

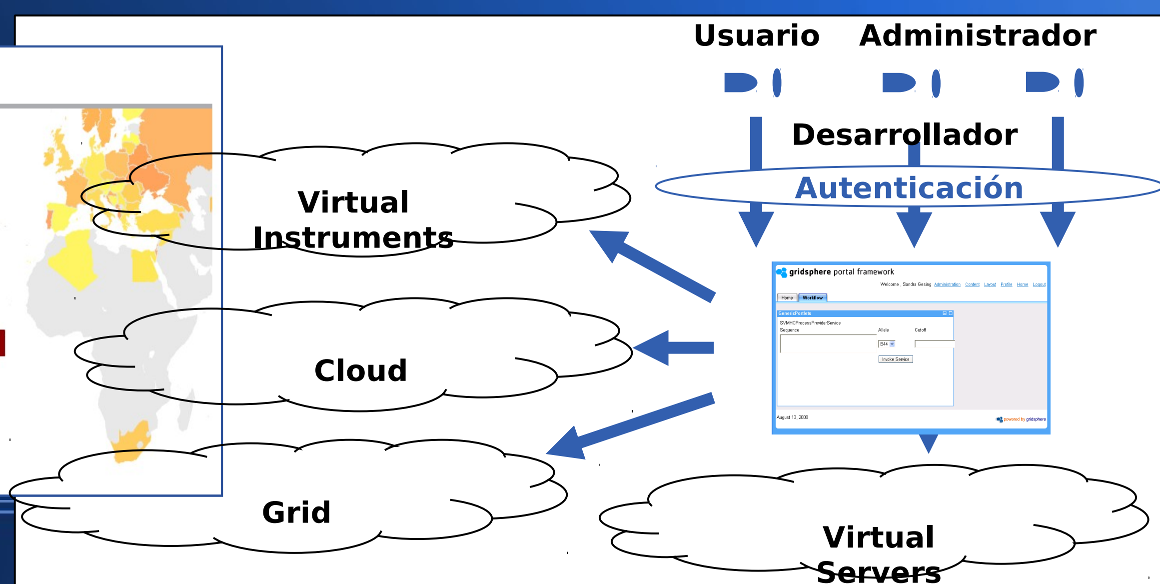
Un access for e-Science

89

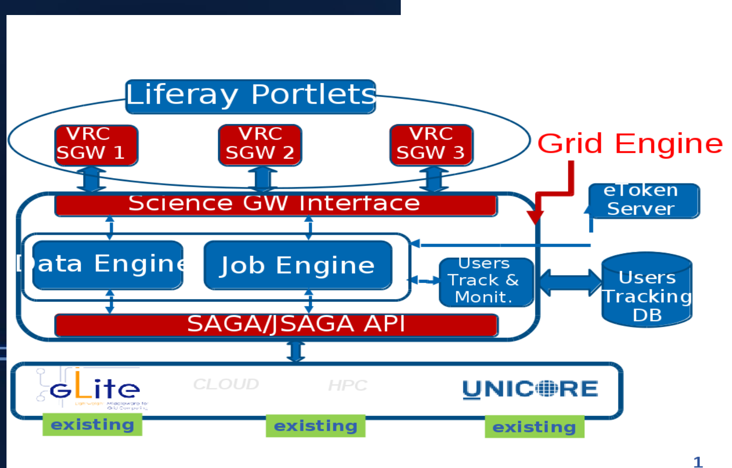
https://gisela-gw.ct.infn.it.

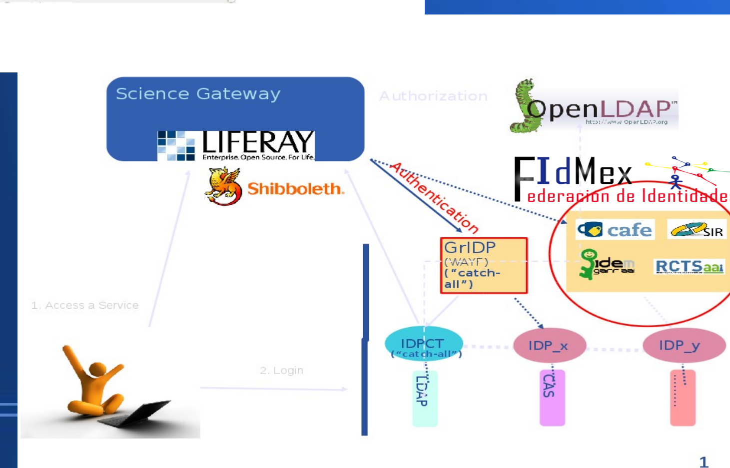






Arquitectura de los Servicios del Science Gateway





Transferencia del SGW para LA: Status

SG in EGI application data base

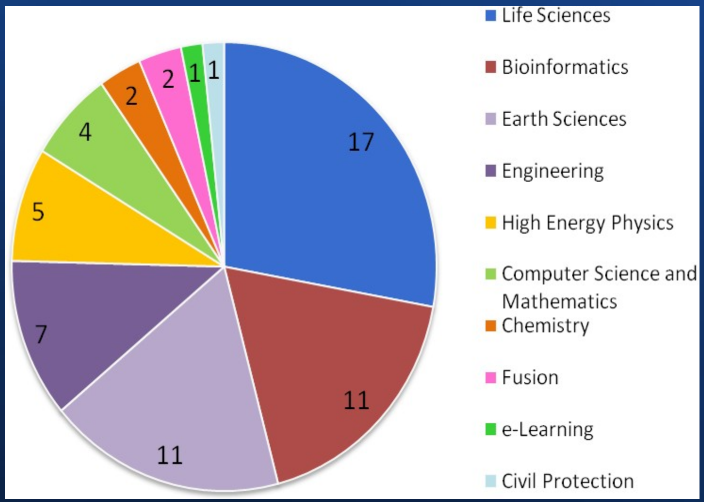
SG Infrastructure for LA+C in progress:

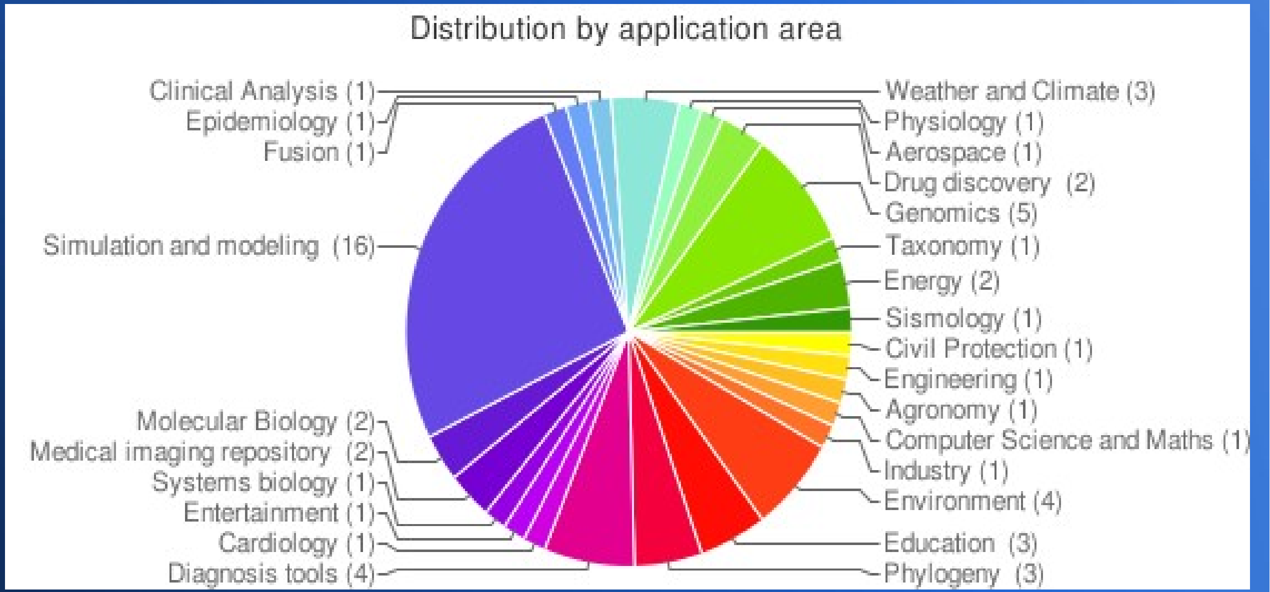
IdP, SP, LDAP, e-Token, Glassfish, Liferay, Grid Engine, applications.

Virtual Research Communities

Antecedentes: Aplicaciones portadas al entorno de Grid
EELA - EELA-2 - GISELA

93





Virtual Research Communities

Identified Communities

Bioinformatics

Industry

Earth Science

Cultural heritage

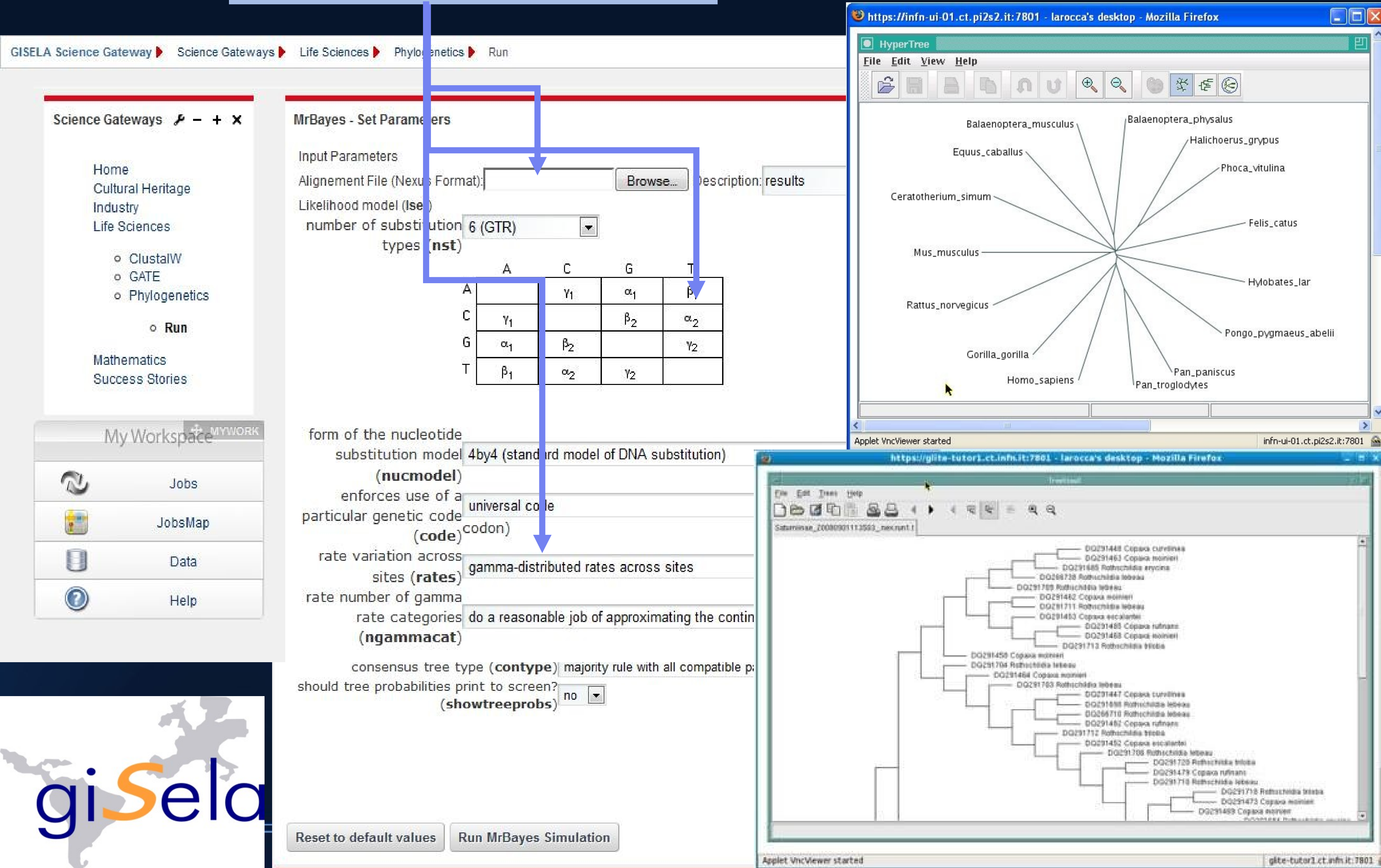
New Communities

The VRC-driven GISELA Science Gateway: Life Science Gateway

The VRC-driven GISELA Science Gateway: Phylogenetics (MrBayes)

96

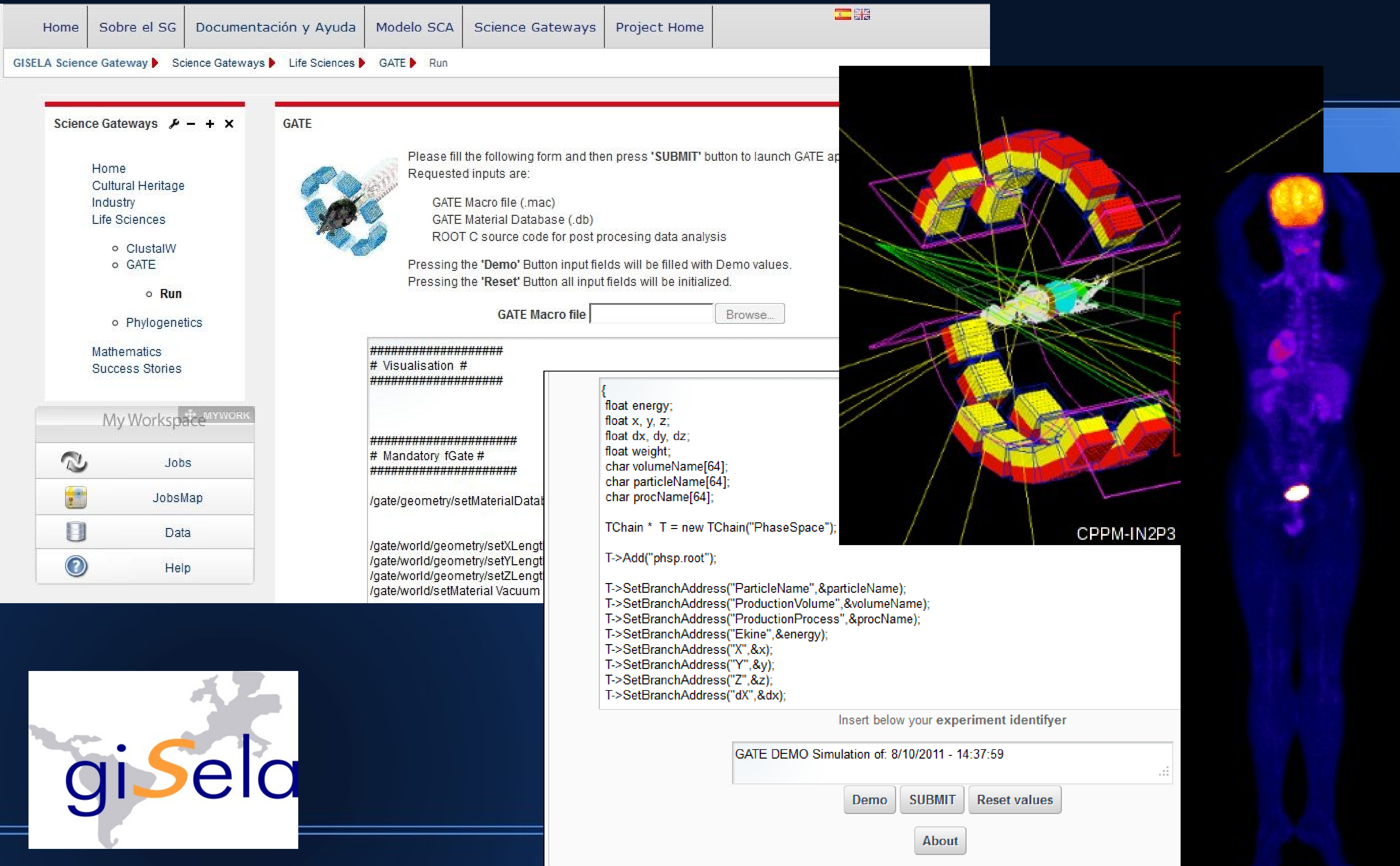
Input Parameters



	A	C	G	T
A		γ_1	α_1	β_1
C	γ_1		β_2	α_2
G	α_1	β_2		γ_2
T	β_1	α_2	γ_2	

The VRC-driven GISELA Science Gateway: GATE

97



giSela

Earth Sciences

Gracias

?

Co-ordination & Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing

A Standard-based Science Gateway Framework to Seamlessly Access HPC, Grid and Cloud Resources Distributed Worldwide
Big Data, Big Network Workshop – 10 October 2013

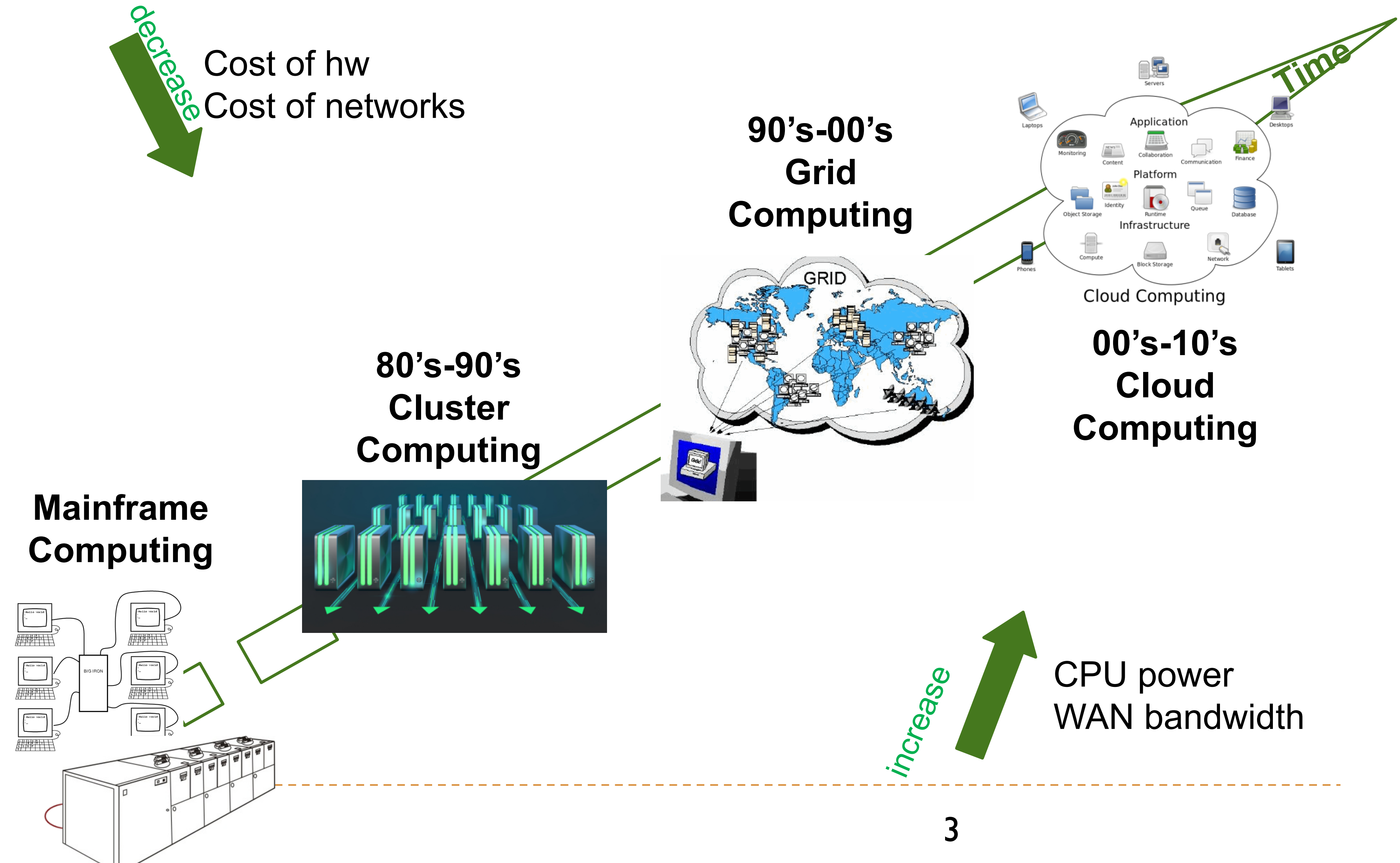
Roberto Barbera and Diego Scardaci – INFN Catania - Italy
{roberto.barbera, diego.scardaci}@ct.infn.it

Research Infrastructures – Proposal n. 306819

Outline

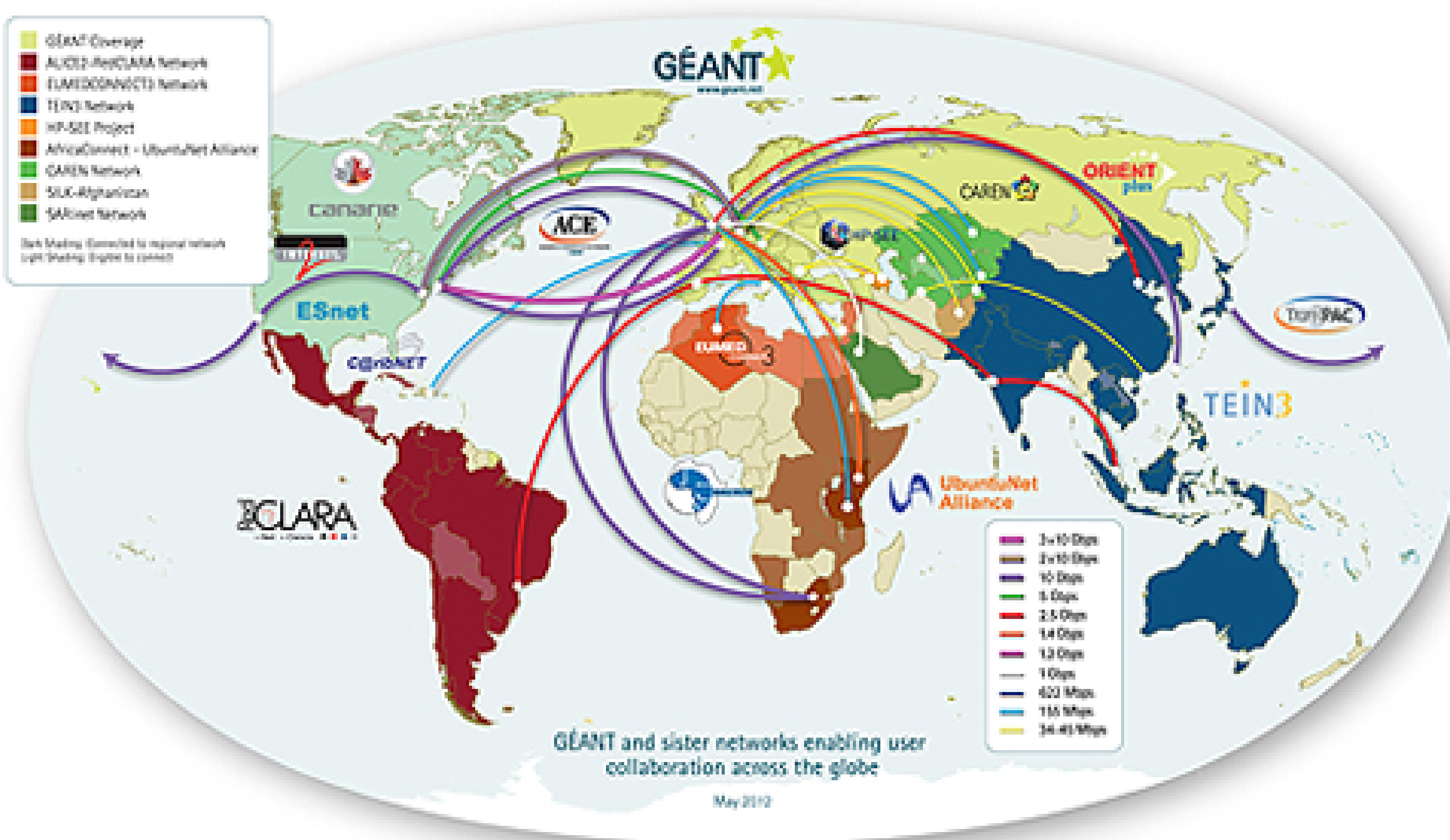
- Introductory concepts and driving considerations
- Vision and use cases
- Current results
- Activities in Latin America
- Summary and outlook

Evolution of distributed computing

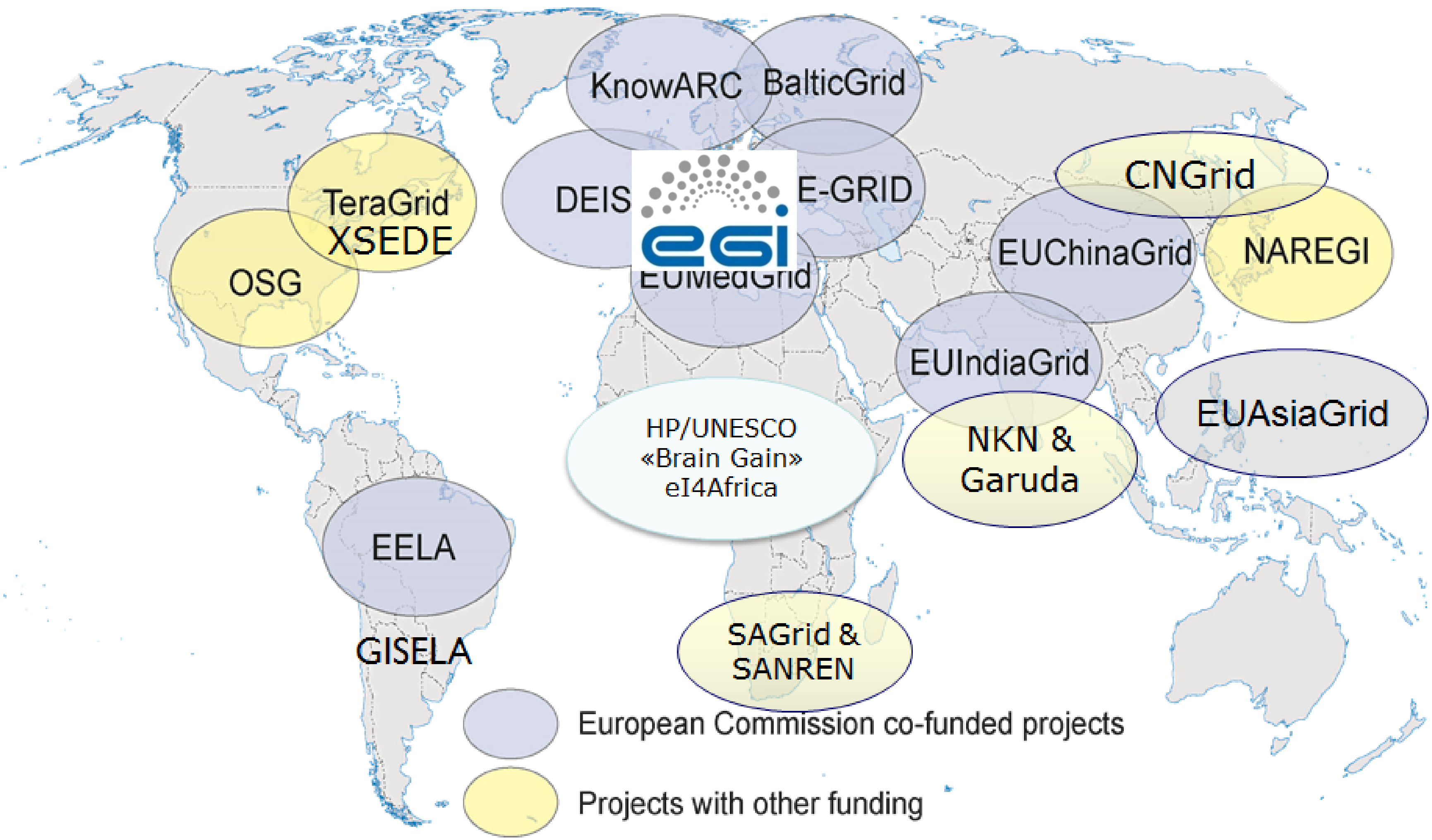


Research Networks at "global" scale

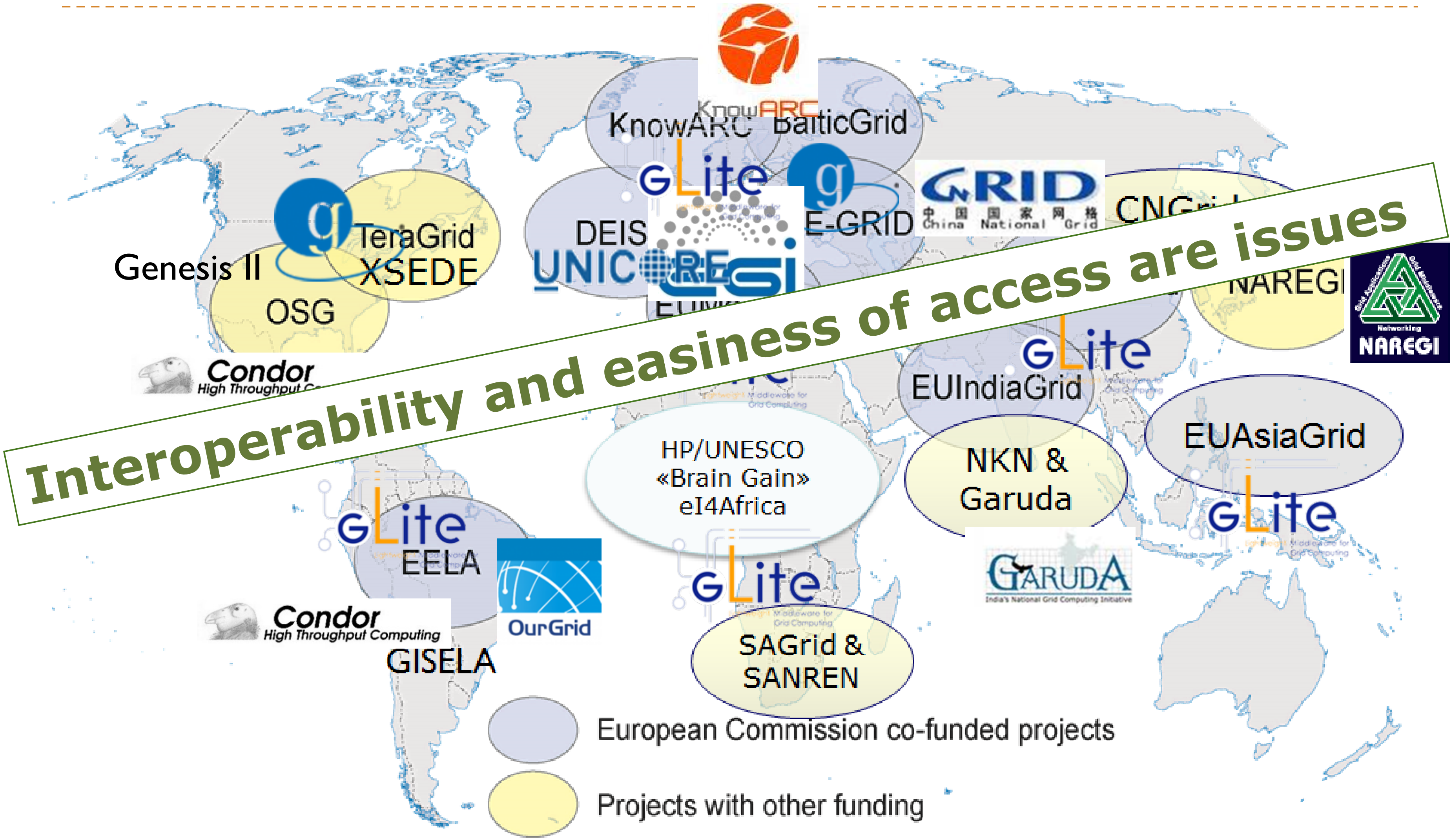
103



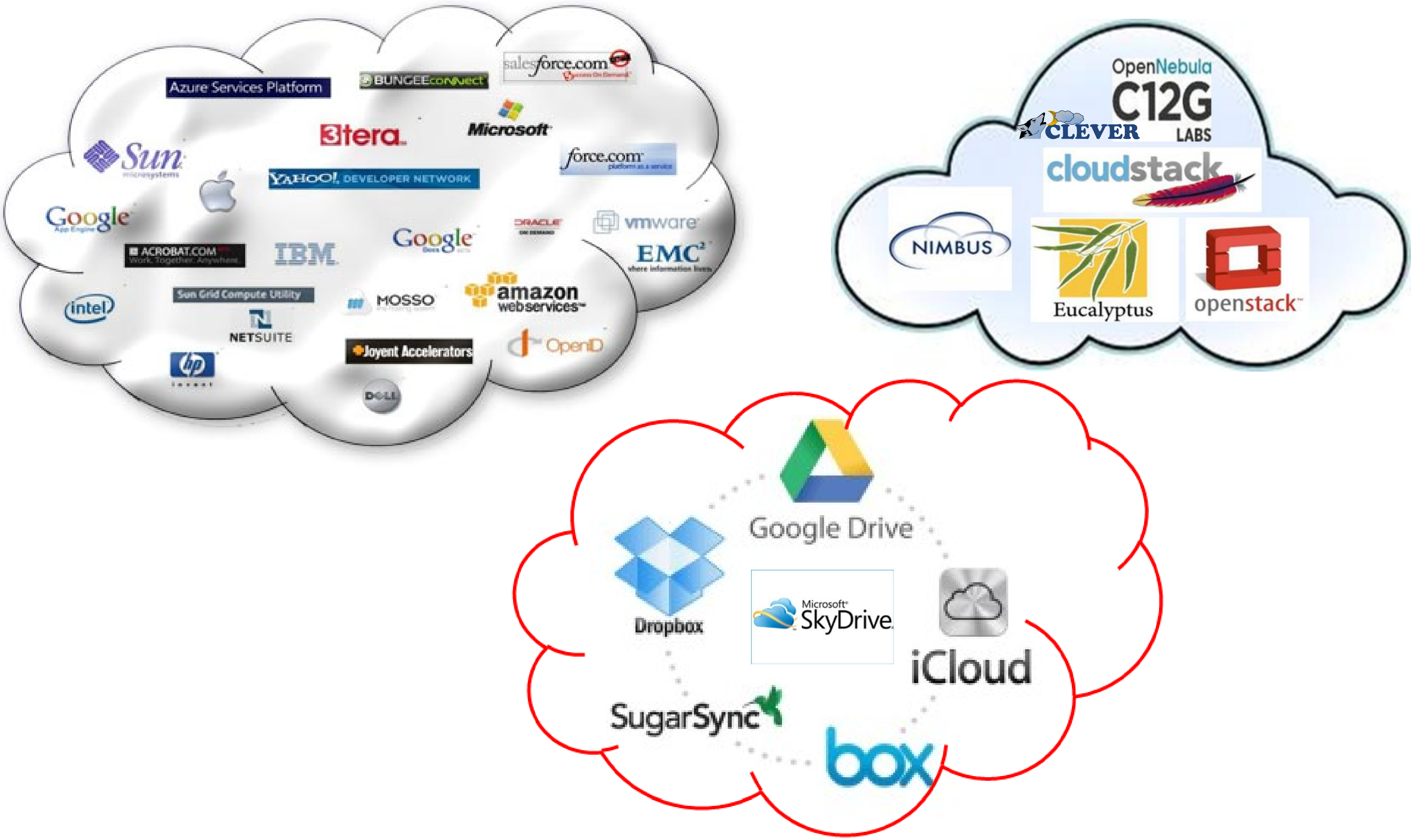
The “Global” Grid



The "non-global" middleware



Cloud «sky» is not less 106 . «cloudy»



The CHAIN-REDS Project
(www.chain-project.eu)

Started: 1 Dec 2012
Duration: 30 months
Targeted regions: Africa, Arab Region, Latin America, China, India, and Far-East Asia

Objectives

- Extend and consolidate the international cooperation of Europe with other regions of the world in the domain of e-Infrastructures for R&E, thus reinforcing the scientific collaboration and broadening the impact of the European Research Area
- Promote, coordinate and support the effort of a critical mass of non-European e-Infrastructures for R&E to collaborate with Europe by addressing interoperability and interoperation of Grids and other DCIs such as potential upcoming Cloud federations and HPC centres
- Study the opportunities of data sharing across different e-Infrastructures and continents thus widening the scope of the existing CHAIN Knowledge Base to Data Infrastructures and Cloud implementations
- Promote trust-building towards open Scientific Data infrastructures across the world regions, including organisational, operational and technical aspects
- Demonstrate the relevance of intercontinental cooperation in several scientific data fields addressing existing and emerging Virtual Research Communities (e.g. Earth Science, Climate Change, Genomics, etc.) and propose pragmatic approaches that could impact the everyday work of the single researcher, even if not structured in the framework of a Virtual Research Community
- Provide guidance and recommendations for roadmaps for long-term global collaboration in e-Infrastructures and harmonisation of existing policies. These are envisaged to act as input to policy and decision-making mechanism, harmonised with the European Digital Agenda and Horizon 2020

Interoperability and interoperation

(source: Wikipedia)

- According to ISO/IEC 2382-01 (Information Technology Vocabulary, Fundamental Terms), **interoperability** is *"The capability to communicate, execute programs, o... transfer data among various functio... that requires the us... knowledge of the uniq... those units“*

Adoption of standards is key

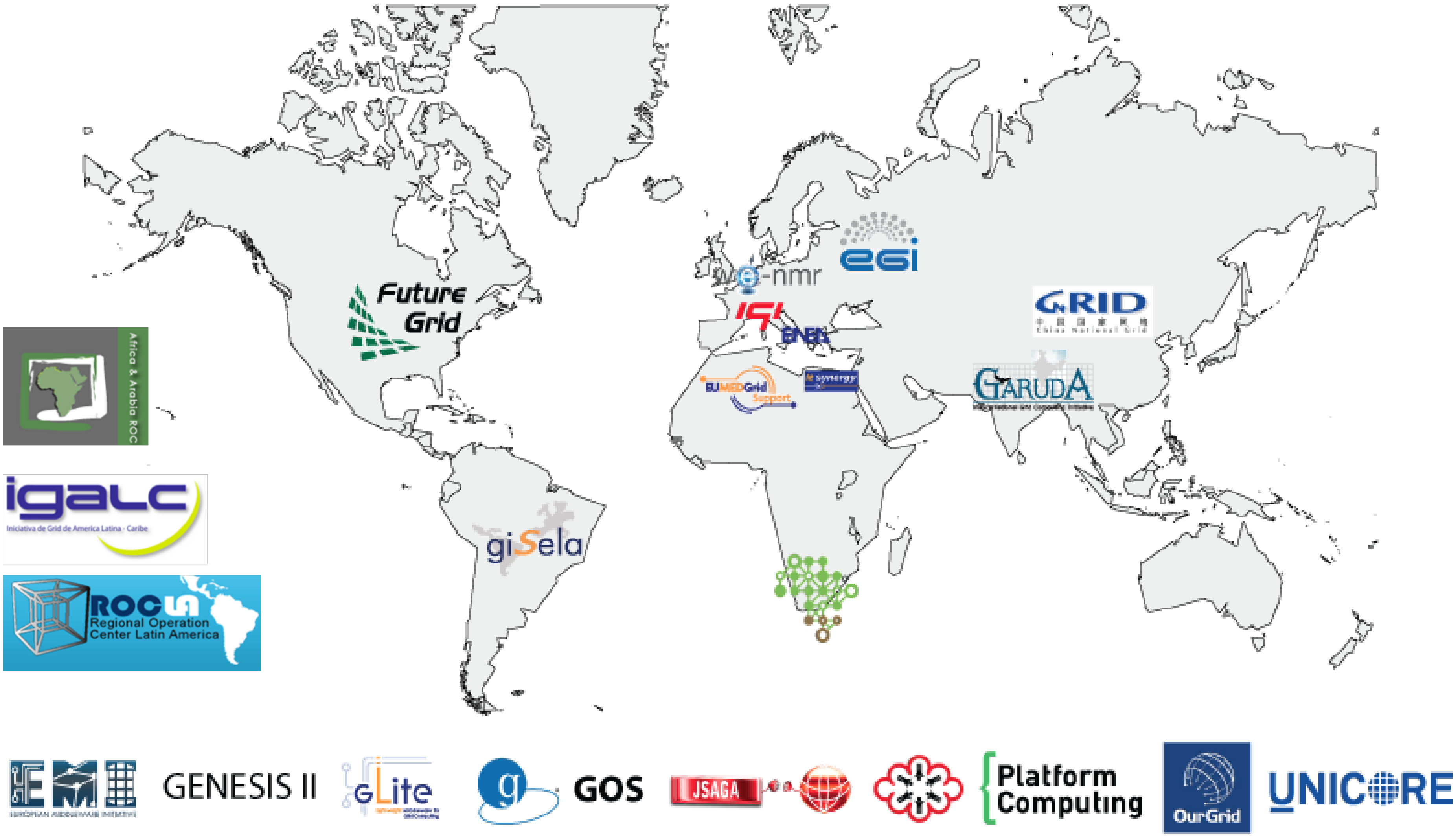
- In engineering, **interoperation** is the setup of ad hoc components and methods to make two or more systems work together as a combined system chain reds

Vision

- A **scientist** can seamlessly run applications on HPC machines, Grids and Clouds based on different middleware
  - (to demonstrate interoperability   use case #1)

- The **cloud tenant** of a real or virtual organisation can seamlessly and easily manage Cloud resources pledged by providers owning infrastructures based on different Cloud middleware stacks
  - (to demonstrate interoperation   use case #2)

CHAIN-REDS Demo contributors



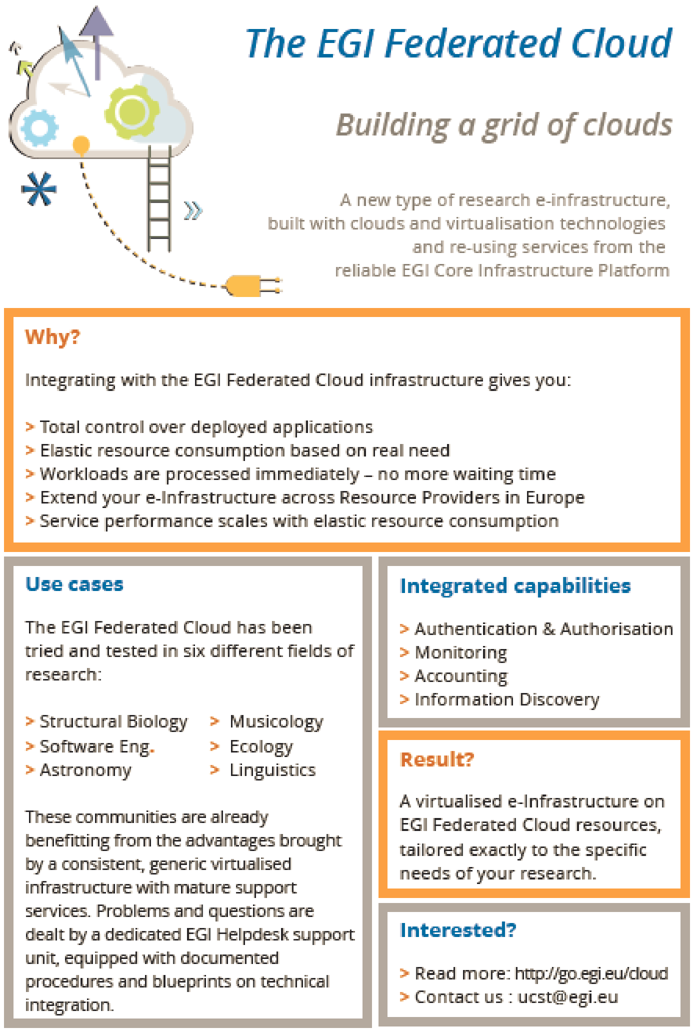
Use case #1 (scientist)

- A user can sign in on a Science Gateway using his/her federated credentials, select an application from a menu and seamlessly execute it on HPC machines, Grids and Clouds
  - The fractions of executions on the three different platforms can be adjusted to simulate the need to "boost" the resources in case of temporary peaks of activity

A reference infrastructure: the EGI Federated Cloud

Web: http://www.egi.eu/infrastructure/cloud/

Wiki: https://wiki.egi.eu/wiki/Fedcloud-tf



The EGI Federated Cloud

Building a grid of clouds

A new type of research e-infrastructure, built with clouds and virtualisation technologies and re-using services from the reliable EGI Core Infrastructure Platform

Why?

Integrating with the EGI Federated Cloud infrastructure gives you:

> Total control over deployed applications
> Elastic resource consumption based on real need
> Workloads are processed immediately – no more waiting time
> Extend your e-Infrastructure across Resource Providers in Europe
> Service performance scales with elastic resource consumption

Use cases

The EGI Federated Cloud has been tried and tested in six different fields of research:

> Structural Biology
> Software Eng.
> Astronomy
> Musicology
> Ecology
> Linguistics

These communities are already benefitting from the advantages brought by a consistent, generic virtualised infrastructure with mature support services. Problems and questions are dealt by a dedicated EGI Helpdesk support unit, equipped with documented procedures and blueprints on technical integration.

Integrated capabilities

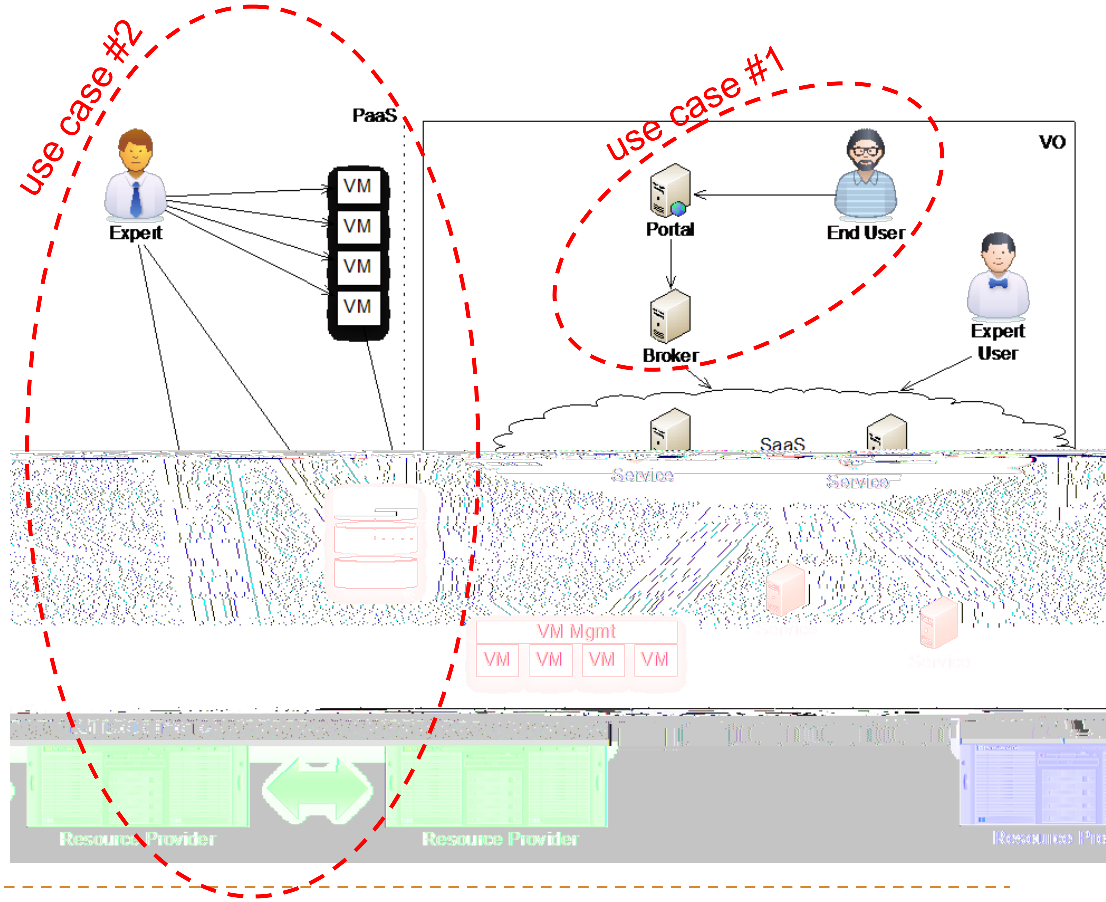
> Authentication & Authorisation
> Monitoring
> Accounting
> Information Discovery

Result?

A virtualised e-Infrastructure on EGI Federated Cloud resources, tailored exactly to the specific needs of your research.

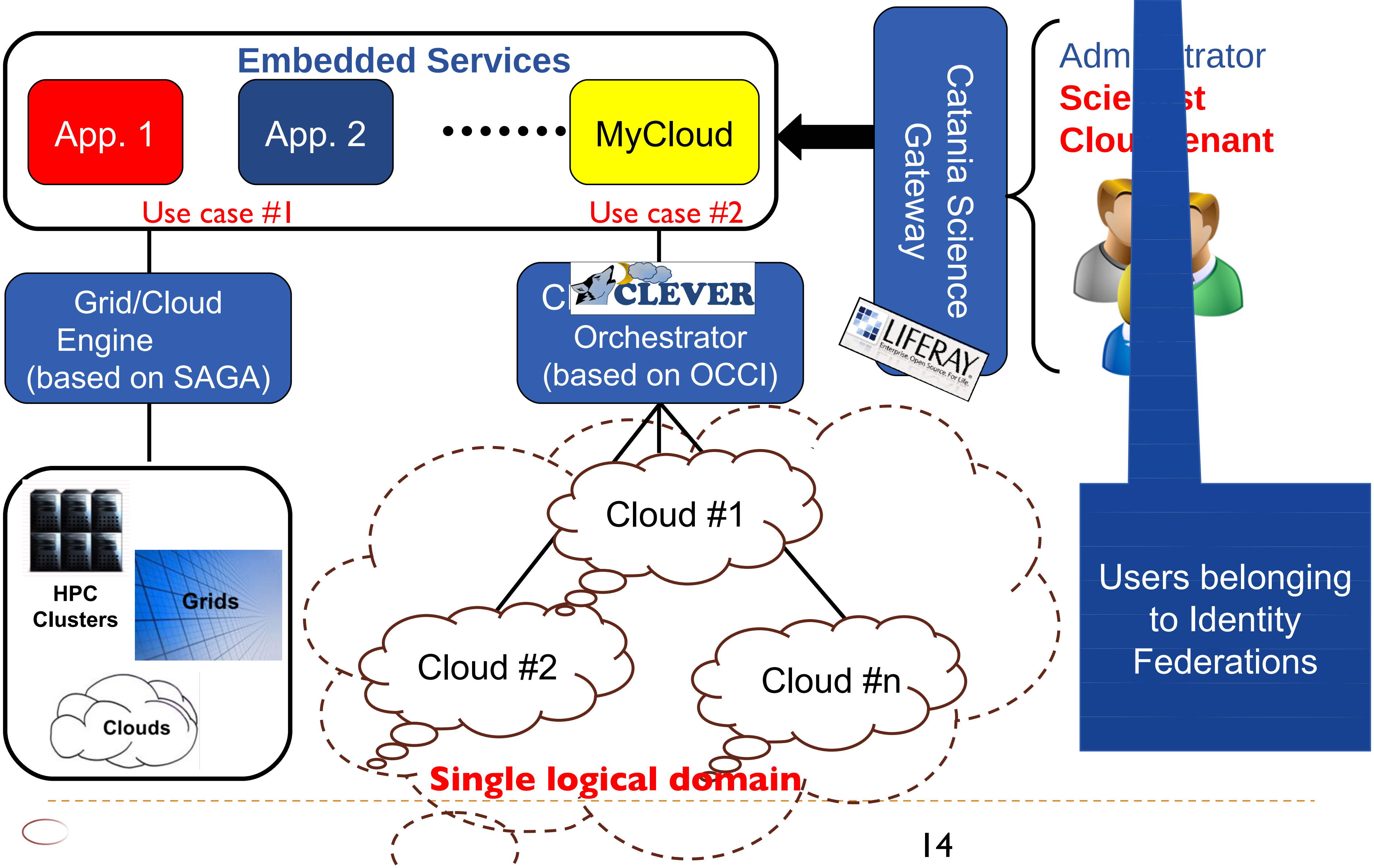
Interested?

> Read more: http://go.egi.eu/cloud
> Contact us : ucst@egi.eu



Catania Science Gateway Framework architecture

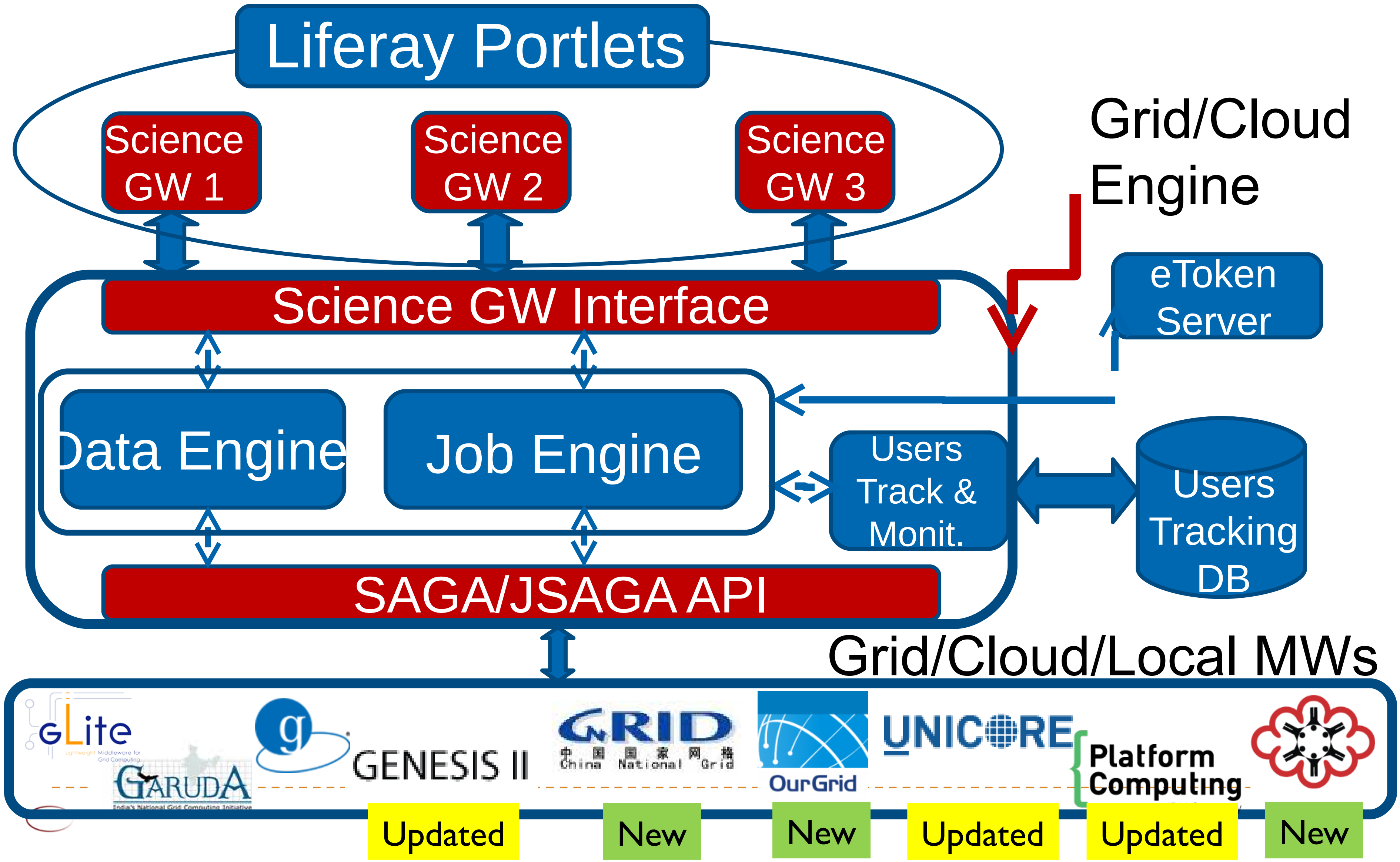
113



Official Identity (Inter-)Federations currently supported by Catania Science Gateways

Other IdPs deployed in Latin America

The Catania Grid & Cloud Engine



Running applications on the EGI Federated Cloud

Co-ordination & Harmonisation of Adv[…]
for Research and Education Data Shar[…]

Science […]

Welcome | Help | Demo Applications | Demo Status

"Hello World"
Computer Science and Mathematics
Life Sciences

By definition, a Scienc[…] […]elopment se[…]
tools, applications, and […] […]al or a suite […]

Cloud Applications

With this service it is possible to execute scientific applications on Virtual Machines (VMs) deployed on […]
Federated Cloud.

The following applications are currently available:

Sequential "Hello World !"
This application just outputs the name of the Virtual Machine where the job has been ex[…]

GNU Octave
GNU Octave is a high-level interpreted language, primarily intended for numerical compu[…] solution of linear and nonlinear problems, and for performing other numerical experiment[…] data visualization and manipulation.

GNU R Environment
R is a free software environment for statistical computing and graphics. R provides a wi[…] classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical techniques, and is highly extensible.

MyWorkspace

My Workspace
- Jobs
- JobsMap
- Data
- Help

Run Scientific Applications on the OCCI-complaint federated clouds

Input Form

1 Portlet Settings

The current portlet has been configured to access

☑ The EGI Federated Cloud Testbed

2 The Cloud Testbed

See the Cloud Resource(s) where you can run your application

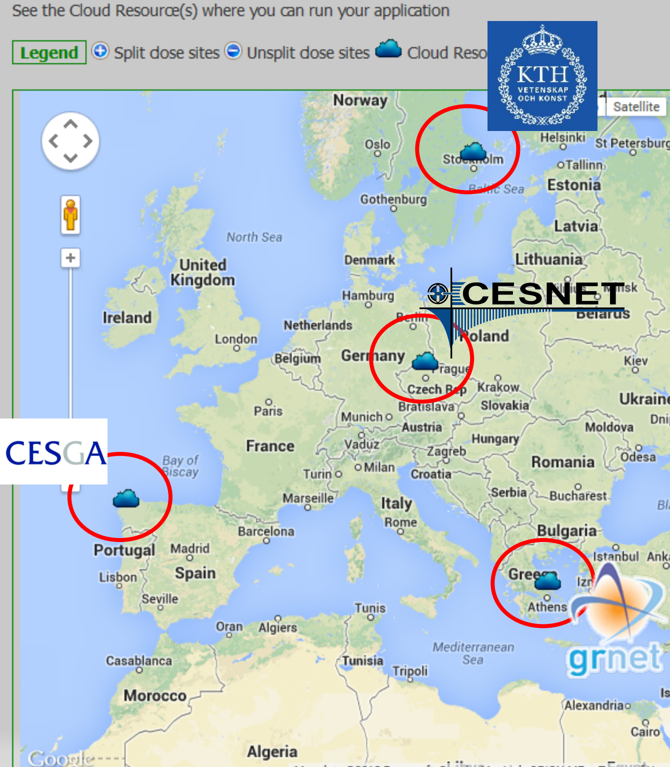
Legend: Split dose sites · Unsplit dose sites · Cloud Reso[…]

Instructions fo[…]
~ With this ser[…] deployed on st[…]

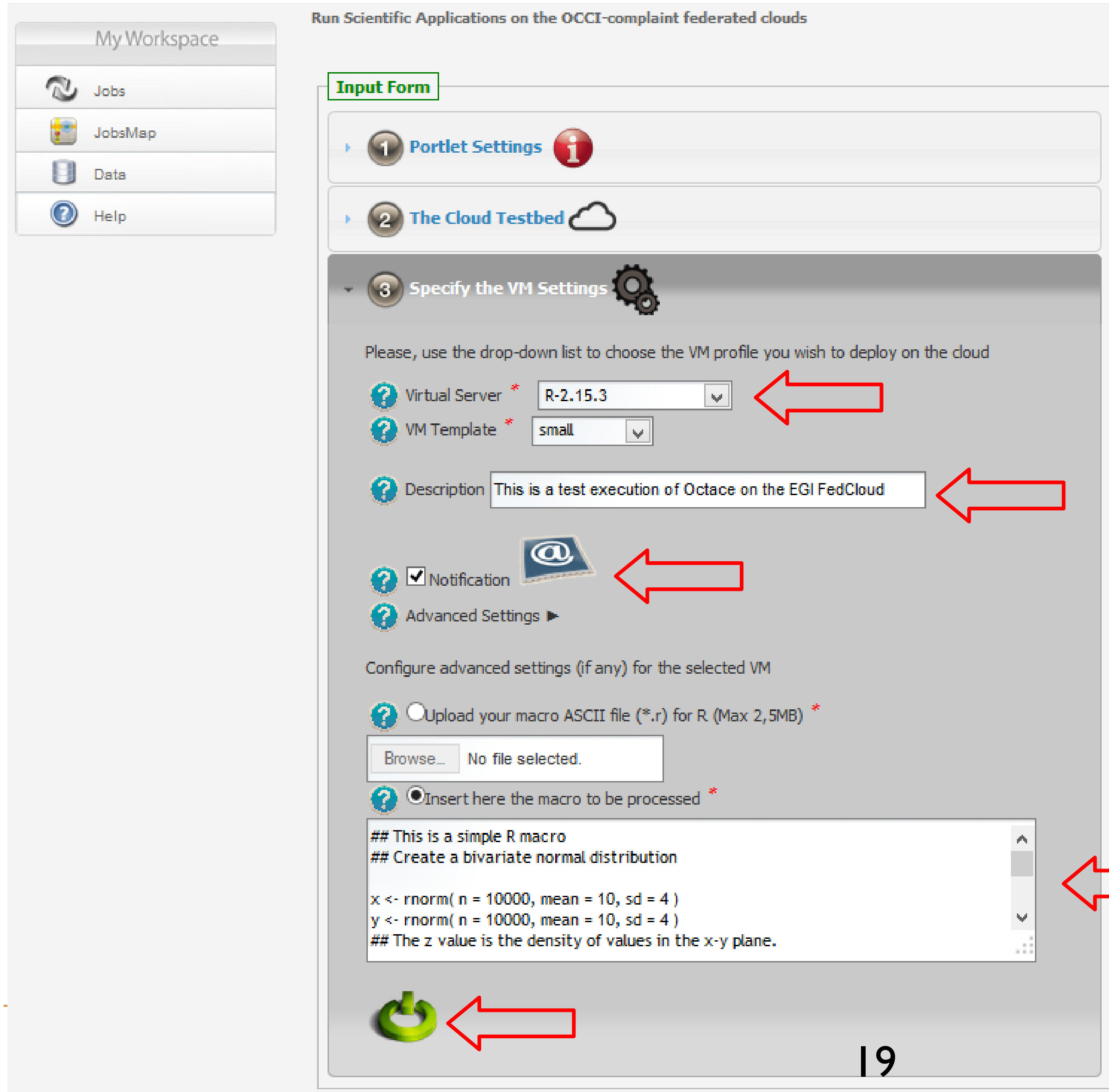
~ The present […]

~ The followin[…]
GNU Octave-3.[…]
GNU Octave i[…] numerical com[…] linear and n[…] exper[…] visual[…]

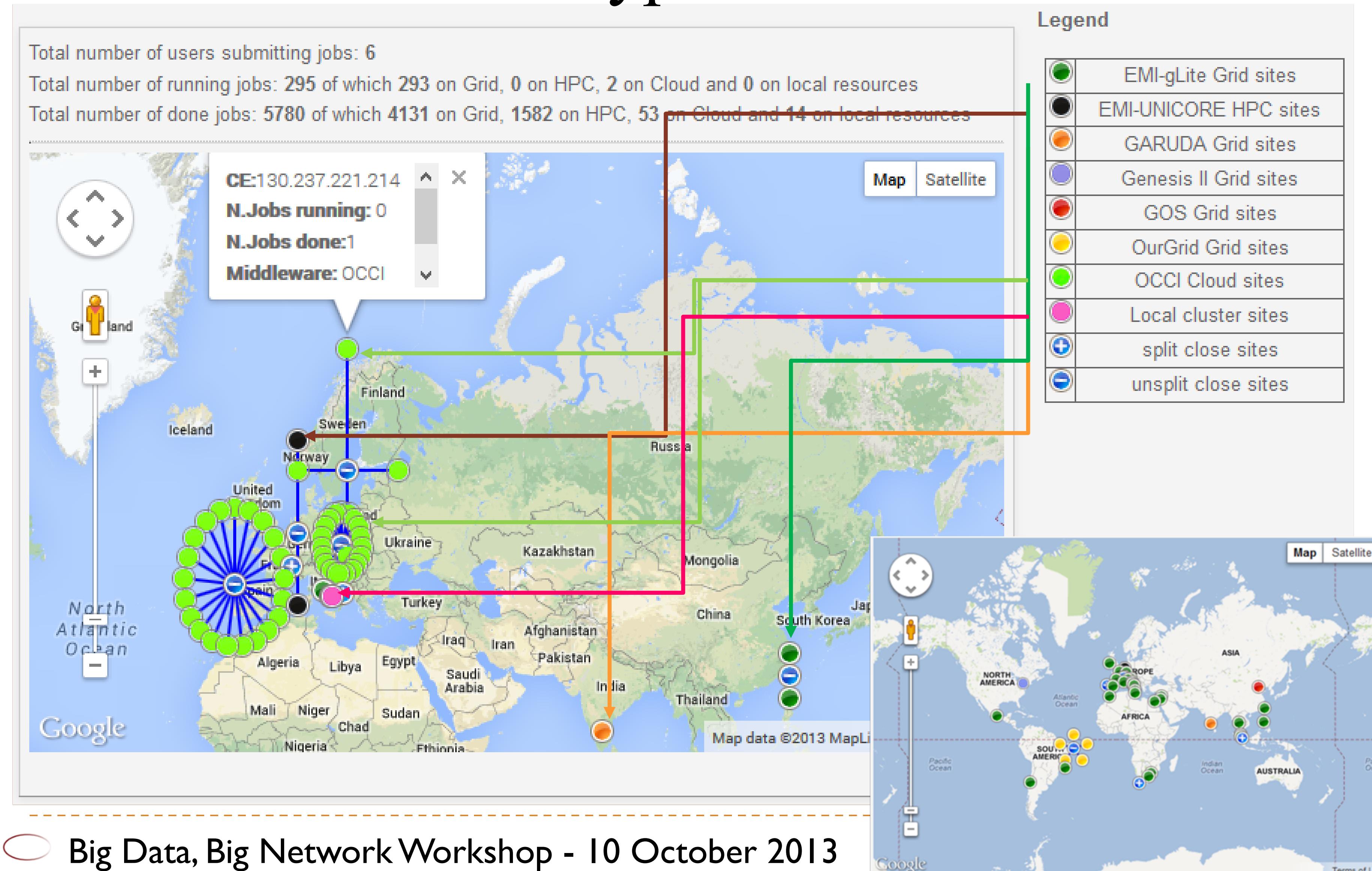
The GNU R-2.1[…]
R is a free sof[…] provides a wid[…] statistical test[…] graphical techn[…]



Running applications on the EGI Federated Cloud



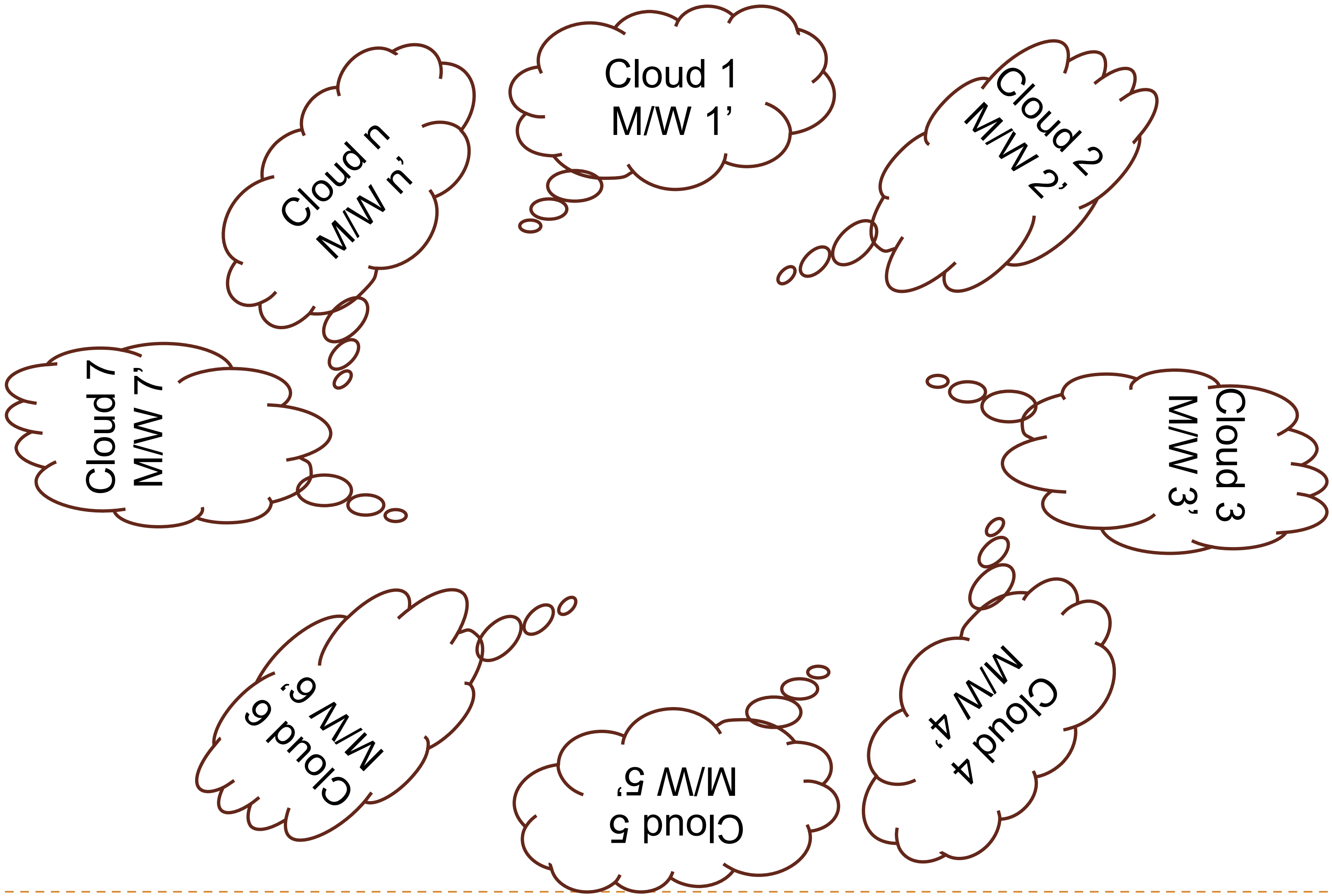
Running applications on various types of e-Infrastructures



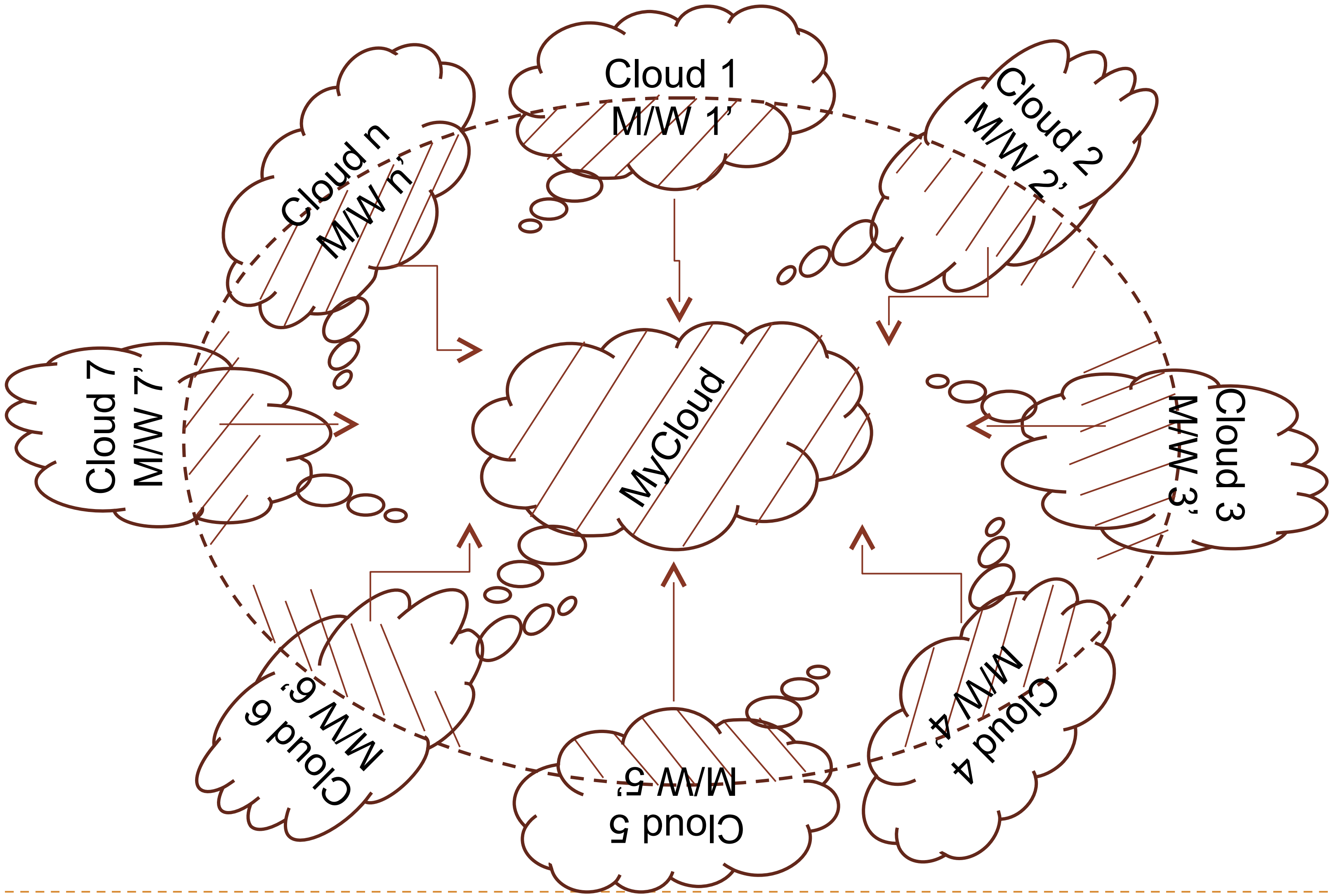
Use case #2 (cloud tenant)

- The cloud tenant of a real or virtual organisation can sign in on a Science Gateway using his/her federated credentials, select virtual machine(s) from a geographically shared repository and deploy/move/copy it/them across his/her personal cloud
  - The graphic user interface will be very intuitive including point & click and drag & drop functionalities
  - The virtual machine(s) will belong to the same domain name (chain-project.eu in the particular case) independently of the site where it/they will be instantiated and of the underlying Cloud middleware stack

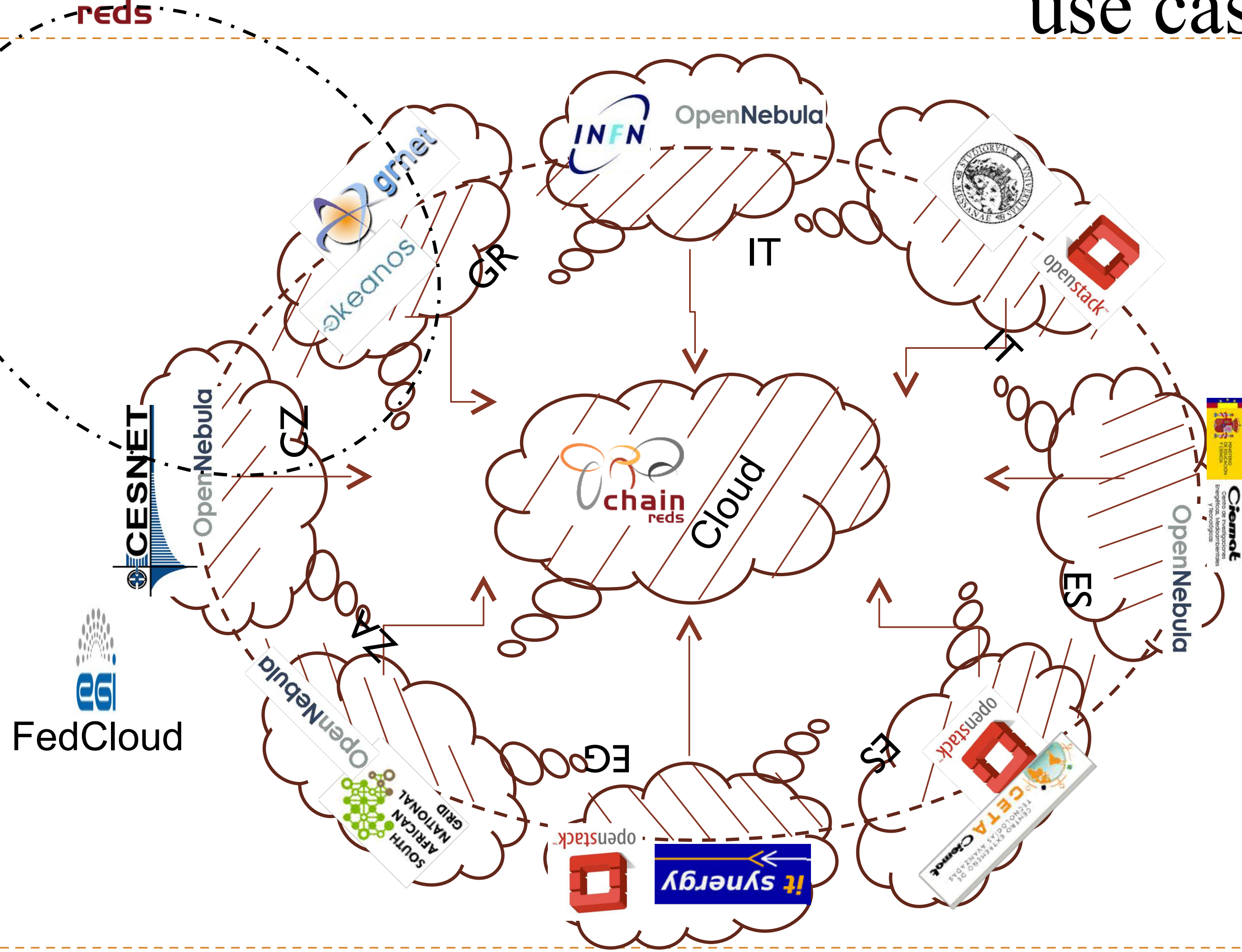
Scenario of use case #2



Scenario of use case #2



Actual testbed configuration for use case #2



8 clouds
6 countries
3 m/w stacks
1 SME

MyCloud @ the CHAIN-REDS Science Gateway

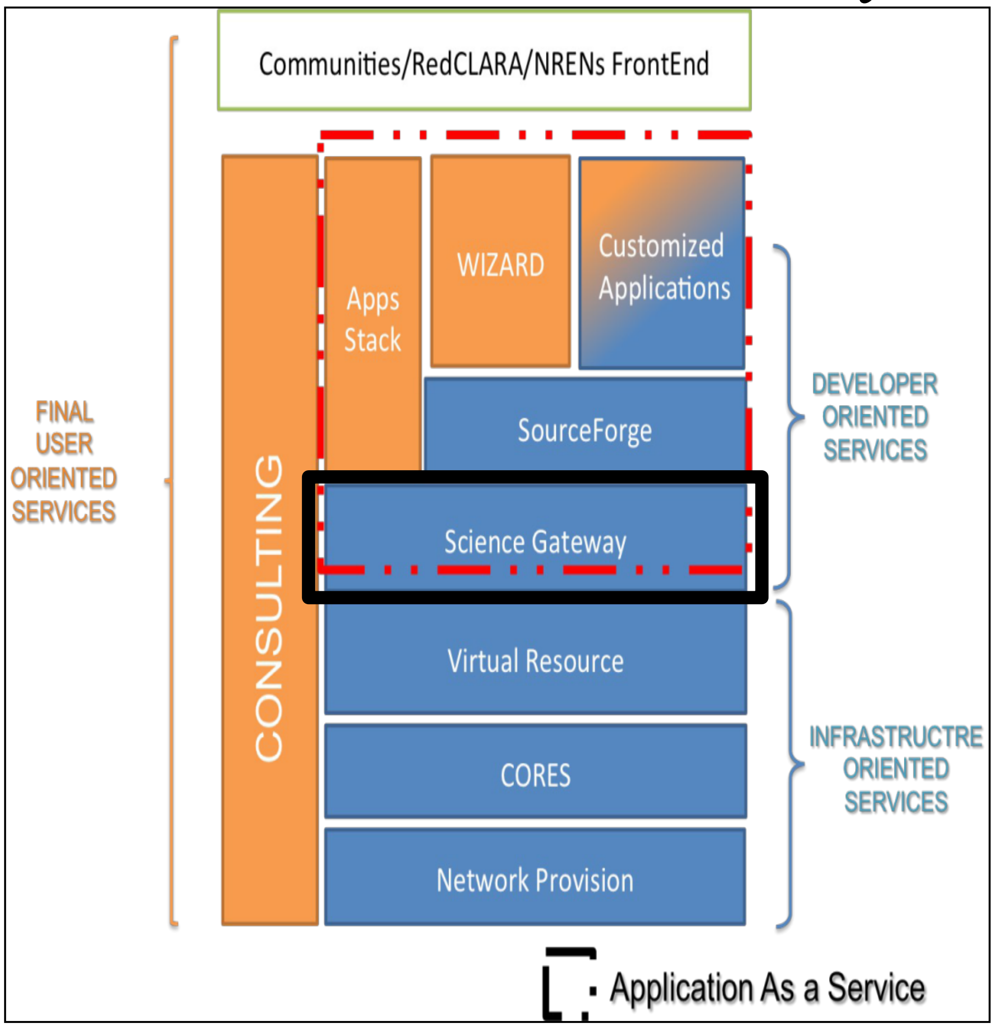
Current functionalities:

- Federated authentication
- Fine-grained authorisation
- Single/multi-deployment of VMs on a cloud and across clouds
- Single/multi-move of VMs across clouds
- Single/multi-deletion of VMs on a cloud and across clouds
- SSH connection to VMs
- Direct web access to VMs hosting web services

Science Gateways in Latin America: the GISELA Science Gateway

.it)

El Servicio de Computación Avanz América Latina y el Caribe



- As part of the SCALAC model, the GISELA SG has been re-installed and updated at UNAM (Mexico) and named SCALAC SG
- The SCALAC SG is in progress to be relocated at: https://sgw.scalac.redclara.net
- As part of the SCALAC commitments, CUDI (the Mexican NREN) is financing two pilot projects for new services: access to HPC and to Cloud infrastructures via the SCALAC Science Gateway
- The SCALAC Science Gateway will be operational by the end of 2013
- The SCALAC SG will be connected to the “Identity Provider of the Mexican Advanced Computing Services for e-Science” and to all other IdPs being established in Latin America in the context of the ELCIRA project

https://comunidades.redclara.net/wiki/scalac/index.php/SCALAC

Summary and outlook

- Standard-based interoperability should be enabled not only across middleware but also and more importantly across computing paradigms (Grid, HPC, Cloud, local clusters, desktops, etc.) in order to exploit big networks as much as possible
- Catania Science Gateways successfully bridge e-Infrastructures developed according to different models architectures and make them interoperable at user application level thanks to standard adoption (SAGA, SAML, OCCI, CDMI, JSR286, etc.)
- The MyCloud service allows seamless multi-cloud service operation across different OCCI-compliant middleware stacks on many sites worldwide
- Next steps are:
  - Creation of the shared storage infrastructure to support stateful VMs
  - Allow deployed VMs to «find themselves» in MyCloud
  - Fostering the deployment of cloud infrastructures in the regions addressed by CHAIN-REDS to widen the testbed both in size and geographic coverage
  - Promotion of the EGI FedCloud model and possible extension of its infrastructure to other regions in order to support global VRCs

Co-ordination & Harmonisation of Advanced e-Infrastructures for Research and Education Data Sharing

Questions ?

Research Infrastructures – Proposal n. 306819

Tom DeFanti* and Larry Smarr **

*Research Scientist, Qualcomm Institute,
California Institute for Telecommunications and Information Technology, UCSD
Distinguished Professor Emeritus of Computer Science, University of Illinois at Chicago
** Director, California Institute for Telecommunications and Information Technology
Harry E. Gruber Professor,
Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD

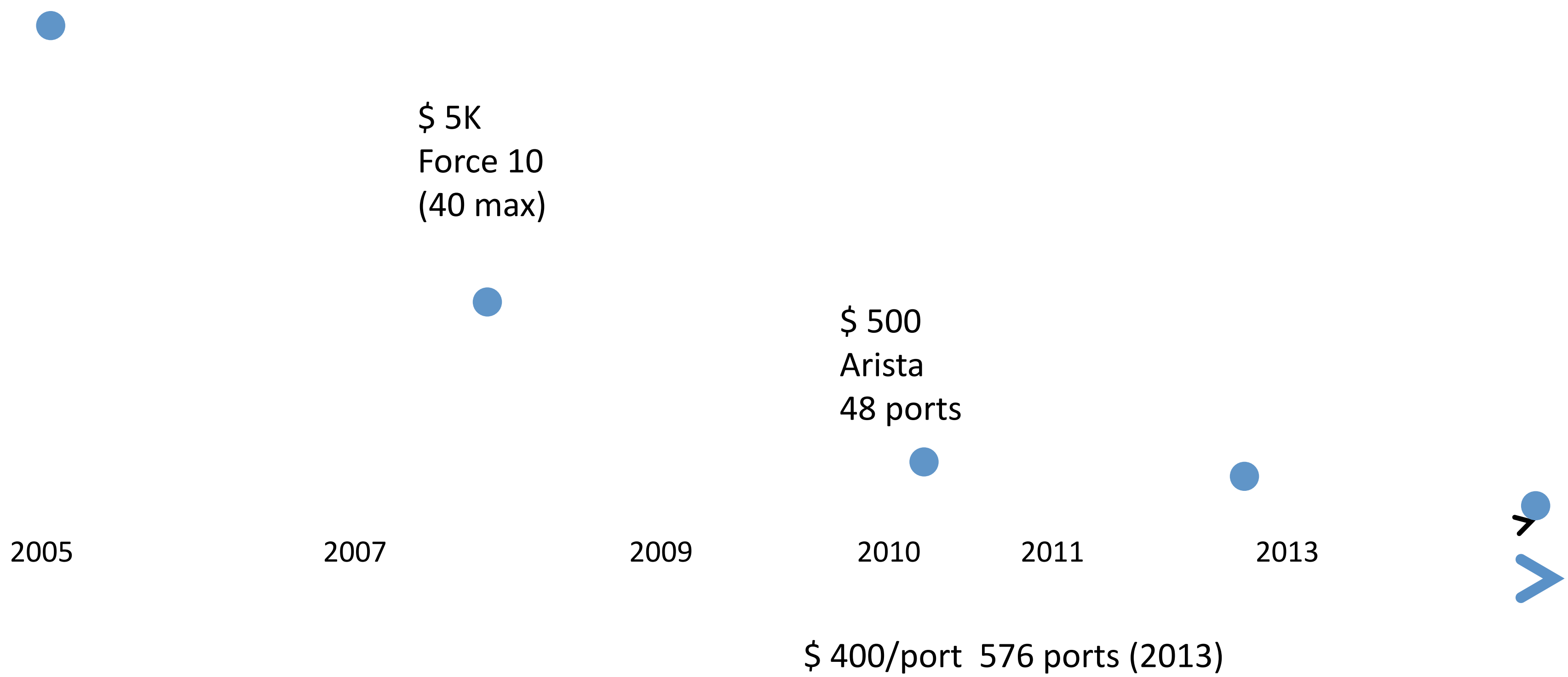
Big Displays For Big Data Served by 40G and 100G Networks

Great Displays For Great Data Served by Great Networks

Rapid Evolution of 10Gb Port Prices Makes Campus-Scale 10Gbps CI Affordable

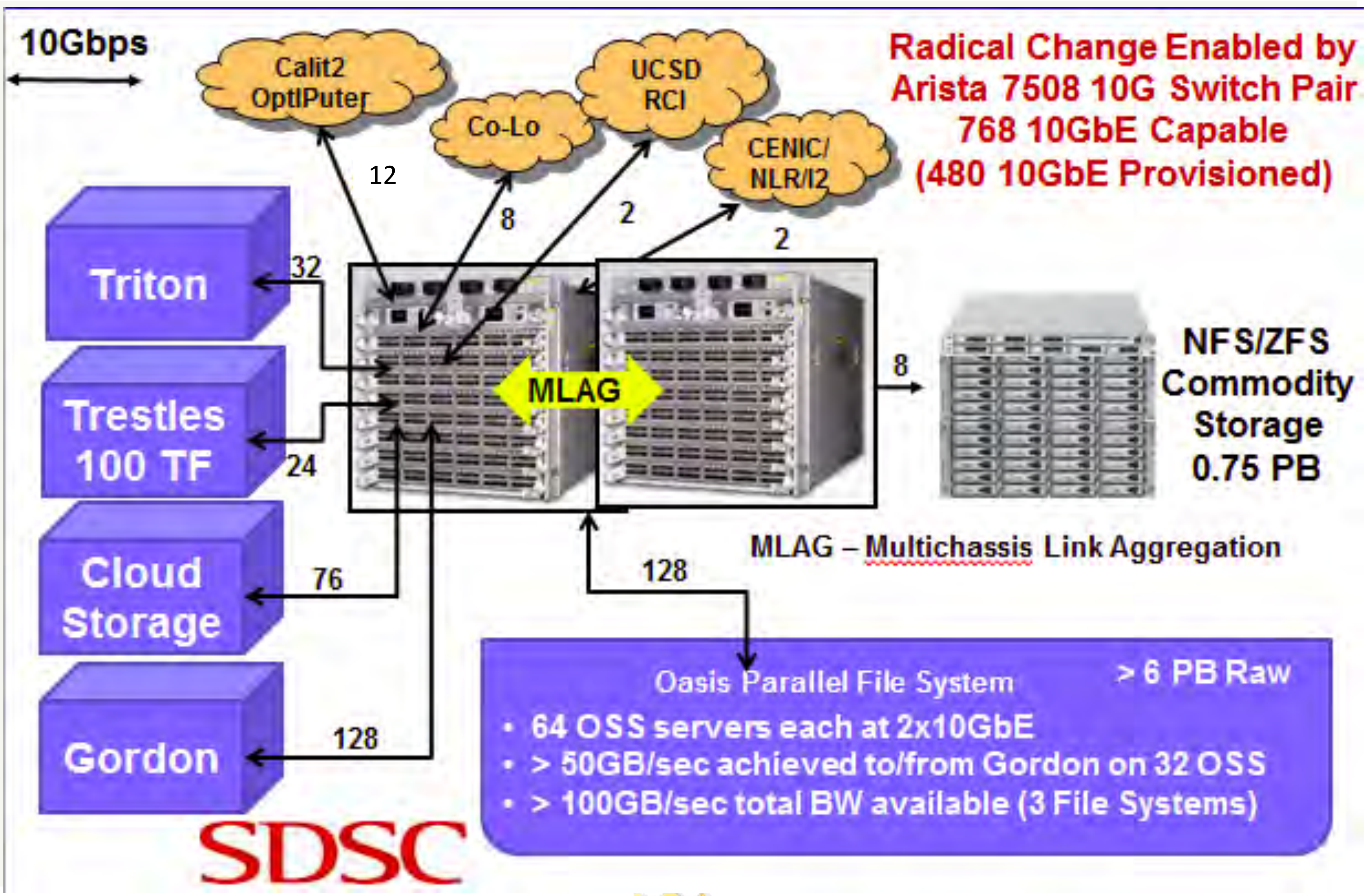
- Port Pricing is Falling
- Density is Rising – Dramatically
- Cost of 10Gb NICs same as graphics cards!

$80K/port
Chiaro
(60 Max)



Source: Philip Papadopoulos, SDSC/Calit2

Arista Switch Enables SDSC's Massively Parallel 10G Switched Data Analysis Resource



Gordon Bests Previous Mega I/O per Second by 25x

SDSC

NSF

Many Disciplines Beginning to Need Dedicated High Bandwidth on Campus

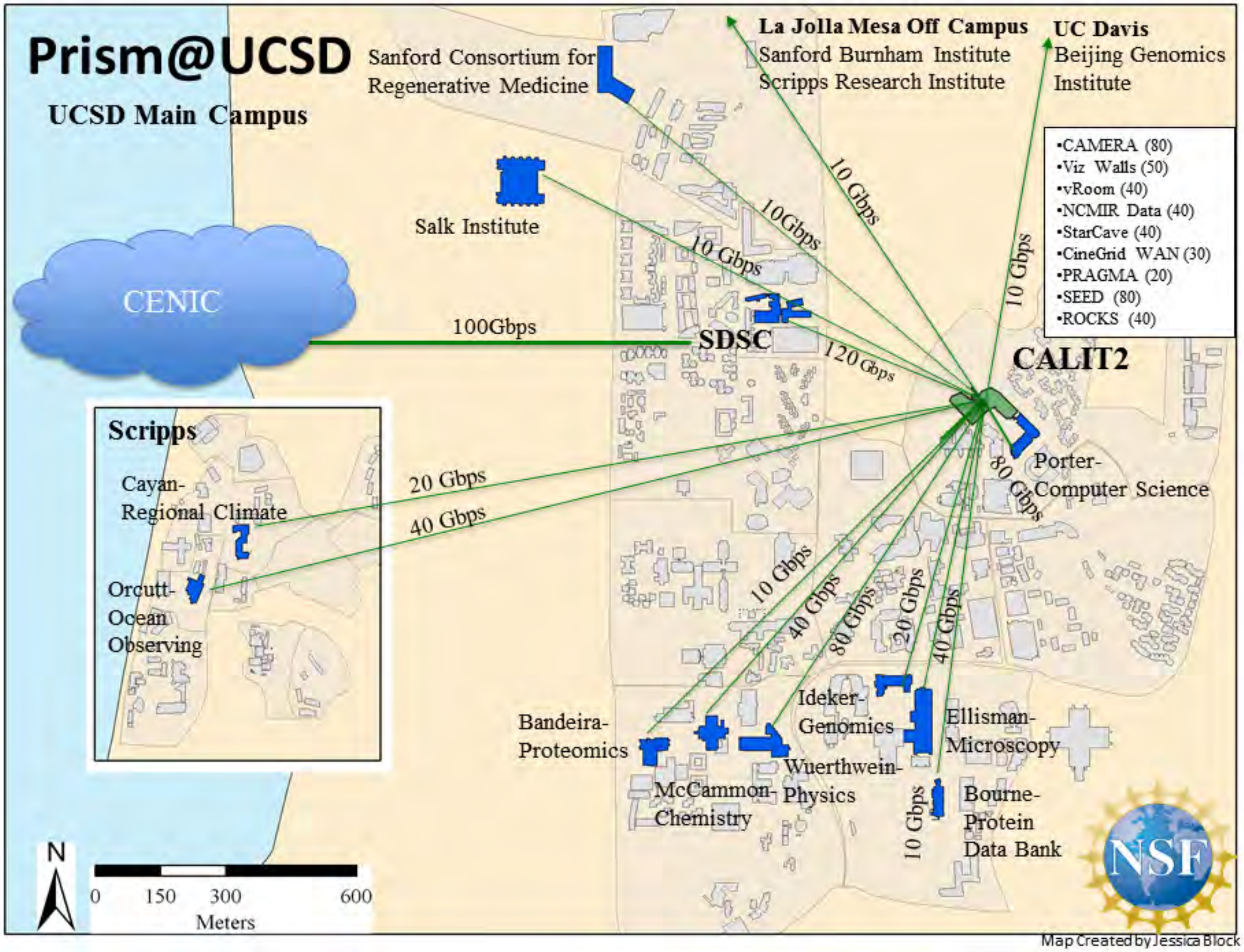
How to Terminate a CENIC 100G Campus Connection

- Remote Analysis of Large Data Sets
  - Particle Physics
- Connection to Remote Campus Compute & Storage Clusters
  - Ocean Observatory
  - Microscopy and Next Gen Sequencers
- Providing Remote Access to Campus Data Repositories
  - Protein Data Bank and Mass Spectrometry
- Enabling Remote Collaborations
  - National and International

SDSC

Creating a "Great Data Freeway" System Connecting Instruments, Computers, & Storage

Phil Papadopoulos, PI
Larry Smarr co-PI

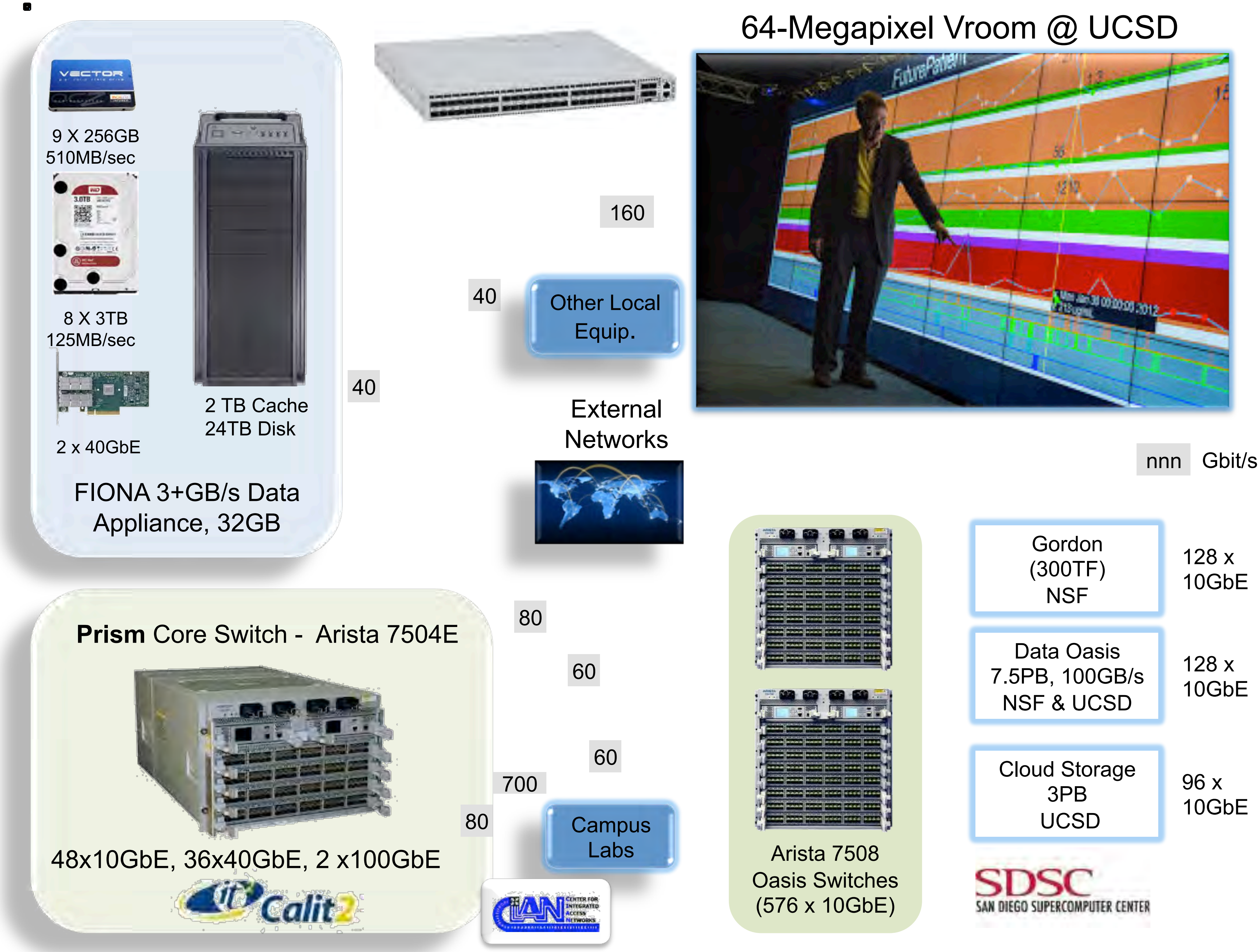


PRISM @UCSD

Start Date 1/1/13

Great Data to Photons Challenges

- Optimizing transfer via 40G, 100G and rdma
  - FIONA disks/flash in Switch Colo to head nodes
  - Or from FIONA Grande flash memory to GPUs
- GPUs synchronizing screens for 16-64 megapixel movies at 30/60 fps (a real 100G problem)
- Understanding resulting bottlenecks as they shift
- Extending this capability nationwide and worldwide
- FIONA is our interim impedance matcher and test system (see next slides)



64-Megapixel Vroom @ UCSD
9 X 256GB
510MB/sec
8 X 3TB
125MB/sec
2 x 40GbE
2 TB Cache
24TB Disk
FIONA 3+GB/s Data Appliance, 32GB
40
160
40
Other Local Equip.
External Networks
nnn Gbit/s
Prism Core Switch - Arista 7504E
48x10GbE, 36x40GbE, 2 x100GbE
Calit2
80
60
60
700
80
Campus Labs
Arista 7508 Oasis Switches (576 x 10GbE)
Gordon (300TF) NSF
128 x 10GbE
Data Oasis 7.5PB, 100GB/s NSF & UCSD
128 x 10GbE
Cloud Storage 3PB UCSD
96 x 10GbE
SDSC
SAN DIEGO SUPERCOMPUTER CENTER

OptIPortables @ UCSD

18 x 256GB
510MB/sec

16 x 3TB
125MB/sec

4 x 10GbE

4 TB Cache
48 TB Disk
Nvidia GPUs

FIONA Grande
6+GB/s 32GB

4K/8K/16K Video

80

Other Local Equip.

80

External Networks

nn =Gbit/s

Prism Core Switch - Arista 7504E

48x10GbE, 36x40GbE, 2x100GbE

80

60

60

700

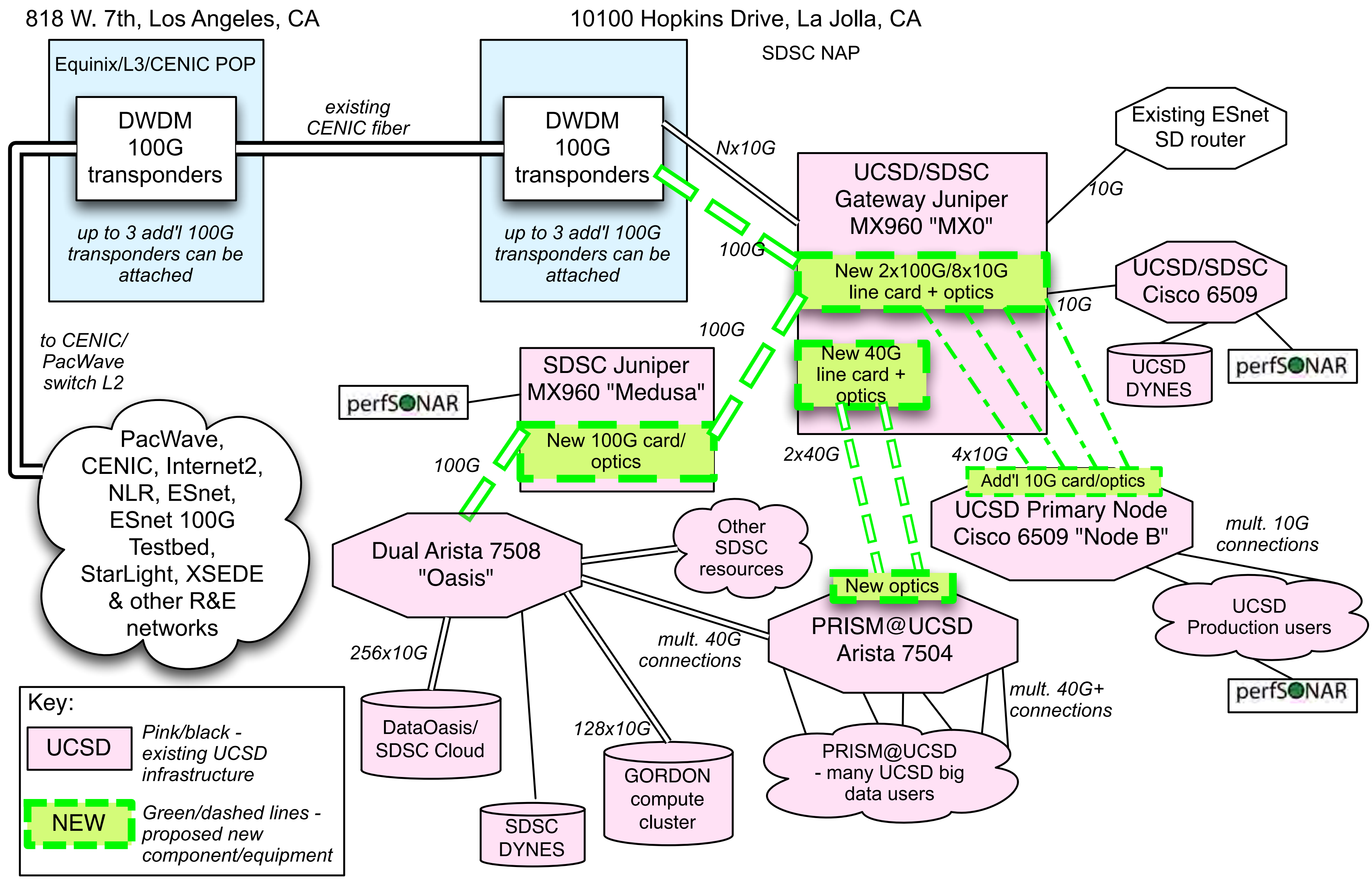
80

Campus Labs

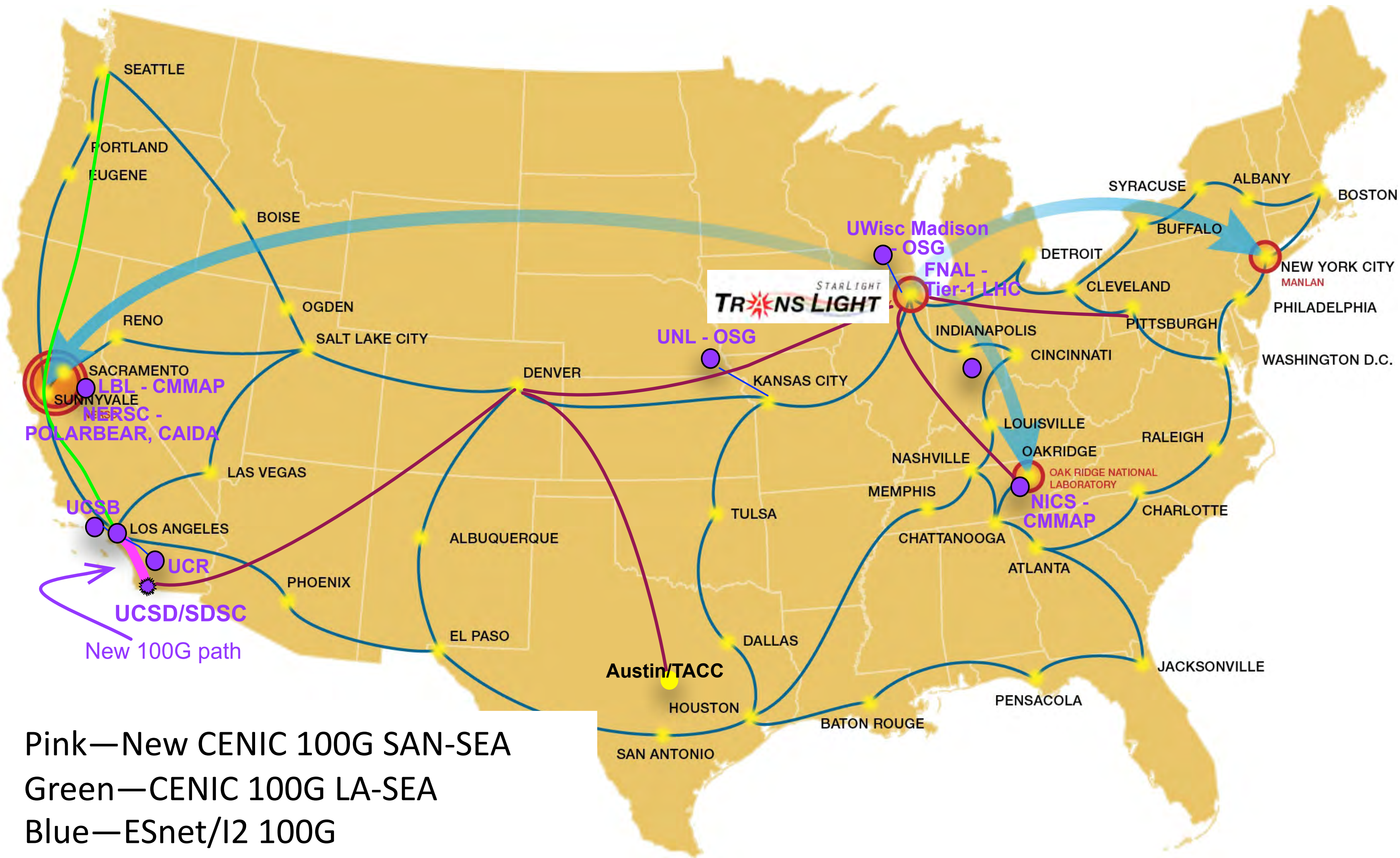
Arista 7508
Oasis Switches
(576 x 10GbE)

Gordon (300TF) NSF	128 x 10GbE
Data Oasis 7.5PB, 100GB/s NSF & UCSD	128 x 10GbE
Cloud Storage 3PB UCSD	96 x 10GbE

SDSC SAN DIEGO SUPERCOMPUTER CENTER



Great Data over Great Networks



Pink—New CENIC 100G SAN-SEA
Green—CENIC 100G LA-SEA
Blue—ESnet/I2 100G
100G to Europe and Shhh!

Great 110” Foldable 4K Stereo on Wheels



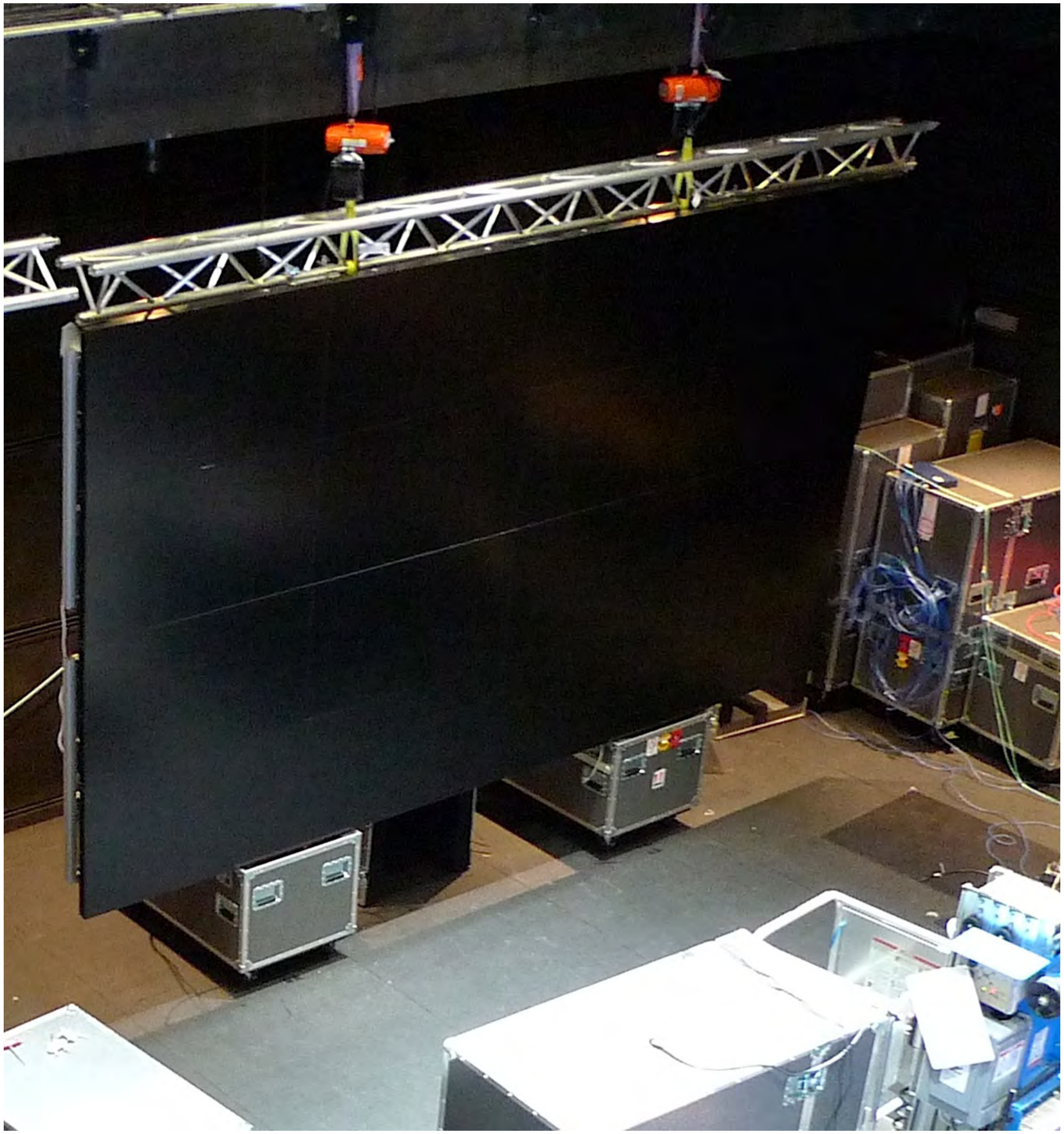


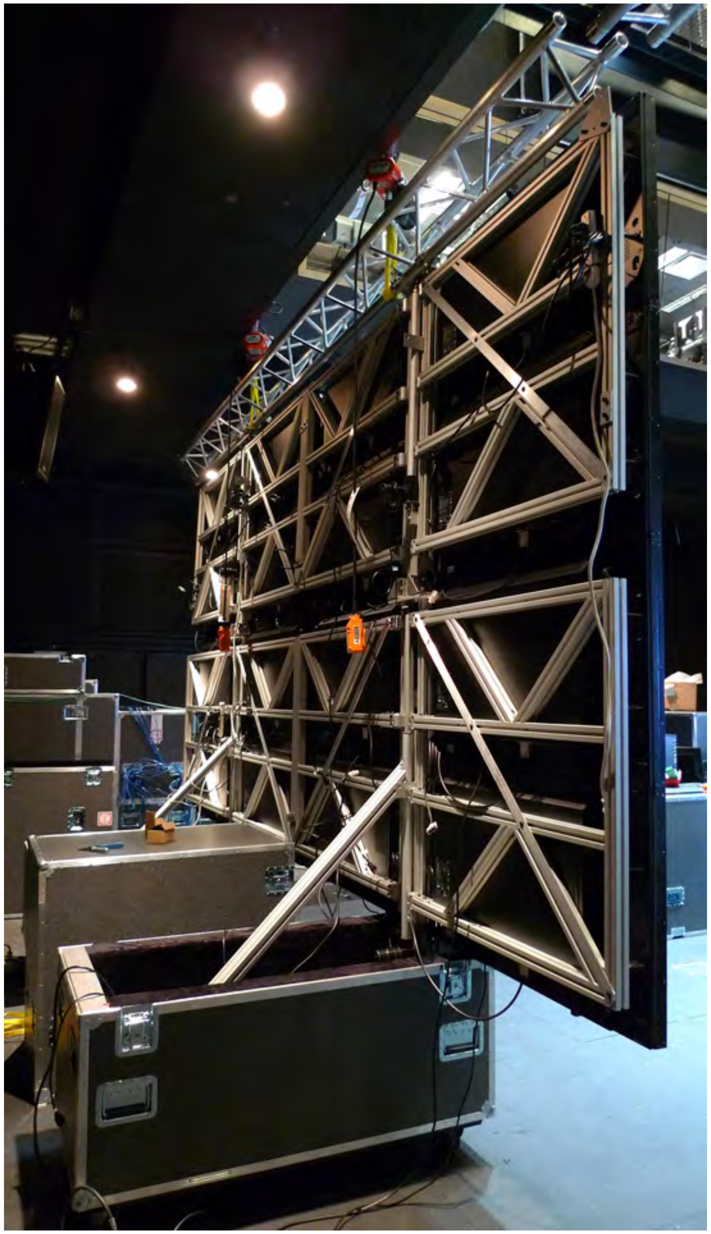

Shipping Great Displays for Collaboration Worldwide

Great Displays can be as Wide as You Want

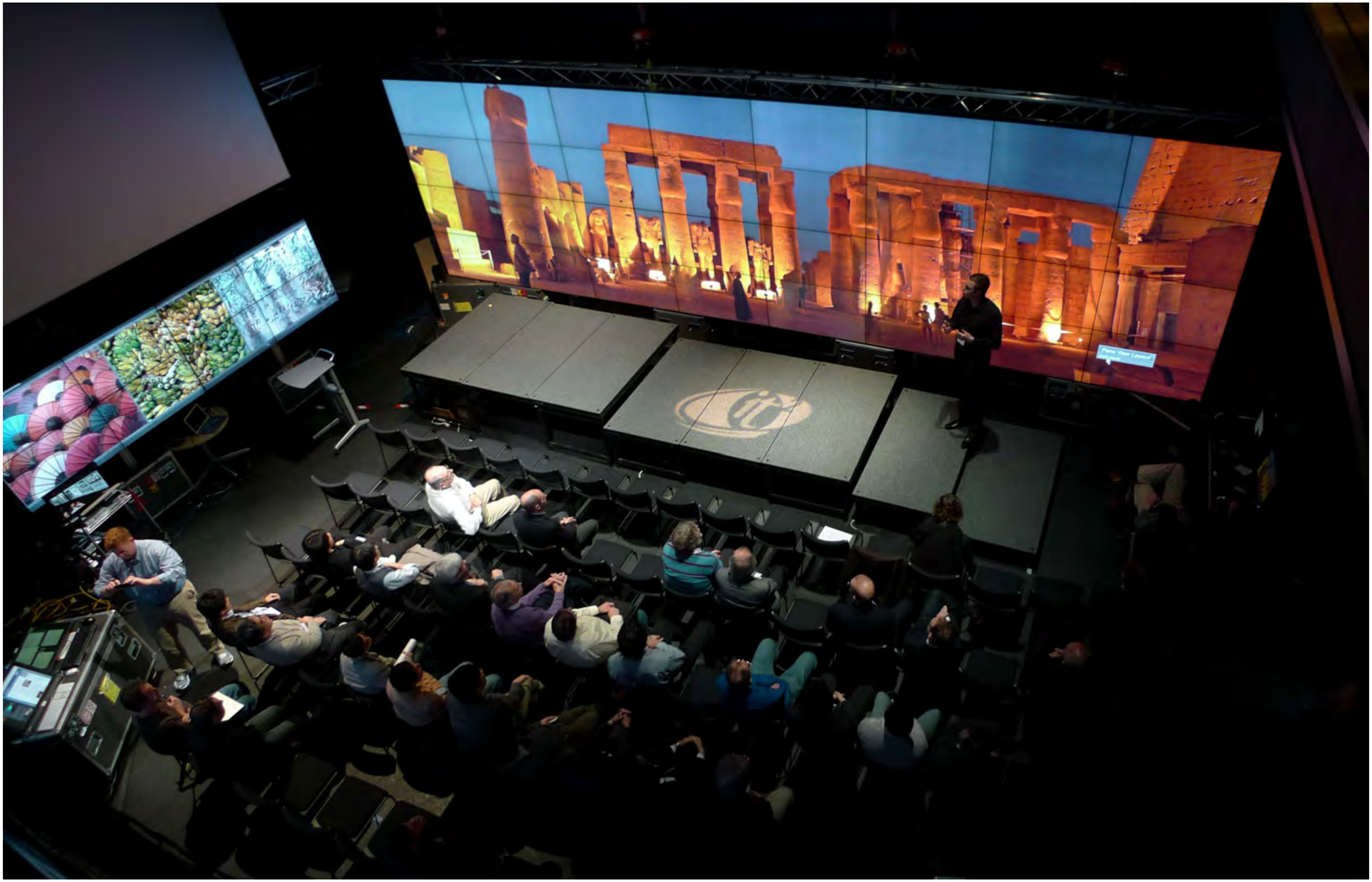
They can be in 2D or 3D, with or without tracking, gesture recognition, and audio

Great Displays Stack



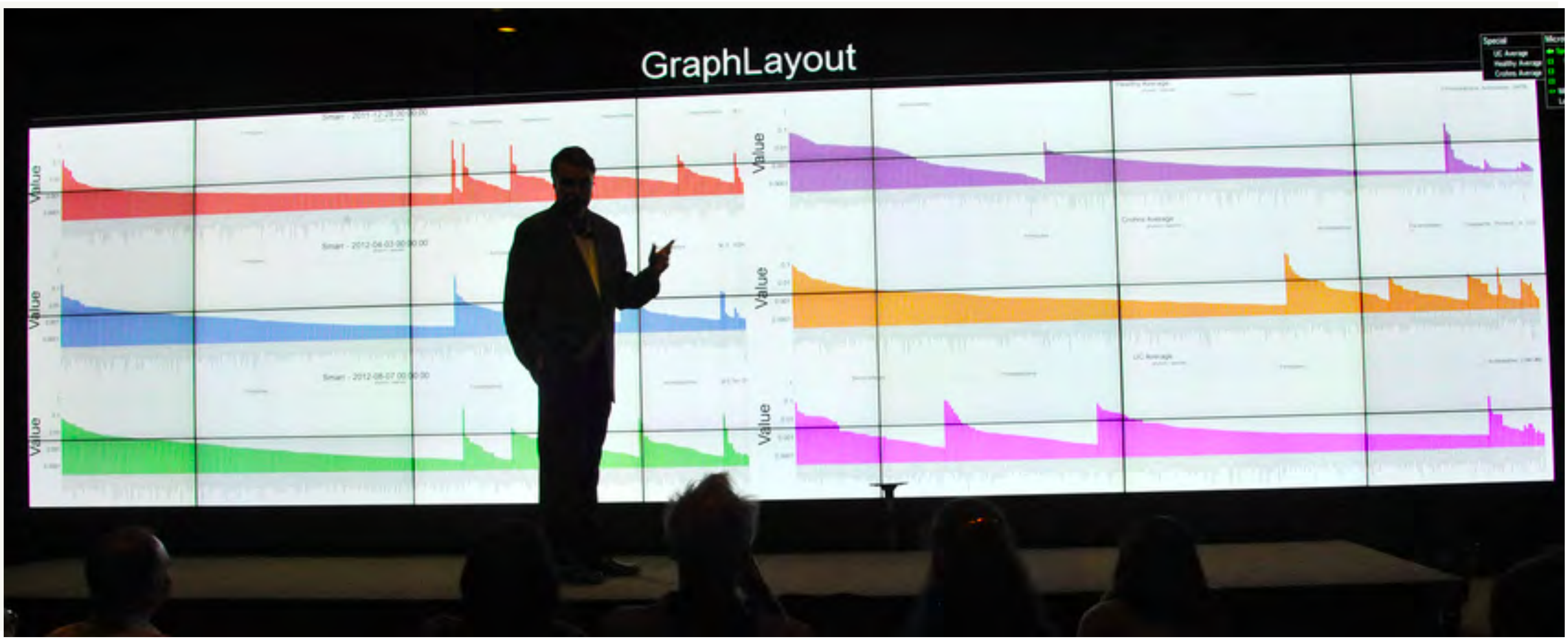


32 Screens Stacked in QI's VRoom



Dynamically Showing Billions of XYZ Lidar Points

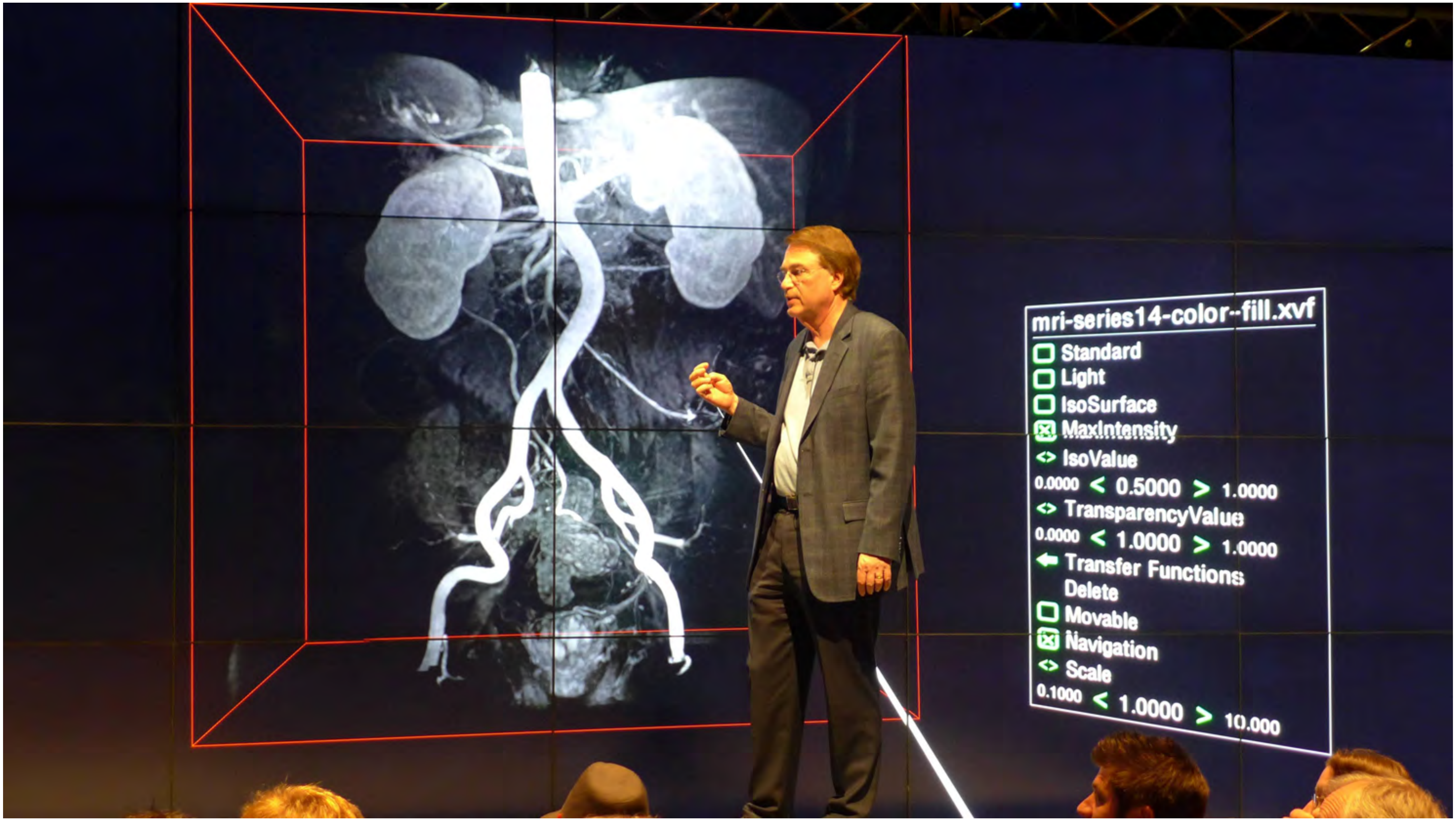
SDSC Gordon Supercomputer Analysis of LS Gut Microbiome Displayed on Calit2 VROOM



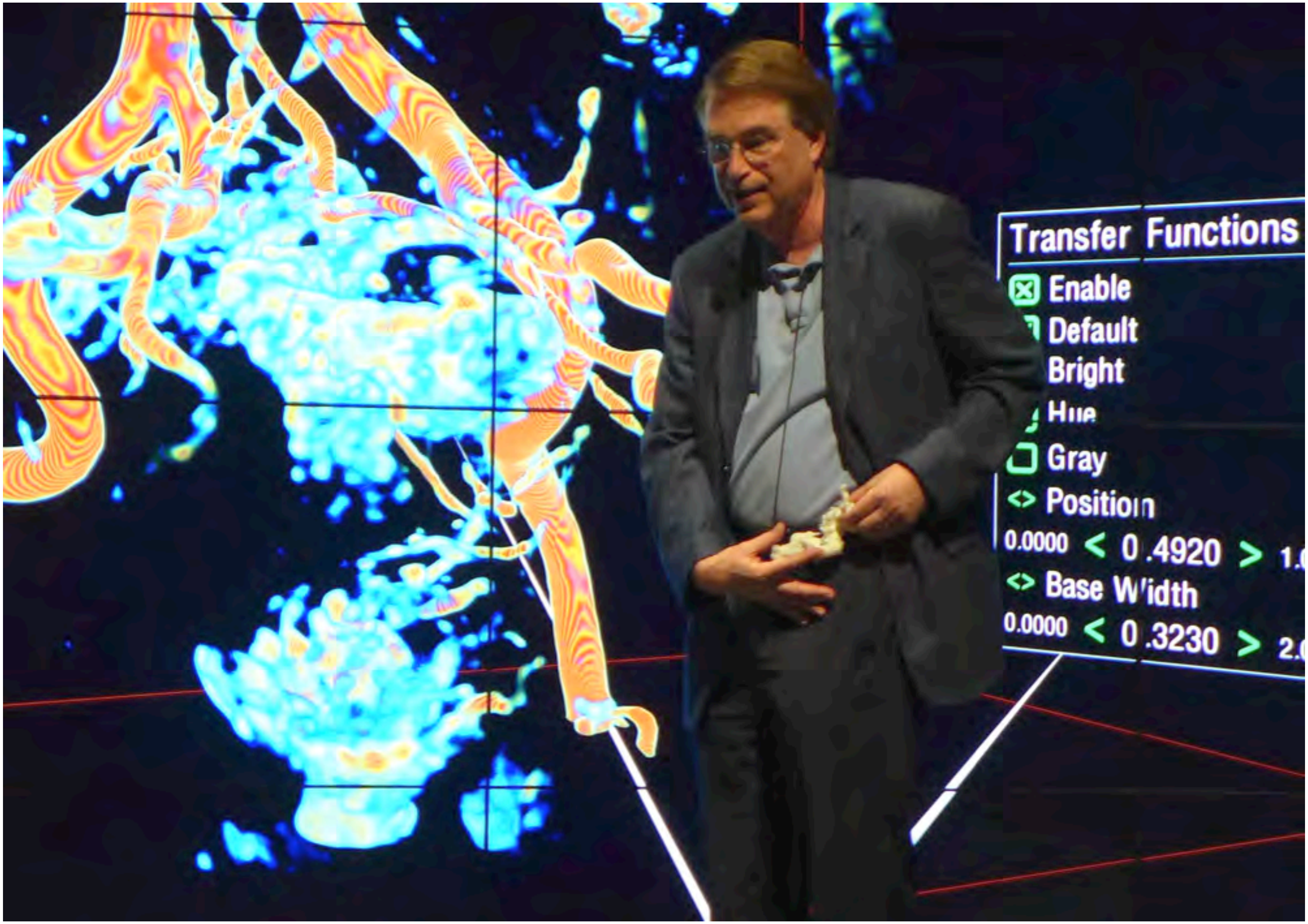
Calit2 VROOM-FuturePatient Expedition

Collaborative Future Patient Analysis

Collaborative Future Patient Analysis



Collaborative Future Patient Analysis



4K Panels next step!

64 Megapixel (32K x 8K) Digital Cinema

Tele-Collaboration for Audio Post-Production
Real-Time Picture & Sound Editing Synchronized Over IP

Skywalker Sound@Marin Calit2@San Diego

Tele-Collaboration

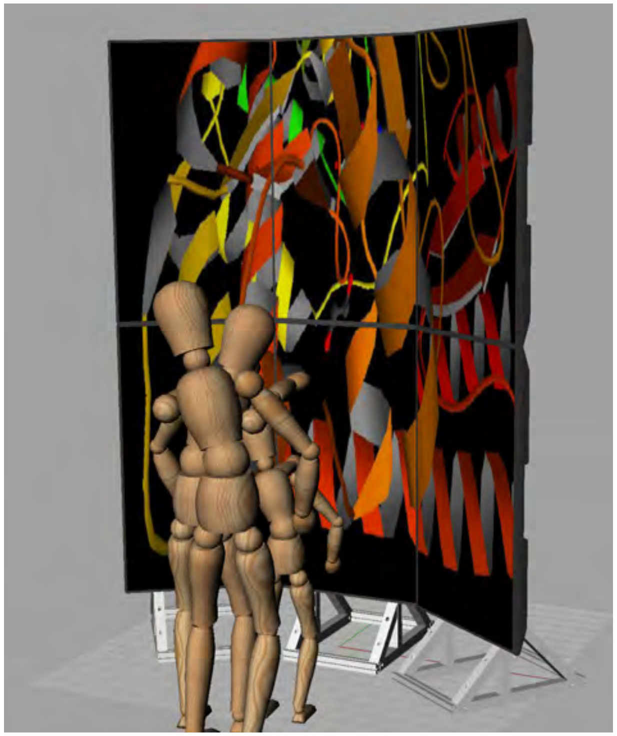
Disney + Skywalker Sound + Digital Domain + Laser Pacific
NTT Labs + UCSD/Calit2 + UIC/EVL + Pacific Interface

Collaboration Between QI's Vroom and EVL's CAVE2

Great Collaboration: UCSD is Linked to CICESE at 10G Coupling OptIPortals via SAGE at Each Site

Our TourCAVE is a Great Portable VR Display designed for Medical Visualization and Cultural Heritage Collaboration

TourCAVE with 65” LG Panels

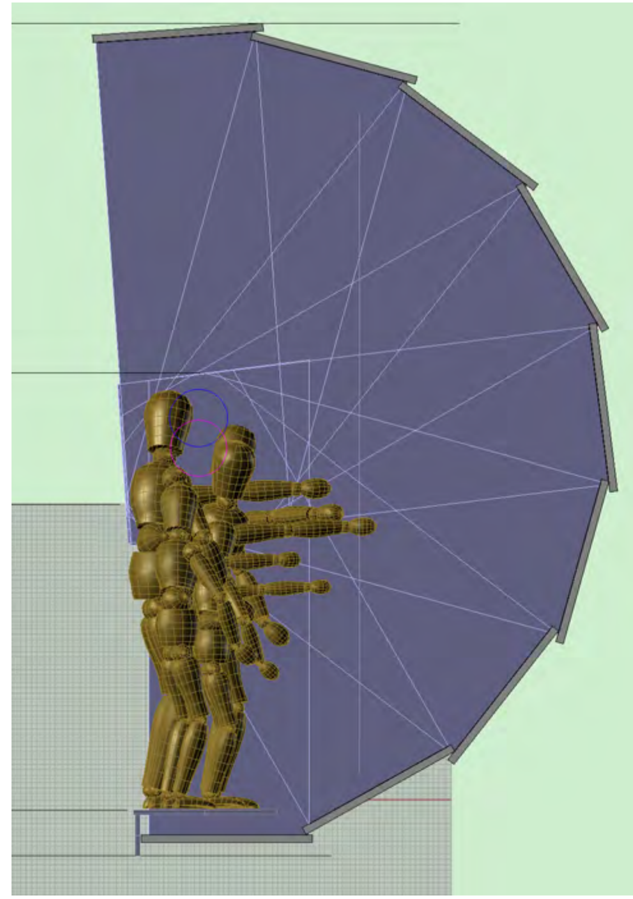


TourCAVE with 55” LG Panels

CAD Design and Rendering by Greg Dawe

3D VR: 28 Megapixel TourCAVE

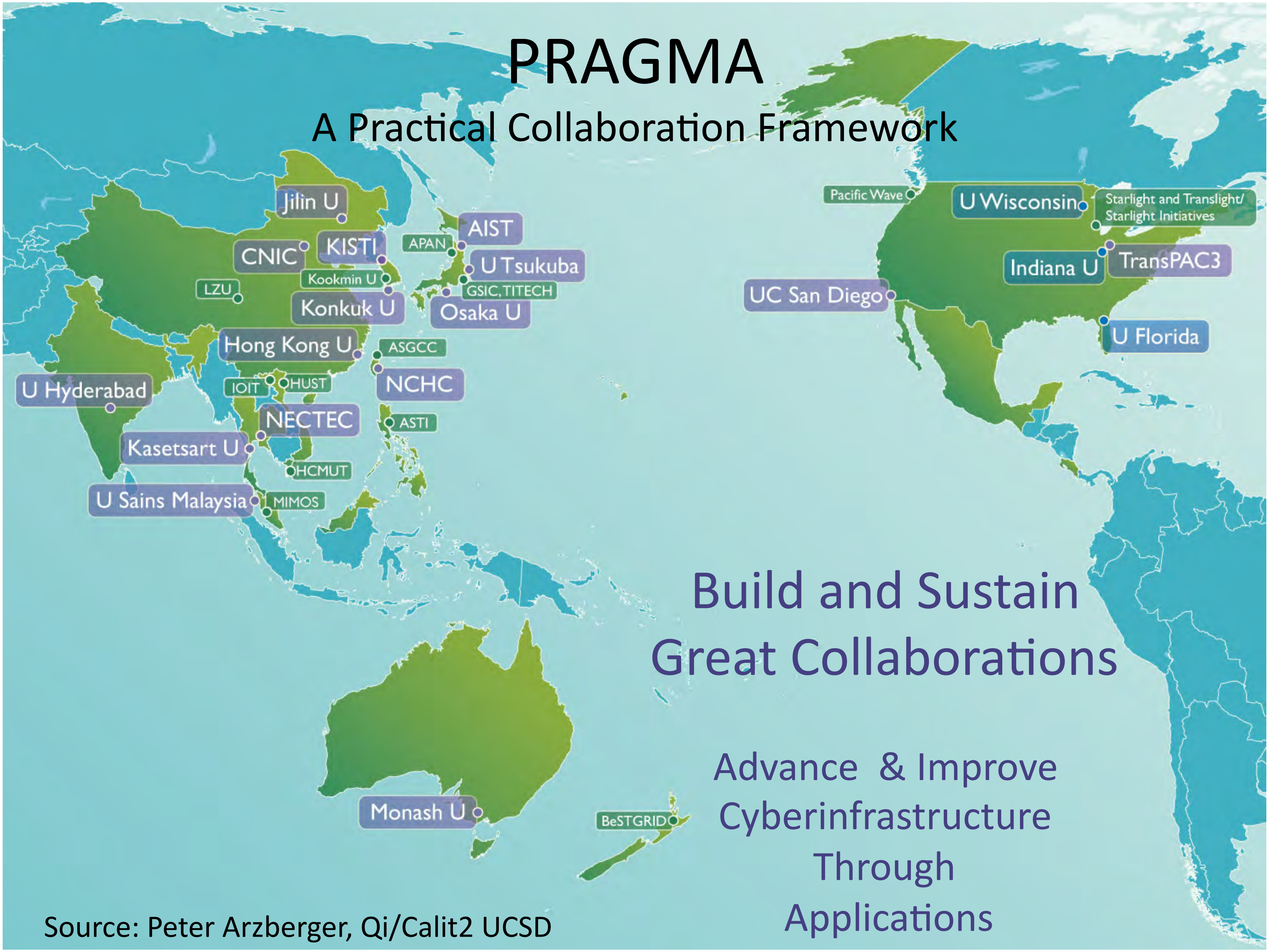
The Great New WAVE—70 Megapixels



WAVE VR display 7 high by 5 wide HD panels built for the UCSD SME Building, UCSD, October2013

CAD Design by Greg Dawe, as Always

70 Megapixel 3D—The WAVE


PRAGMA
A Practical Collaboration Framework
Jilin U
CNIC
KISTI
AIST
APAN
U Tsukuba
Kookmin U
GSIC, TITECH
LZU
Konkuk U
Osaka U
Hong Kong U
ASGCC
NCHC
U Hyderabad
IOIT
HUST
NECTEC
ASTI
Kasetsart U
HCMUT
U Sains Malaysia
MIMOS
Monash U
BeSTGRID
Pacific Wave
U Wisconsin
Starlight and Translight/
Starlight Initiatives
Indiana U
TransPAC3
UC San Diego
U Florida
Build and Sustain
Great Collaborations
Advance & Improve
Cyberinfrastructure
Through
Applications
Source: Peter Arzberger, Qi/Calit2 UCSD

Our Great Displays for Great Data over Great Networks planning, research, and education efforts are made possible by funding from:

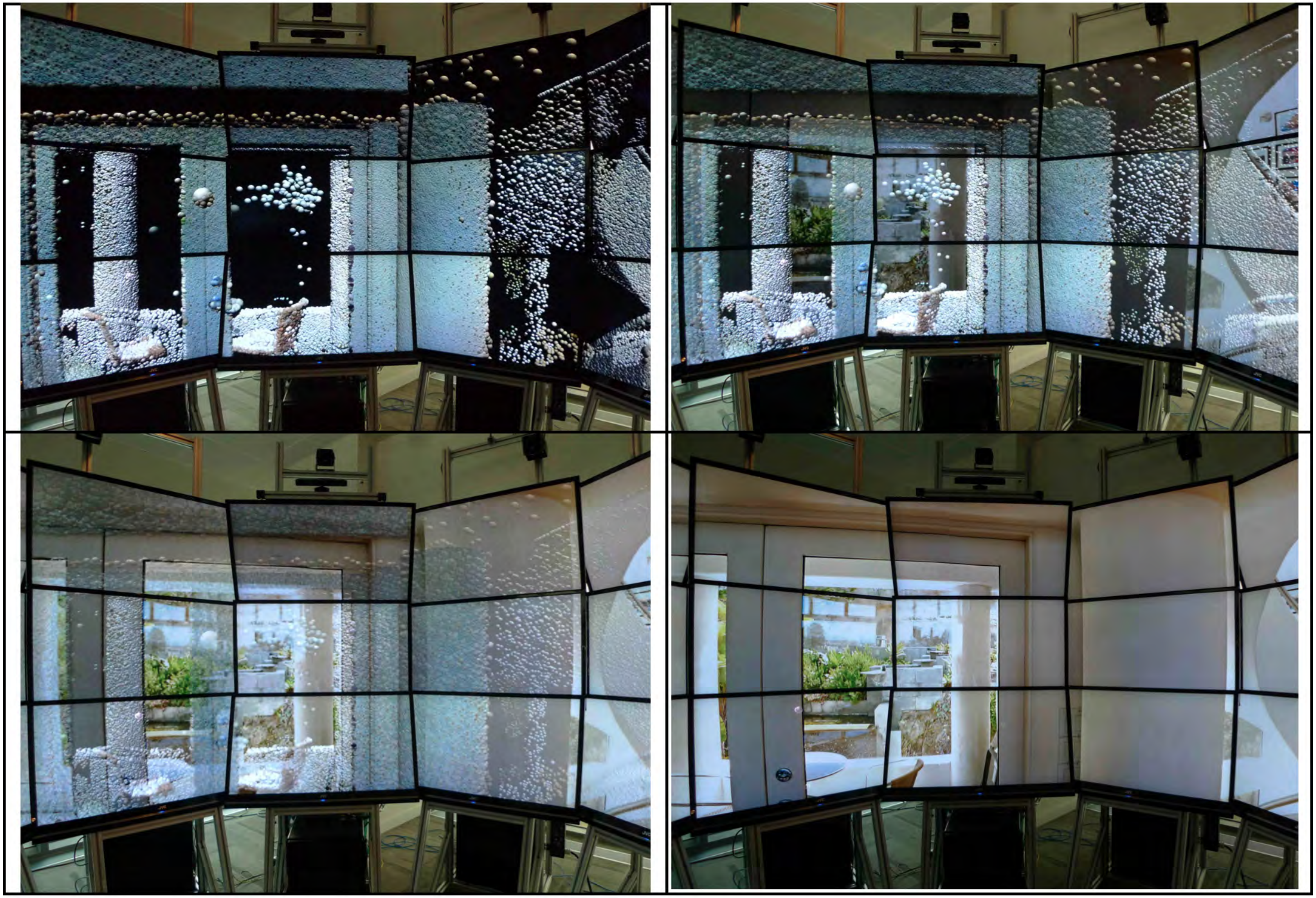
- US National Science Foundation (NSF) awards ANI-0225642, EIA-0115809, SCI-0441094, and CNS 0821155
- Calit2 Director's Office
- Calit2 UCSD Division (Qualcomm Institute)
- King Abdullah University of Science and Technology (KAUST)
- King Abdulaziz City for Science and Technology (KACST)
- University of Illinois at Chicago, Argonne National Laboratory, and Northwestern University for StarLight networking and management
- Pacific Wave and CENIC
- NTT Network Innovations Lab
- Cisco Systems, Inc.
- Pacific Interface, Inc.
- Many corporate donors

The CAVEcam: 3D Stereo Panoramic Photography—Adding Photorealism to VR

With the CAVEcam, in addition to photographically capturing an image that says:
"See what I have seen",
we can now also capture a complete 3D photo environment, recreate it in a VR setting, and claim:
"Be where I have been."

Source: Dick Ainsworth

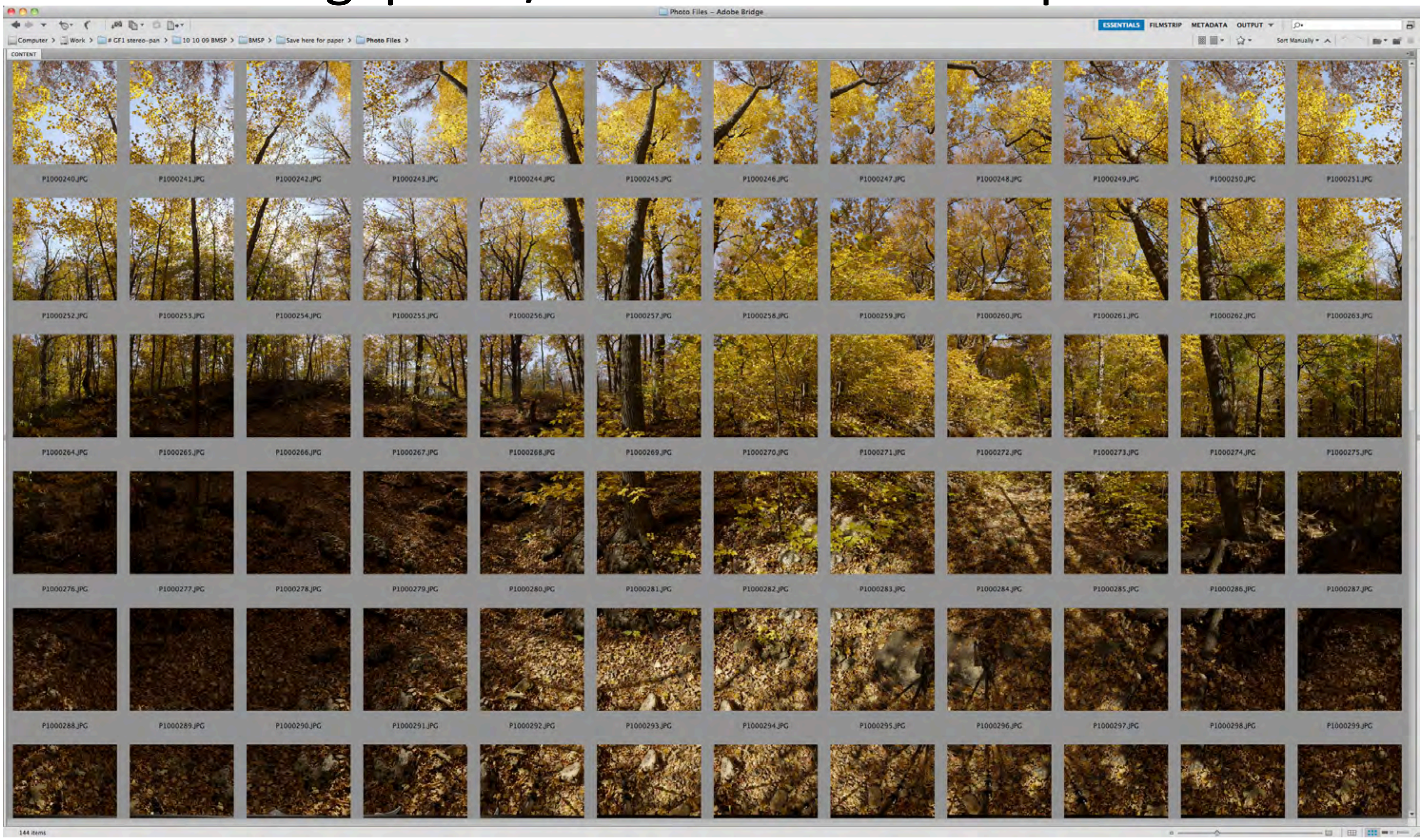

CAVEcam April 29, 2011
Photo: Greg Wickham



Transition between Faro Lidar Scan and CAVEcam of the same 3D scene

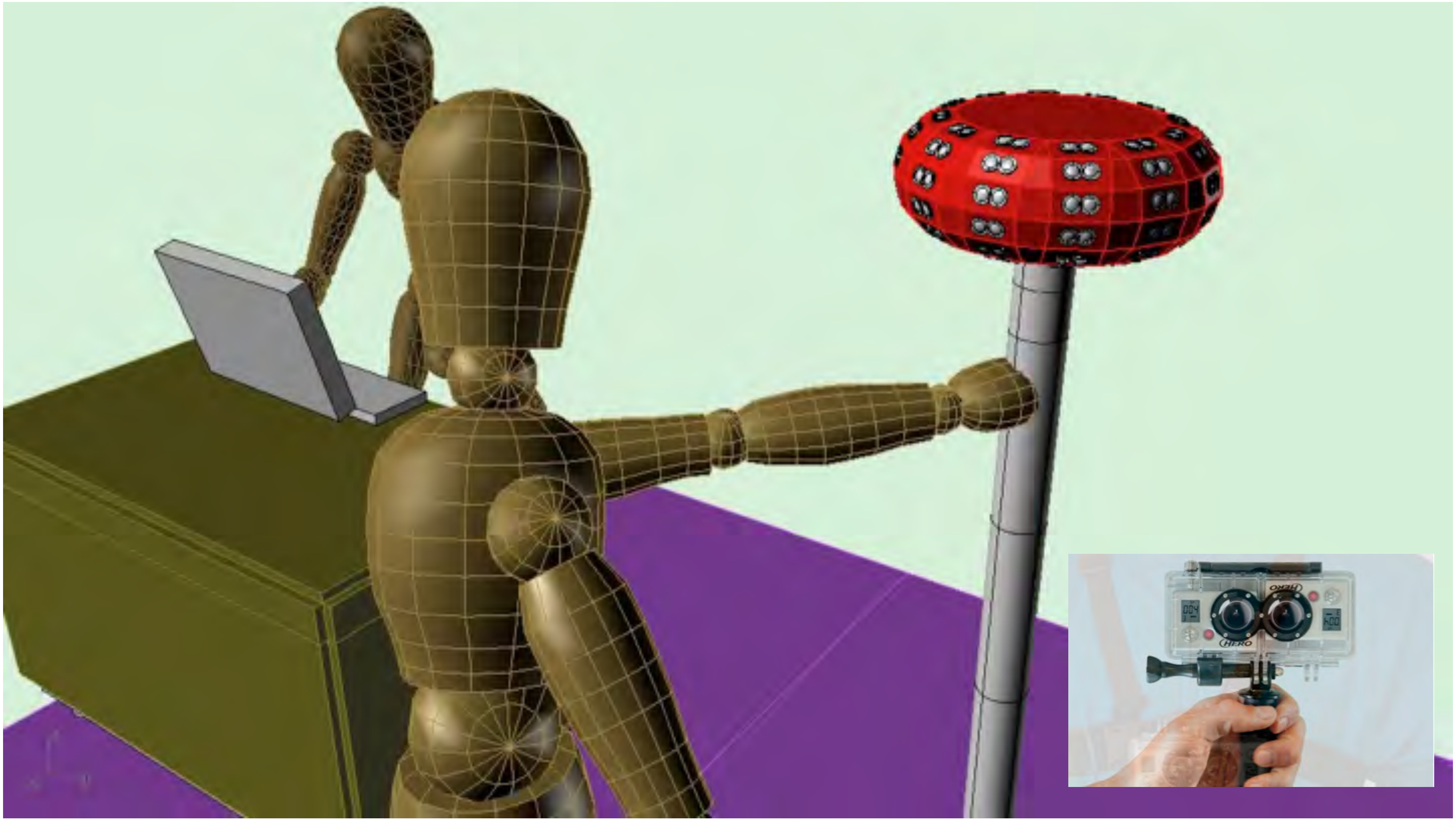
Points to Pans

72 Images—Per Eye: ½ Gigapixel per Eye/Frame 30 Gigapixels/s=~1Terabit Uncompressed



Source: Dick Ainsworth

Camelot Concept

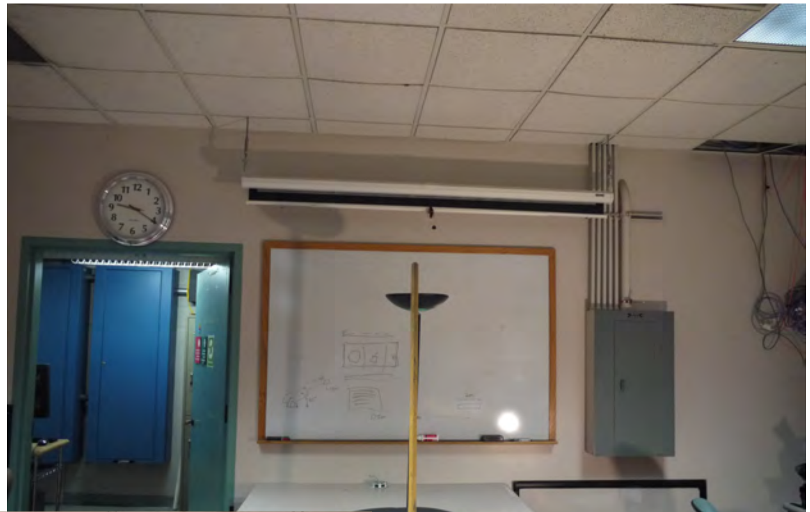


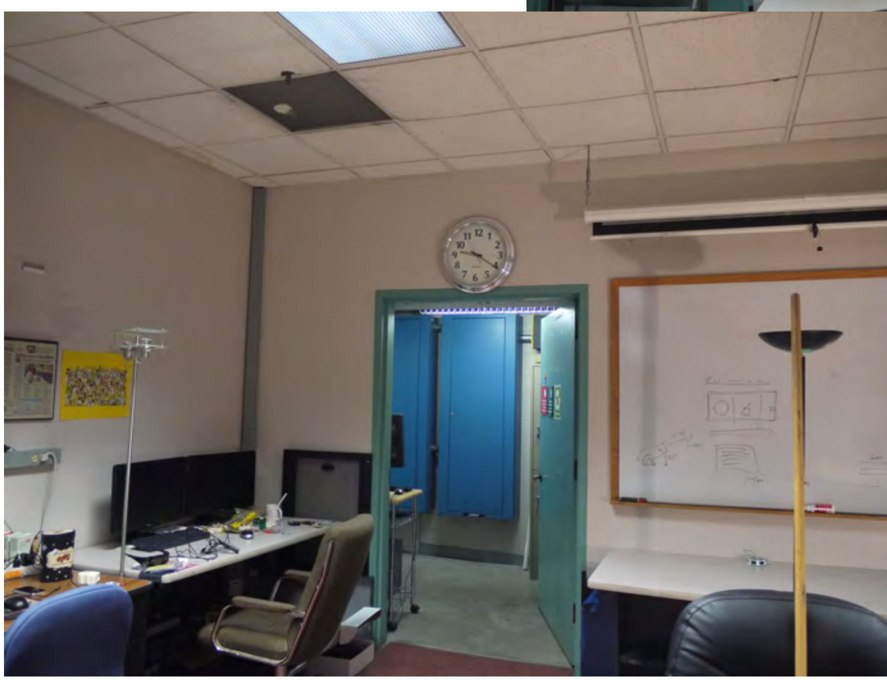
The Problem

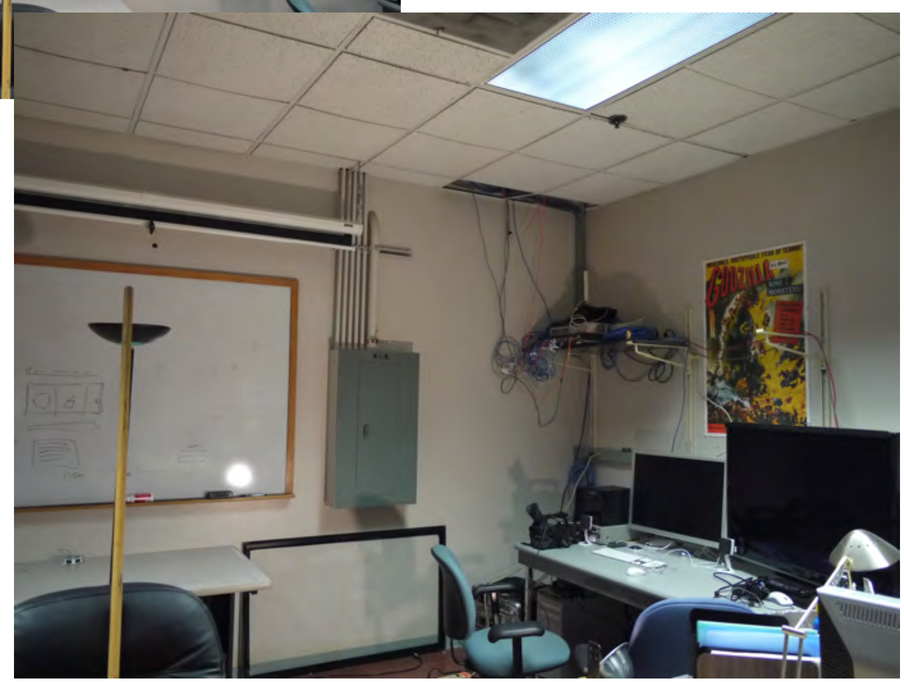
- To make panoramas stitch perfectly
  - Camera has to be rotated about the convergence point of the lens.
  - Our CAVEcam rotates about a point between the 2 points of convergence—this works well enough

  - Can't do this with a surrounding set of cameras, though

Rotation on Center of Convergence
Wood Pole ~5ft Lamp ~10ft AOV 50°

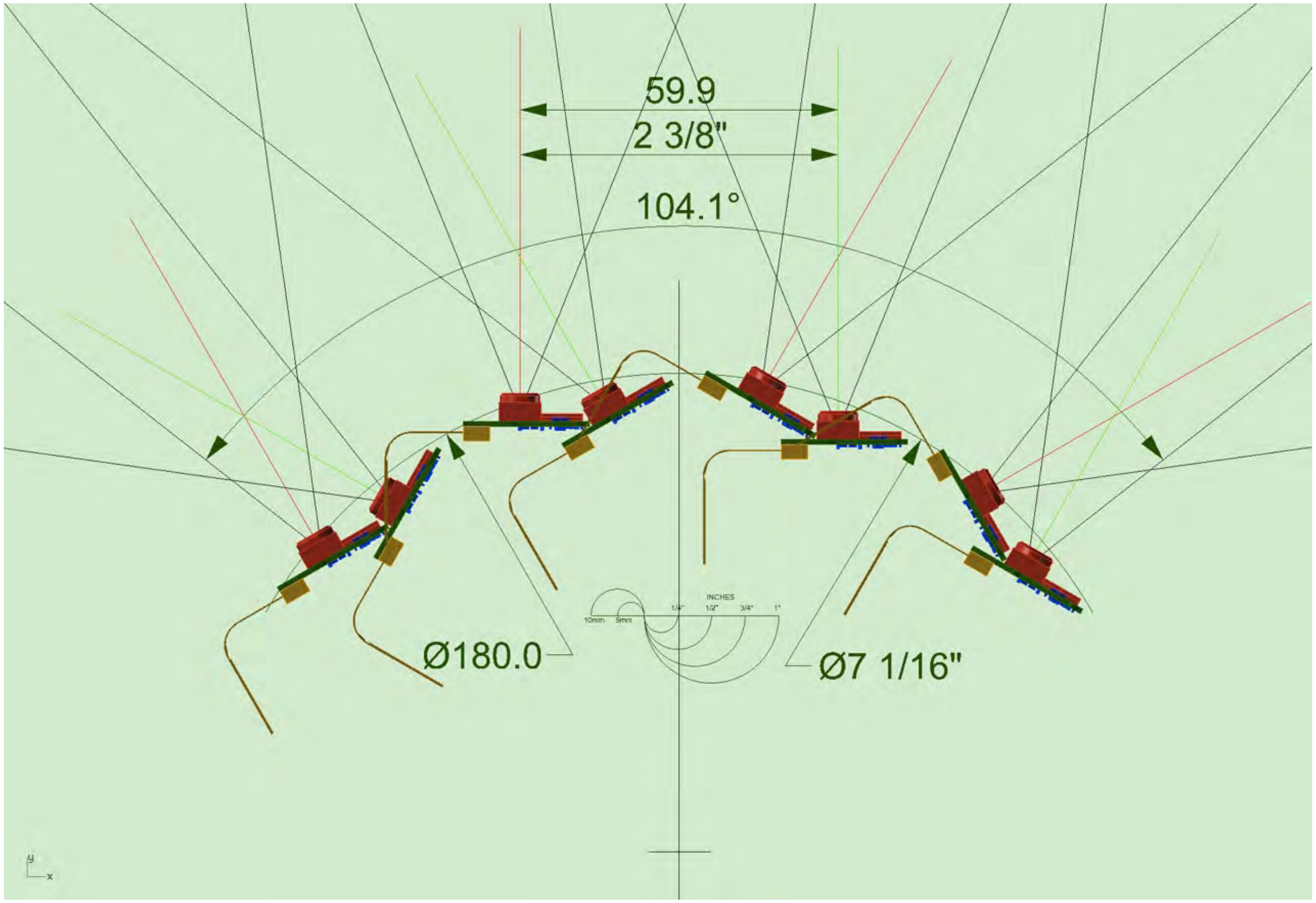






Rotation ≈ 6” From Center of Convergence
Wood Pole ~5ft Lamp ~10ft AOV 50°

Using Raspberry Pi Cameras



For a Video-Capable Camelot
What Do We Need in Cameras?

- Manual external control of exposure & focus
- Fire cameras simultaneously in synch
- Full remote camera control and communication
- Fixed wide angle lens
- Storage and retrieval of images
- Power and communication with cameras
- Very small size
- Cell phone camera with developer kit?

What are the End Goals of the Camelot Project?

- 15 Gigapixel/eye/second surround stereo movies
- Integration of photography into virtual environments
- Funding for full physical implementation will be sought from various sources.

Our planning, research, and education efforts are made possible, in major part, by funding from:

- US National Science Foundation (NSF) awards ANI-0225642, EIA-0115809, SCI-0441094, and CNS 0821155
- Calit2 Director's Office
- Calit2 UCSD Division (Qualcomm Institute)
- King Abdullah University of Science and Technology (KAUST)
- King Abdulaziz City for Science and Technology (KACST)
- University of Illinois at Chicago, Argonne National Laboratory, and Northwestern University for StarLight networking and management
- Pacific Wave and CENIC
- NTT Network Innovations Lab
- Cisco Systems, Inc.
- Pacific Interface, Inc.
- Many corporate donors

CineGrid Introduction

Great Data Great Network @ CICESE

October 10, 2013

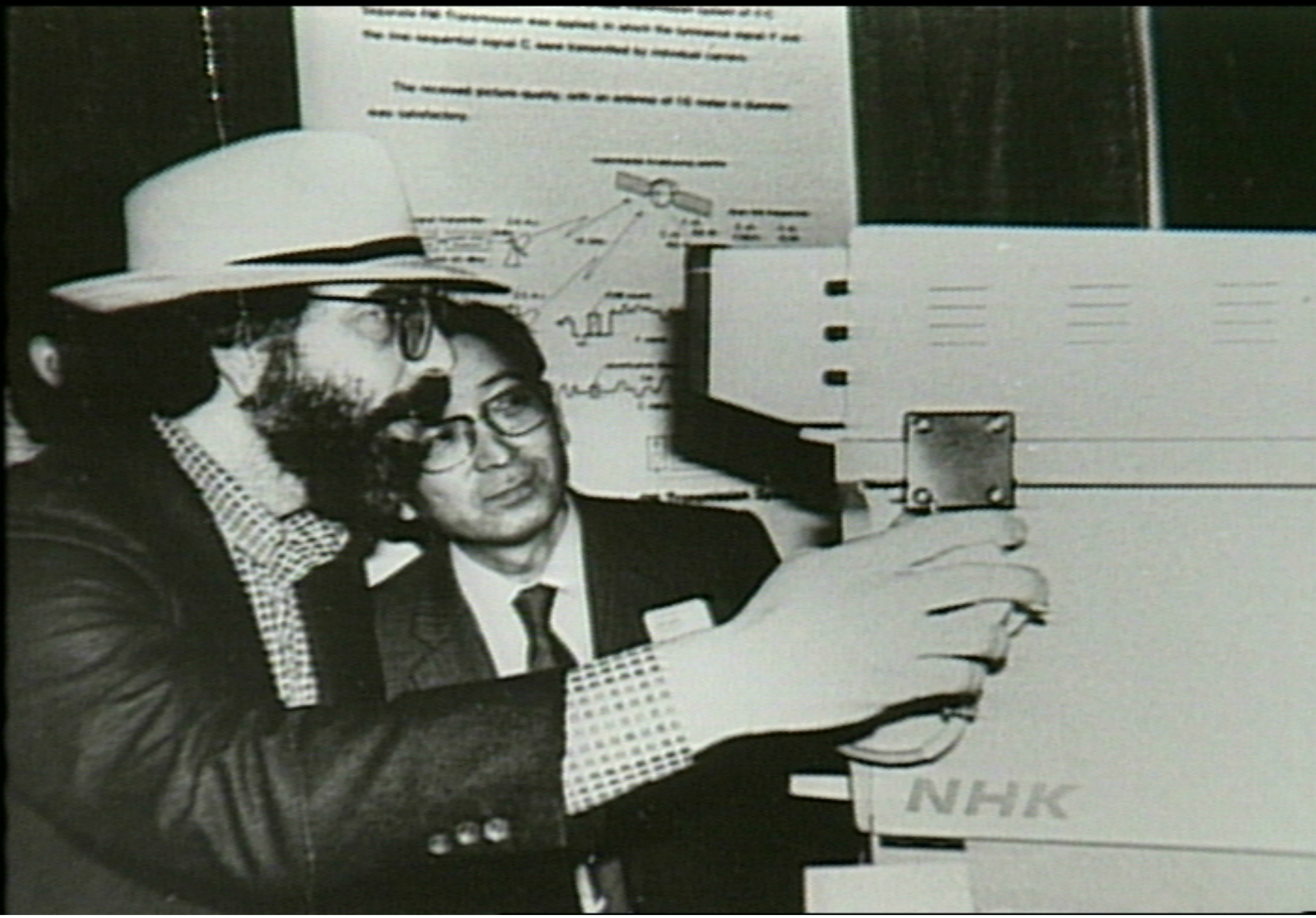
Laurin Herr
President, Pacific Interface Inc.

CineGrid is a non-profit international membership organization, a community of networked collaborators!

1981

Francis Ford Coppola with Dr. Takashi Fujio

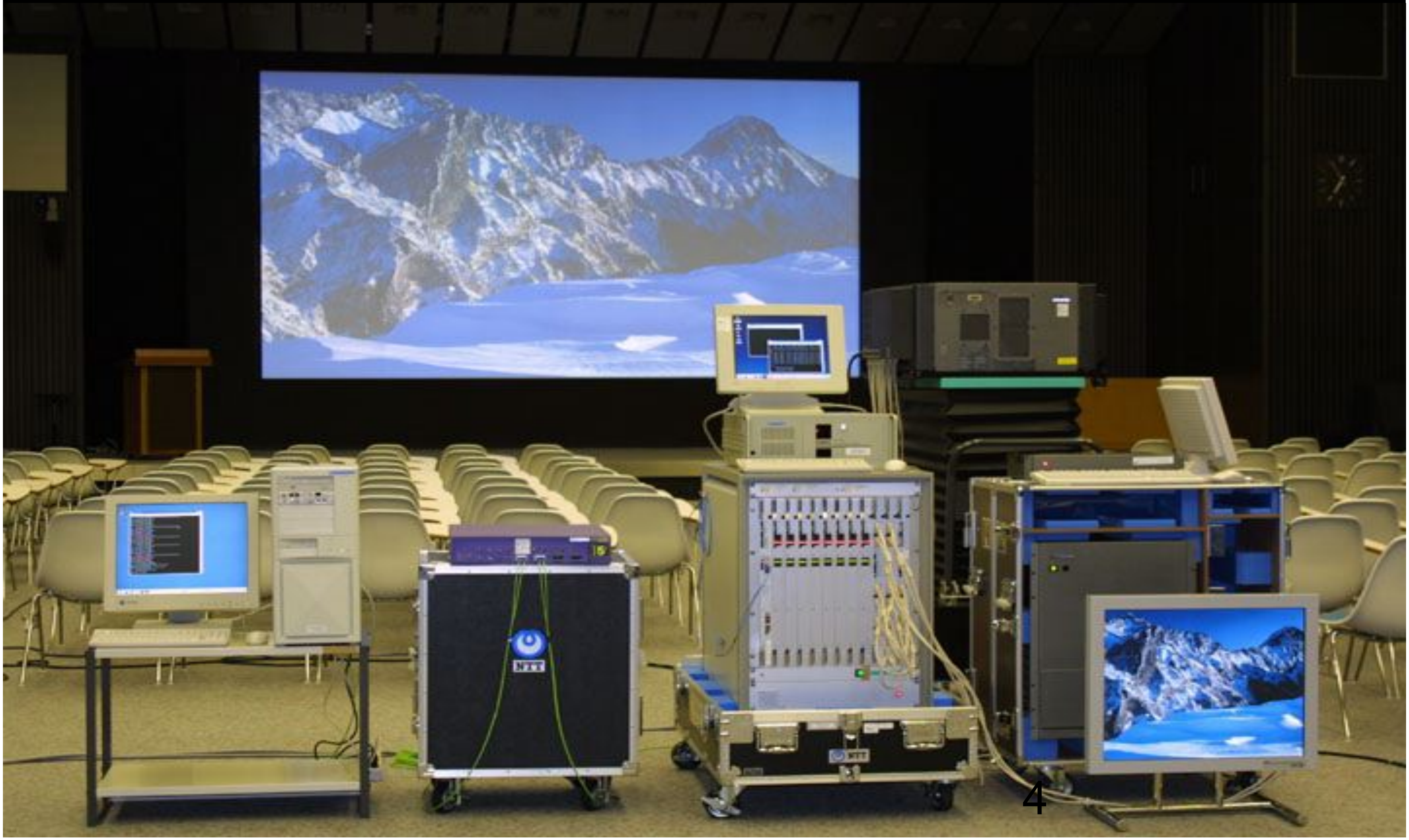
“First Look” at HDTV Electronic Cinema



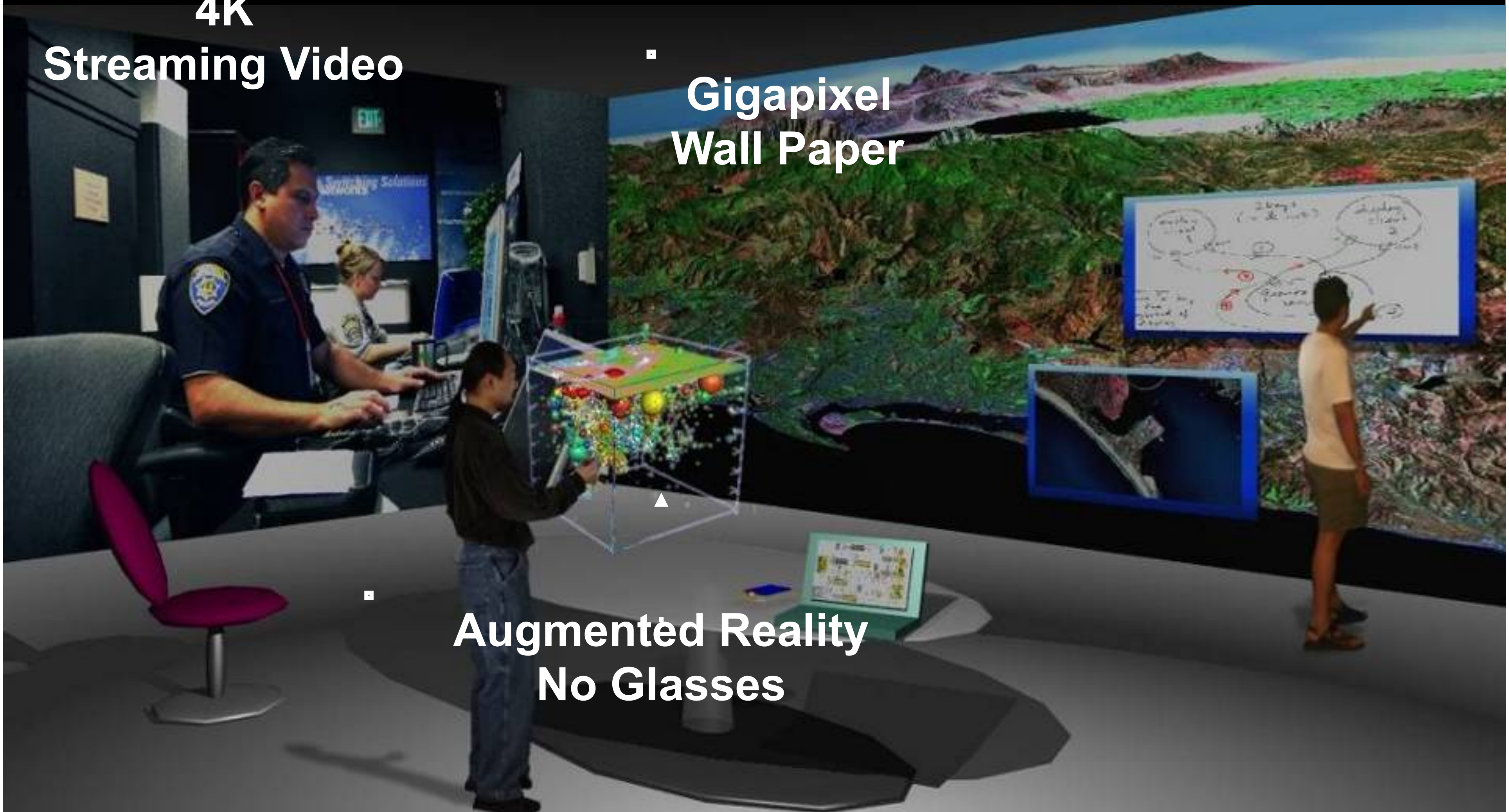
CineGrid

World's First 4K Digital Cinema System

NTT Network Innovations Laboratory (2001)

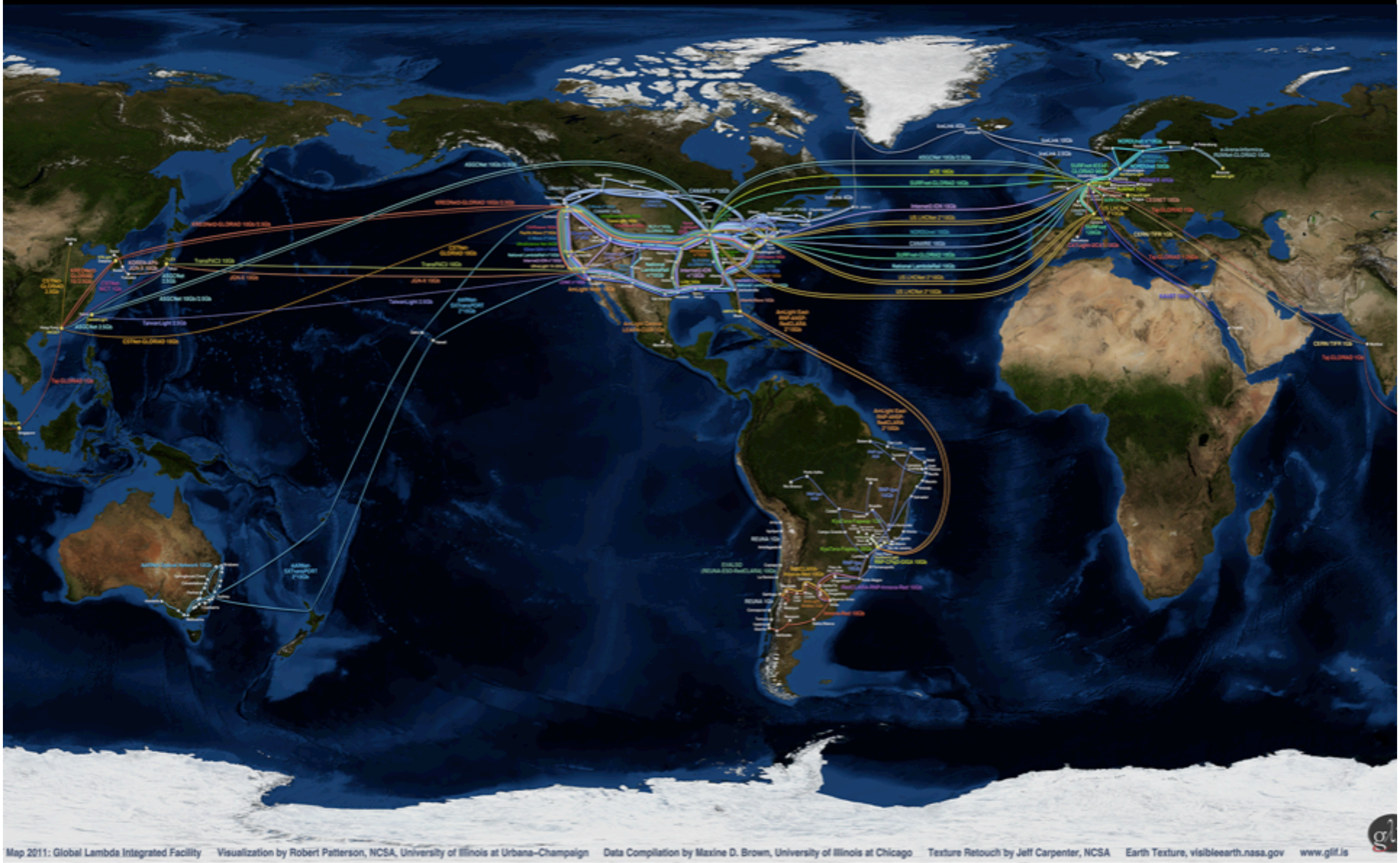


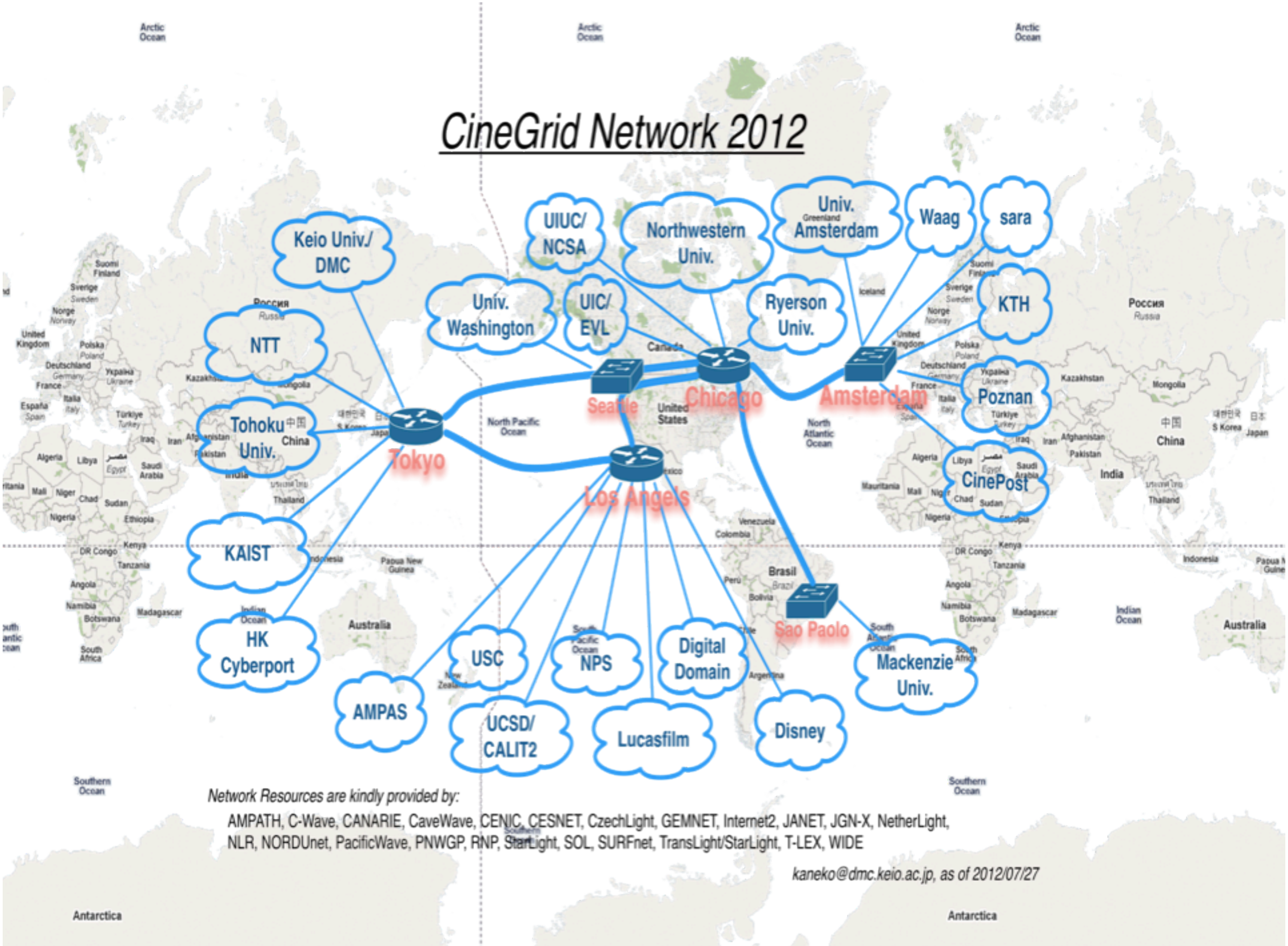
2004 OptIPuter Vision for the Next Decade
Gigapixels @ Terabits/sec



1 GigaPixel x 3 bytes/pixel x 8 bits/byte x 30 frames/sec ~ 1 Terabit/sec!

CineGrid projects run over the Global Lambda Integrated Facility (GLIF)



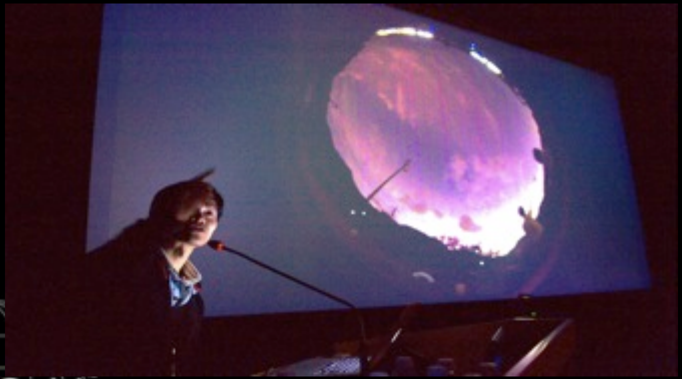

CineGrid Network 2012
Keio Univ./ DMC
NTT
Tohoku Univ.
KAIST
HK Cyberport
UIUC/ NCSA
Univ. Washington
UIC/ EVL
Northwestern Univ.
Univ. Amsterdam
Ryerson Univ.
Waag
sara
KTH
Poznan
CinePost
Tokyo
Seattle
Chicago
Amsterdam
Los Angels
Sao Paolo
Mackenzie Univ.
AMPAS
USC
UCSD/ CALIT2
NPS
Digital Domain
Lucasfilm
Disney
Network Resources are kindly provided by:
AMPATH, C-Wave, CANARIE, CaveWave, CENIC, CESNET, CzechLight, GEMNET, Internet2, JANET, JGN-X, NetherLight, NLR, NORDUnet, PacificWave, PNWGP, RNP, StarLight, SOL, SURFnet, TransLight/StarLight, T-LEX, WIDE
kaneko@dmc.keio.ac.jp, as of 2012/07/27

“Learning by Doing”

2008-2009 CineGrid Projects

Networked Optical Microscopy

Dual 4K Synchronized Remote Sensing



Global Live Streaming of Eclipse

Uncompressed 4K Interactive Cinema Color Grading

CineGrid

Tele-collaboration for audio post-production

realtime picture and sound editing synchronized over IP

Skywalker Sound@Marin ←·······→ Calit2@San Diego

Tele-collaboration for cinema post-production

Disney + Skywalker Sound + Digital Domain + Laser Pacific
NTT Labs + UCSD/Calit2 + UIC/EVL + Pacific Interface

The Growing Documentary

Places + Perspectives

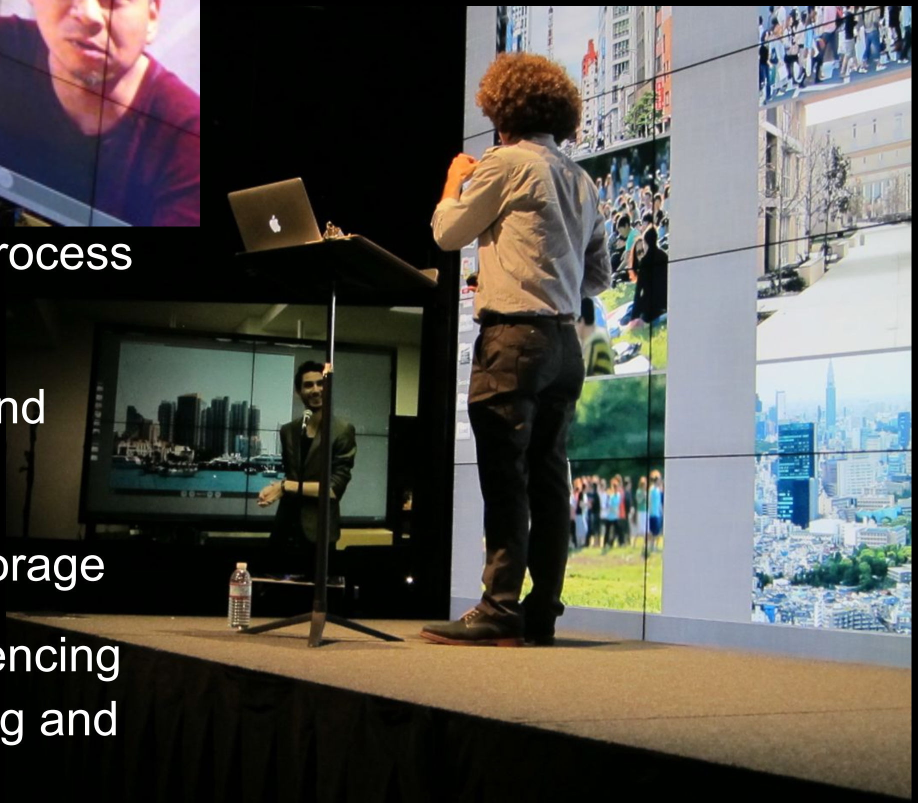
- Second iteration of Growing Documentary focussed on remote collaboration.

- International collaboration between graduate students from Keio Media Design, Keio University and undergraduate students from the Visual Arts Department at the University of California, San Diego.

Places + Perspectives

A Growing Documentary in HD

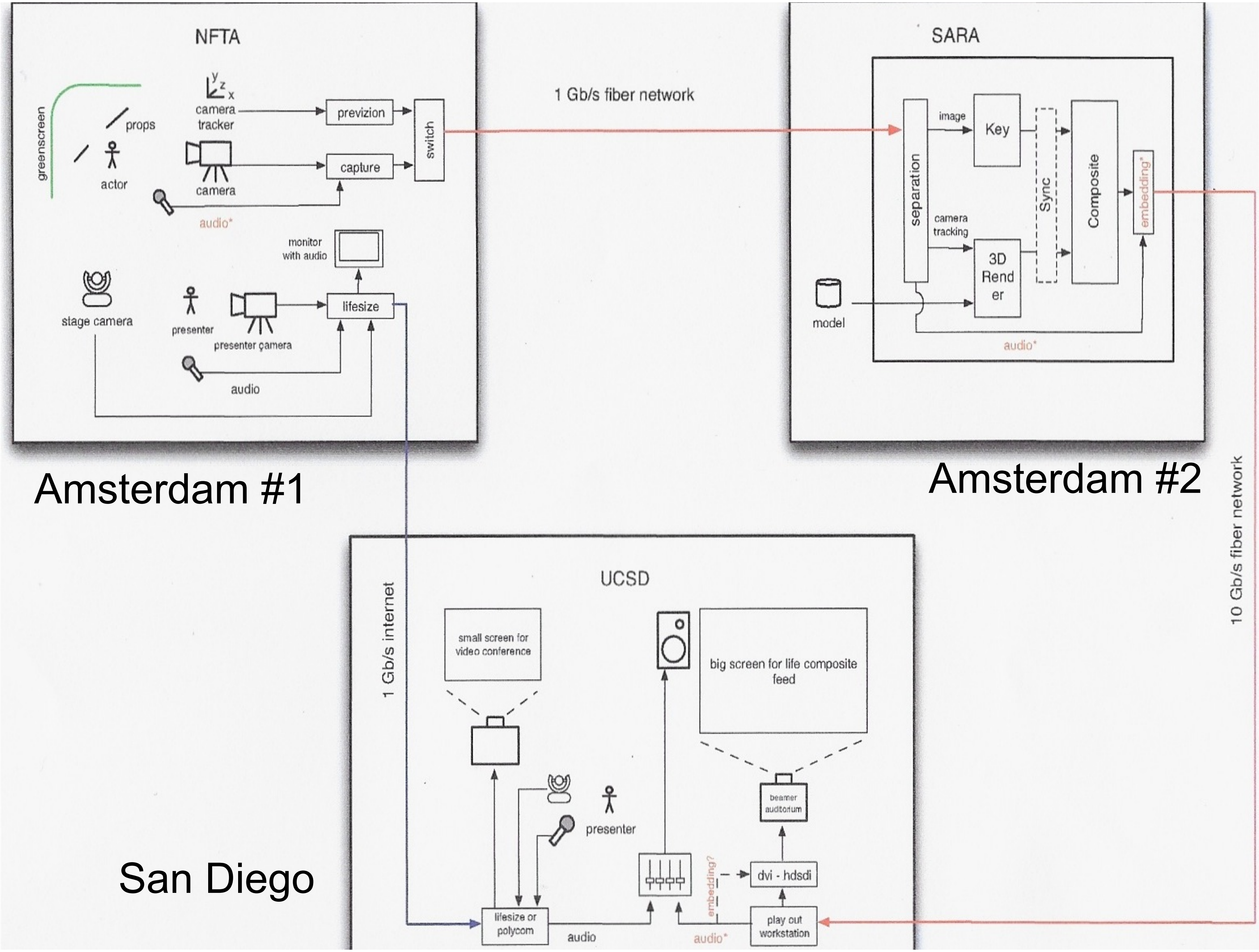
Keio University/KMD @ Hiyoshi
UCSD/Calit2 @ San Diego



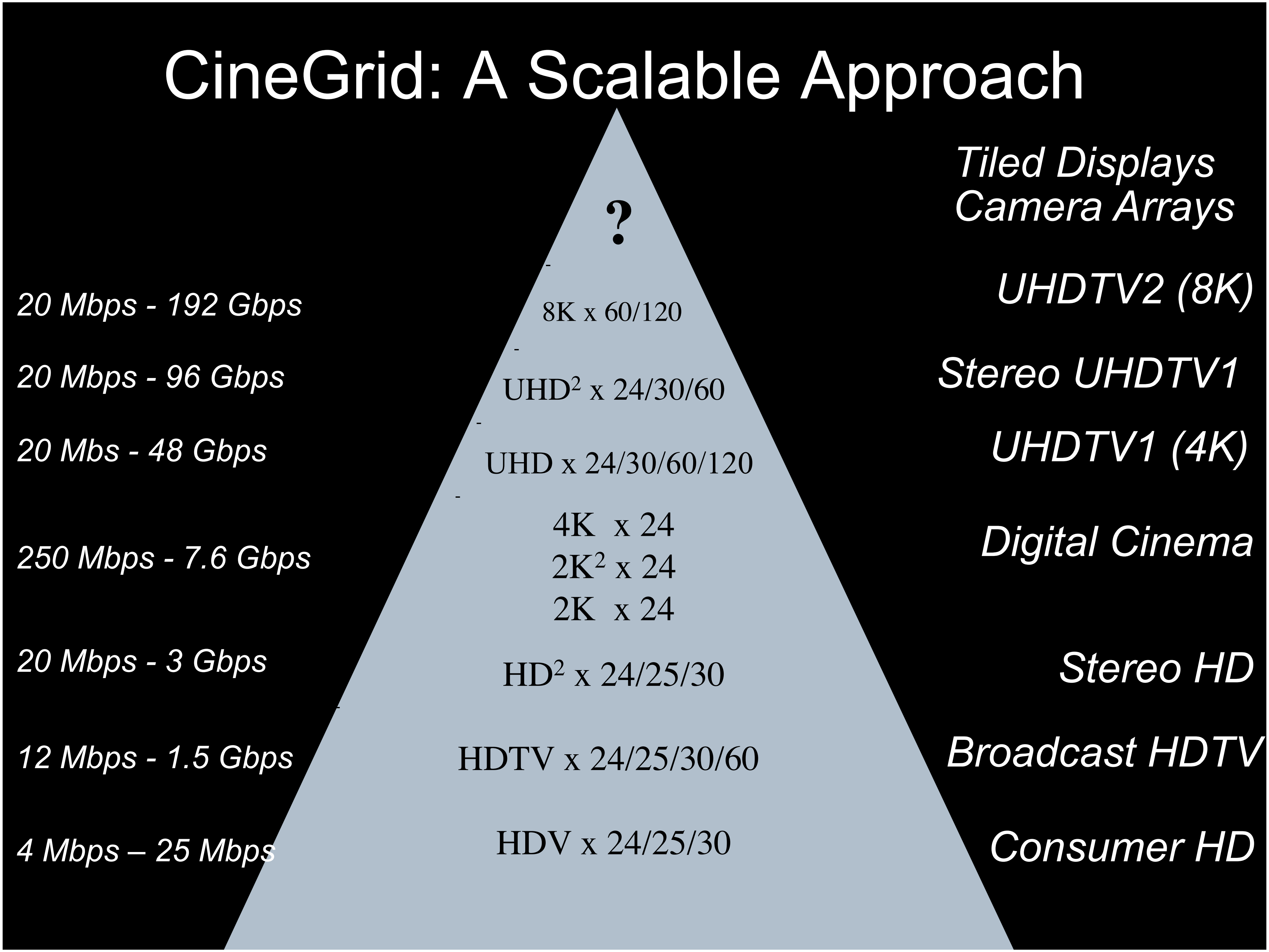
- Explore network-supported collaboration process
- Combine traditional production tools with emerging tools for media sharing, review and critique such as Vroom, CineSAGE & PIX
- Use cloud server for media transfer and storage
- Use multi-channel 4K/HD video teleconferencing for face-to-face discussions, context sharing and project development

Directing Remote Live Shoot of Virtual Set Acting with Live Compositing in the Cloud

Live action camera, actors, green screen at NFTA (Amsterdam #1)
Virtual set compositing at SARA (Amsterdam #2)
Remote viewing and direction at UCSD/Calit2 Vroom (San Diego)


NFTA
greenscreen
props
actor
camera tracker
camera
audio*
previzion
capture
switch
monitor with audio
lifesize
stage camera
presenter
presenter camera
audio
Amsterdam #1
1 Gb/s fiber network
SARA
separation
image
Key
camera tracking
3D Rend er
Sync
Composite
embedding*
model
audio*
Amsterdam #2
10 Gb/s fiber network
UCSD
1 Gb/s internet
small screen for video conference
big screen for life composite feed
presenter
beamer auditorium
dvi - hdsdi
embedding?
lifesize or polycom
audio
audio*
play out workstation
San Diego

Directing Remote Live Shoot of Virtual Set Acting with Live Compositing in the Cloud


CineGrid: A Scalable Approach
Tiled Displays
Camera Arrays
?
20 Mbps - 192 Gbps
8K x 60/120
UHDTV2 (8K)
20 Mbps - 96 Gbps
UHD^2 x 24/30/60
Stereo UHDTV1
20 Mbs - 48 Gbps
UHD x 24/30/60/120
UHDTV1 (4K)
250 Mbps - 7.6 Gbps
4K x 24
2K^2 x 24
2K x 24
Digital Cinema
20 Mbps - 3 Gbps
HD^2 x 24/25/30
Stereo HD
12 Mbps - 1.5 Gbps
HDTV x 24/25/30/60
Broadcast HDTV
4 Mbps – 25 Mbps
HDV x 24/25/30
Consumer HD

Great Media Needs Great Networks

Horizontal Pixels	Vertical Pixels	Frames per Second (nominal)	Total Payload (nominal) 10-bit 4:2:0 10-bit 4:2:2	Total Payload (nominal) 12-bit 4:2:0 12-bit 4:2:2 12-bit 4:4:4 10-bit 4:4:4:4
3840	2160	120	24Gbit/Sec	48Gbit/Sec
3840	2160	60	12Gbit/Sec	24Gbit/Sec
3840	2160	50		
3840	2160	30	6Gbit/Sec	12Gbit/Sec
3840	2160	25		
3840	2160	24		

Table 4 UHDTV1 Image formats and payloads

UHDTV2 = 4x data rates of UHDTV1

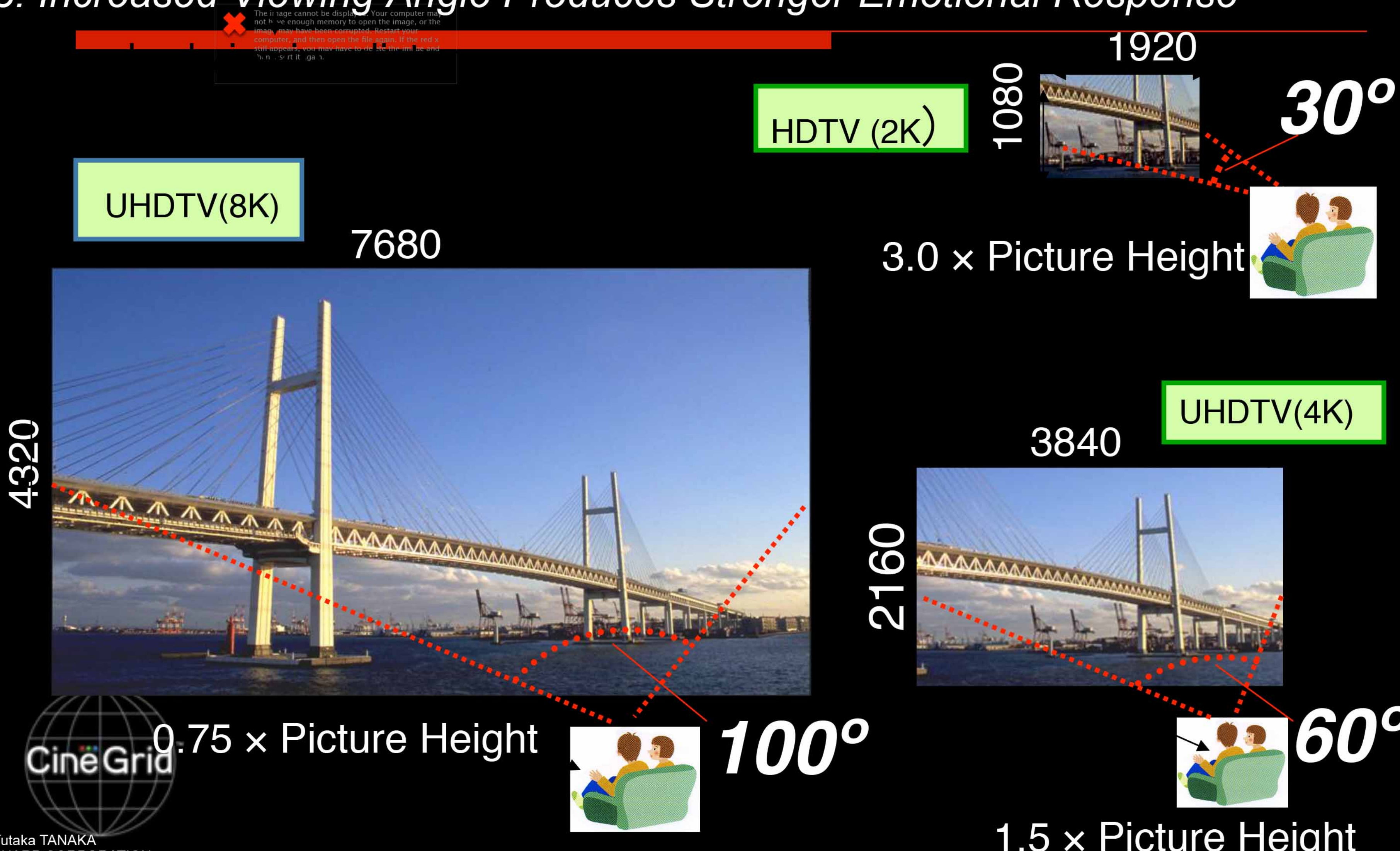
NO SINGLE CABLE REAL TIME INTERFACE DEFINED YET

4K is No Longer Just for Cinema!
Coming: Ultra High Definition (UHD) TV and Gaming at Home!

Picture of Laurin Herr in Shinagawa Station, Tokyo Japan on October 27, 2012

Why is more resolution is better?

1. More Resolution Allows Closer Viewing of Larger Image
2. Closer Viewing of Larger Image Increases Viewing Angle
3. Increased Viewing Angle Produces Stronger Emotional Response



Yutaka TANAKA
SHARP CORPORATION
Advanced Image Research Laboratories

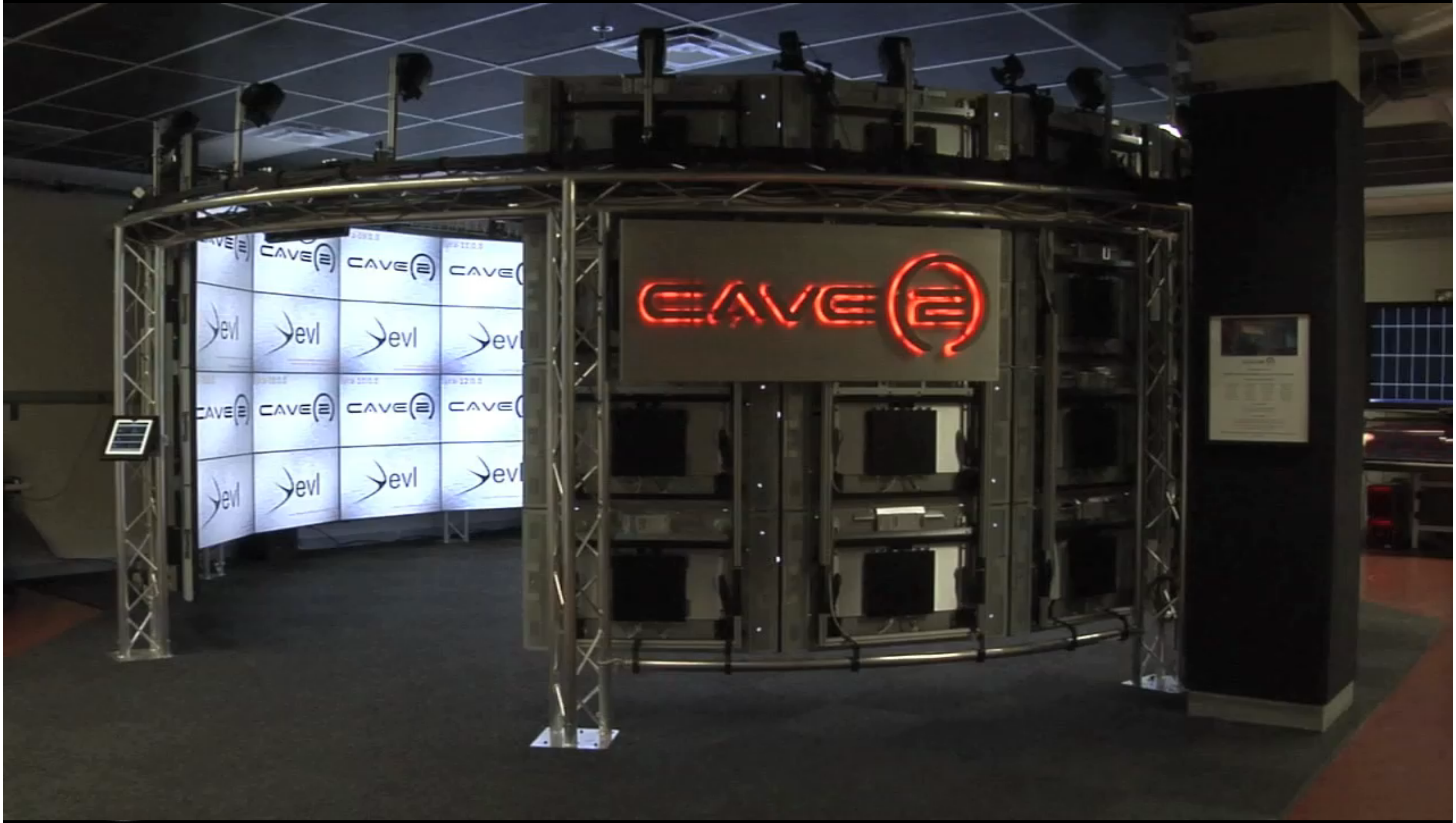
“Vroom” for Networked Scientific Visualization
Dr. Larry Smarr @ Calit2/UCSD

CineGrid

CAVE(2) Advanced Collaboration Environment

72Mpix, 3D stereo, tracking, 20 channels audio

100Gbps network connectivity



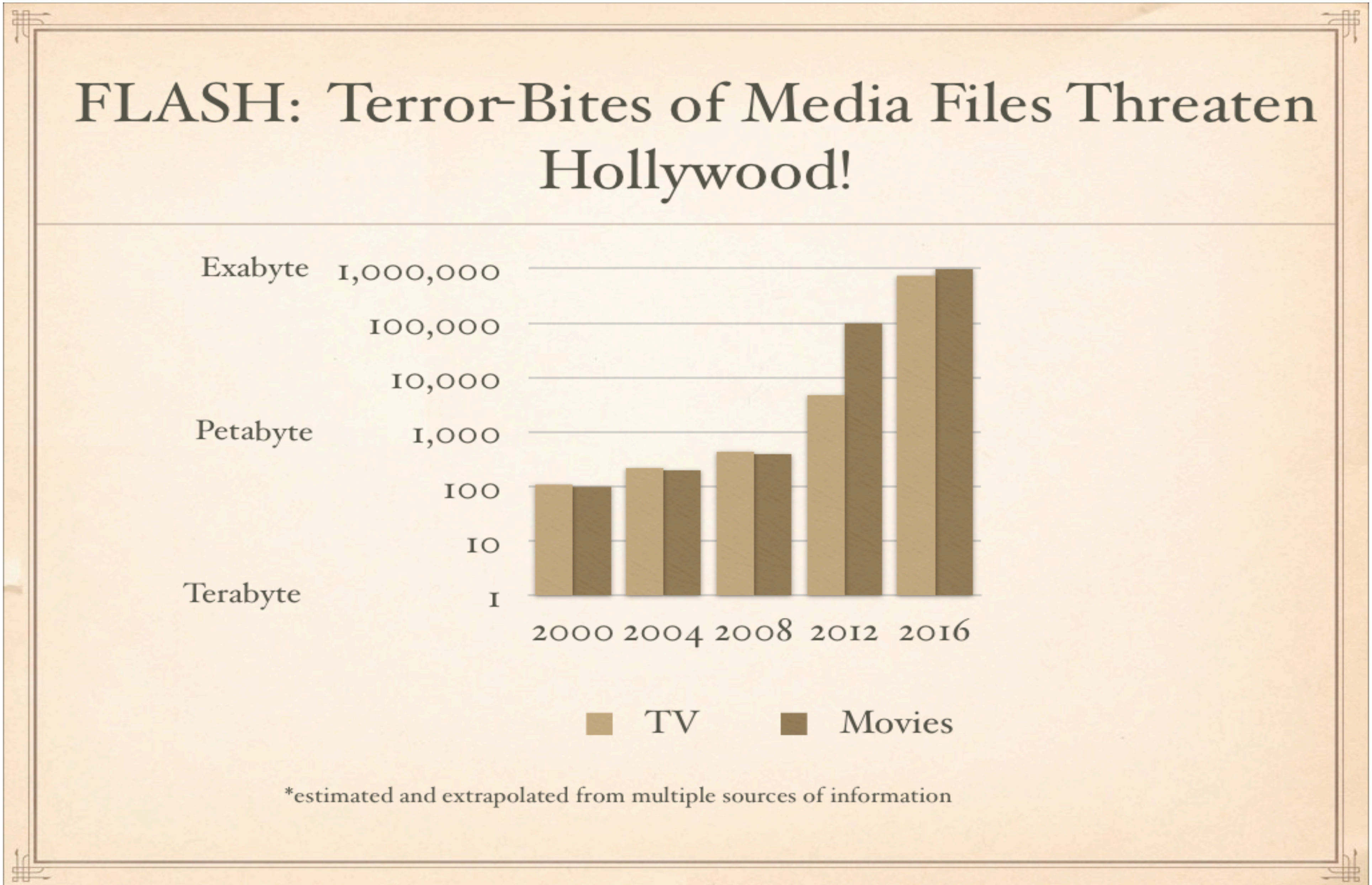
CAVE(2) Advanced Collaboration Environment
Inside the Starship Enterprise

GROWTH OF DATA

- The conversion to digital formats has fueled the increase of media data.
- A TV show (HD) generate 30hrs of media @200Mb/s= 2.7 TB. A season is 60TB. 3D doubles this storage; 4k quadruples the storage (1/4 Petabyte).
- Movies (DCI) generate 1-2 PB per title; 3D more than double (5-6TB); HFR can double or triple these numbers (10-15 TB).
- Distribution. Each new on-line distributor requires their own file formats, usually at 1-4Mb/s. For a 1000hr library, this represents 0.5 to 2 PB for each distribution deal

Thursday, February 21, 13

From presentation at 2013 Hollywood Post Alliance Technology Retreat
“File-based Production Architecture End-to-End” by Setos and DeFilippis



From presentation at 2013 Hollywood Post Alliance Technology Retreat
“File-based Production Architecture End-to-End” by Setos and DeFilippis

THE BIG ISSUES

- The Storage Challenge
- Archiving the data
- File Format anarchy

Thursday, February 21, 13

From presentation at 2013 Hollywood Post Alliance Technology Retreat
“File-based Production Architecture End-to-End” by Setos and DeFilippis

Digital Archive Layers

LIFESPAN	HARDWARE	SOFTWARE
3 ➞ 5 YEARS	HOST COMPUTER	• APPLICATION SOFTWARE • OPERATING SYSTEM • DEVICE DRIVERS
5 ➞ 10+ YEARS	PHYSICAL INTERFACE	• INTERFACE FIRMWARE
3 ➞ 5 YEARS	MEDIA DRIVE	• DRIVE CONTROL FIRMWARE
.5 ➞ 10 YEARS	MEDIA	• FILE SYSTEM • DATA FILE FORMAT • PHYSICAL RECORDING FORMAT
VARIES	TRAINED PERSONNEL	
VARIES	FUNDING	

CineGrid 2013 International Workshop
December 9 – 11, 2013
San Diego, CA

CineGrid™

www.cinegrid.org

Numerical Modeling for Tropical Cyclone Landfall in Mexico

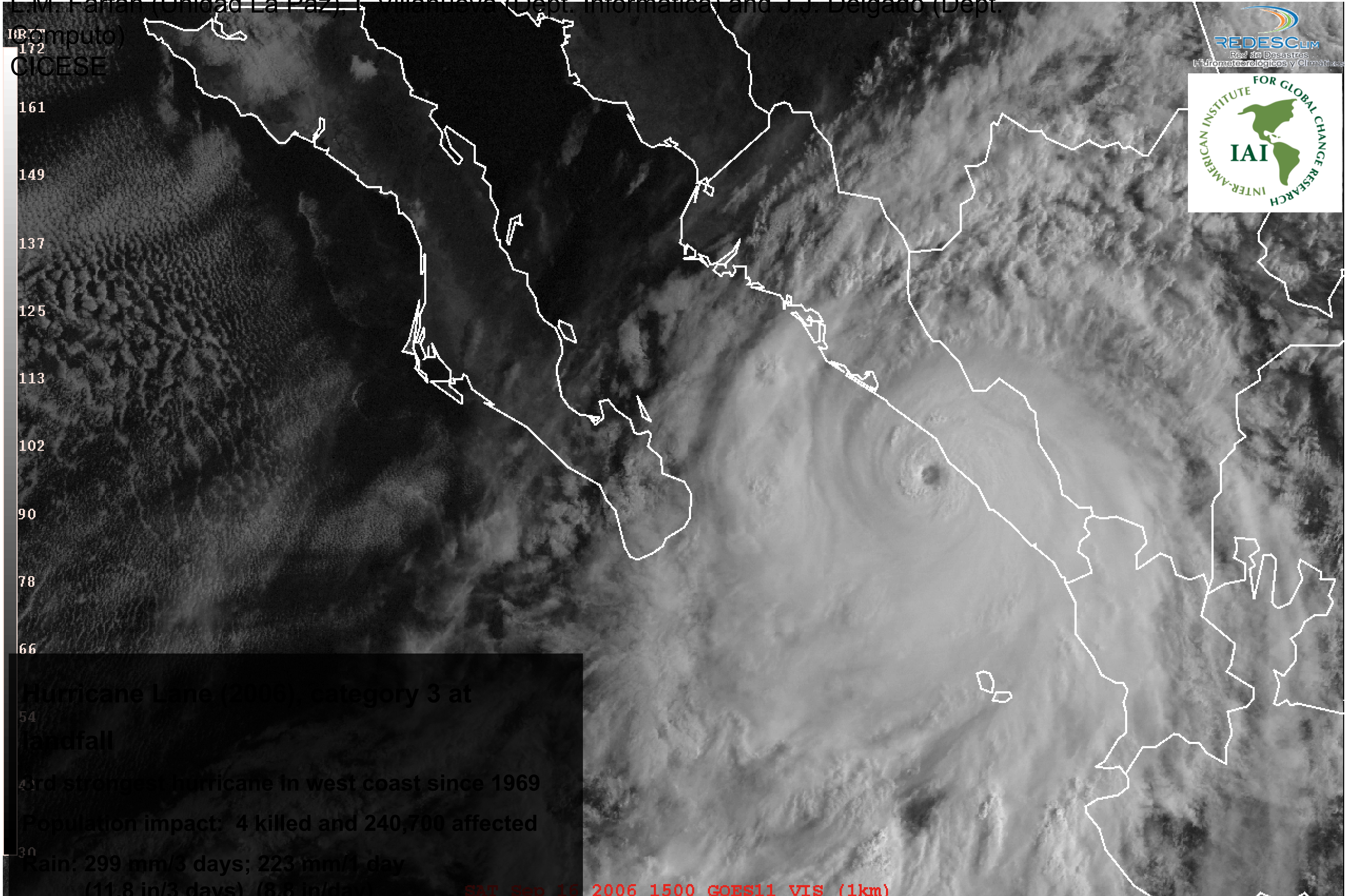
205

CONACYT

REDESClim
Red de Desastres Hidrometeorológicos y Climáticos

INTER-AMERICAN INSTITUTE FOR GLOBAL CHANGE RESEARCH
IAI

L.M. Farfán (Unidad La Paz), I. Villanueva (Dept. Informática) and J.J. Delgado (Dept. Cómputo)
CICESE



Hurricane Lane (2006): category 3 at landfall

3rd strongest hurricane in west coast since 1969

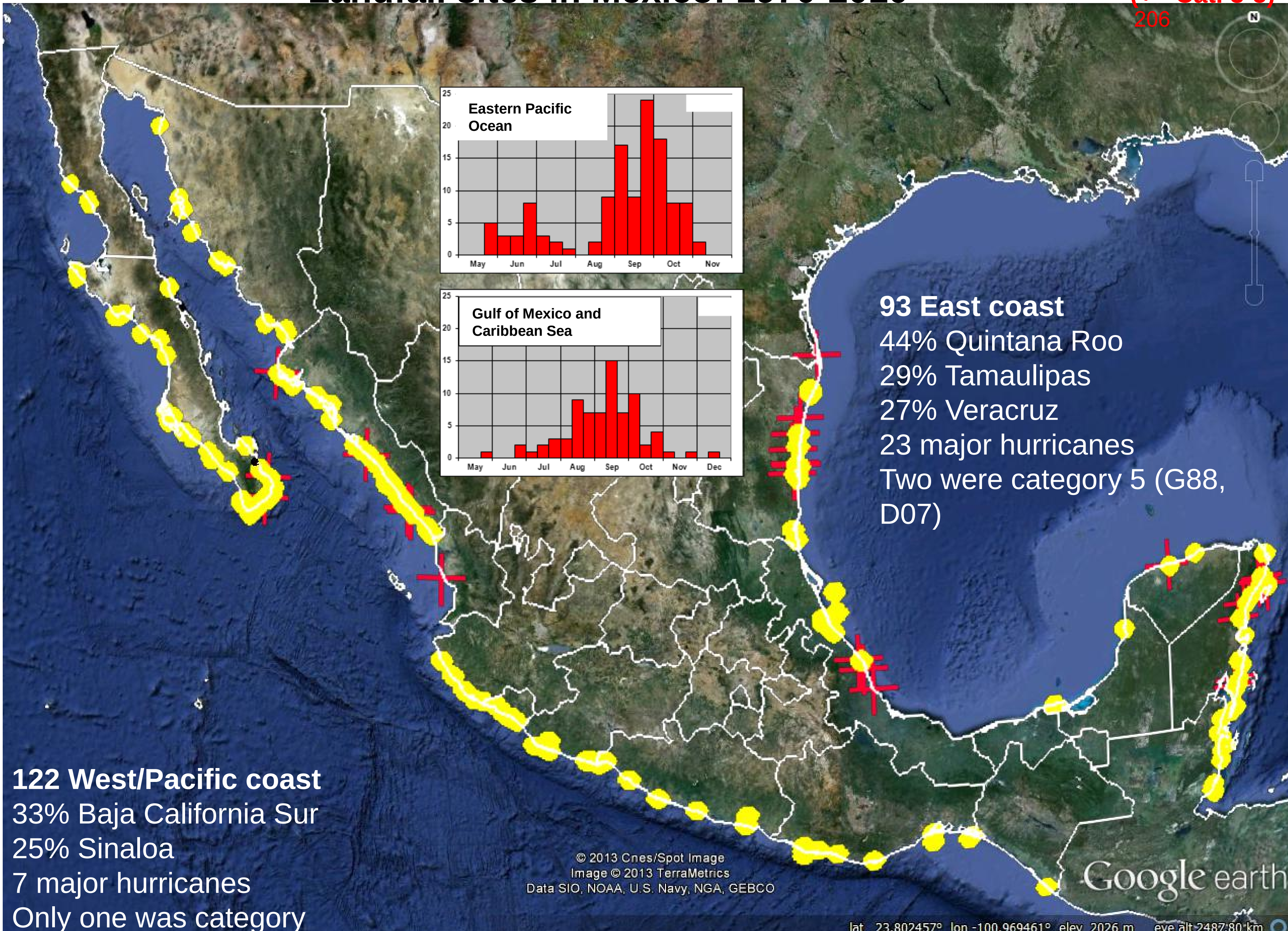
Population impact: 4 killed and 240,700 affected

Rain: 299 mm/3 days; 223 mm/1 day
(11.8 in/3 days) (8.8 in/da

Landfall sites in Mexico: 1970-2010

(+= Cat. 3-5)

206



93 East coast

44% Quintana Roo

29% Tamaulipas

27% Veracruz

23 major hurricanes

Two were category 5 (G88, D07)

122 West/Pacific coast

33% Baja California Sur

25% Sinaloa

7 major hurricanes

Only one was category

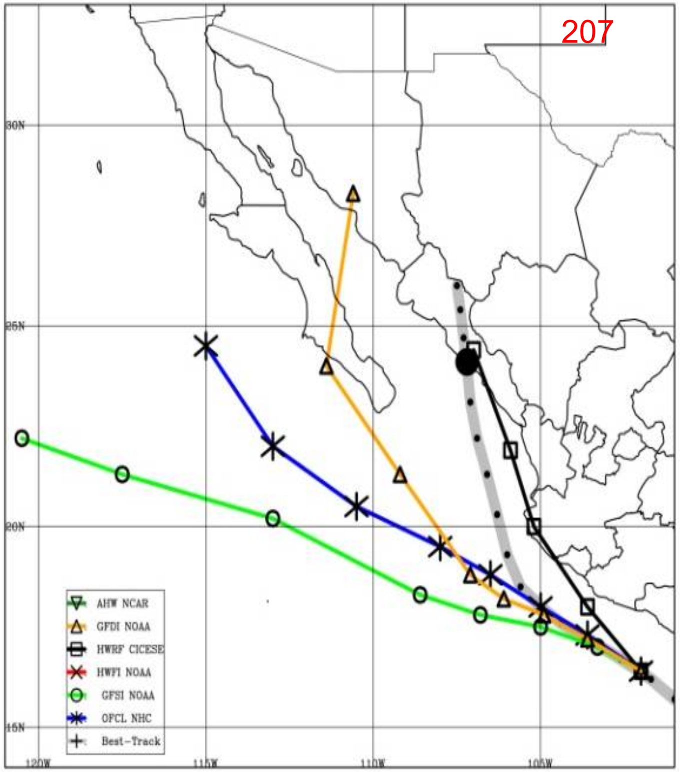
Observations and forecast tracks from

→

-National Hurricane Center

-Official forecast (National Hurricane Center)
-Operational models (NOAA)

-HWRF model applied at CICESE (black line/boxes)



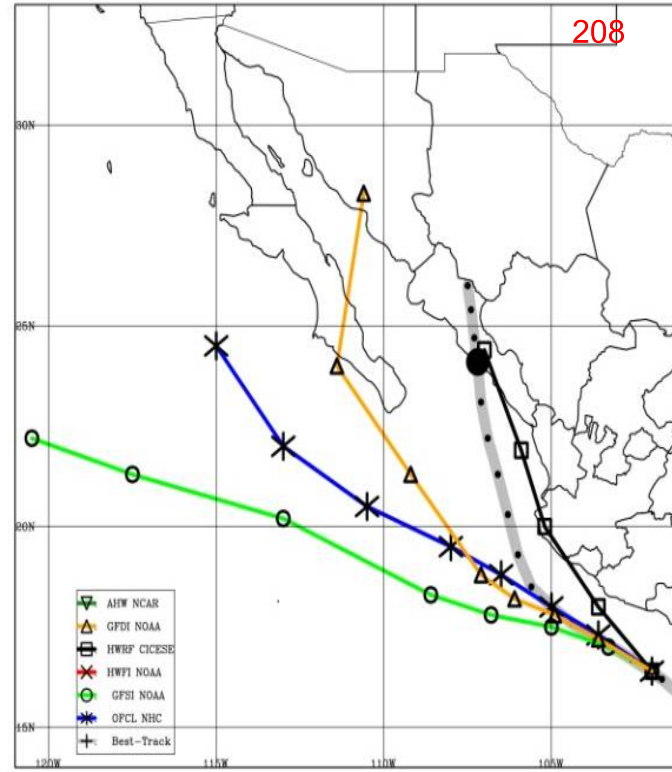
Observations and forecast tracks from

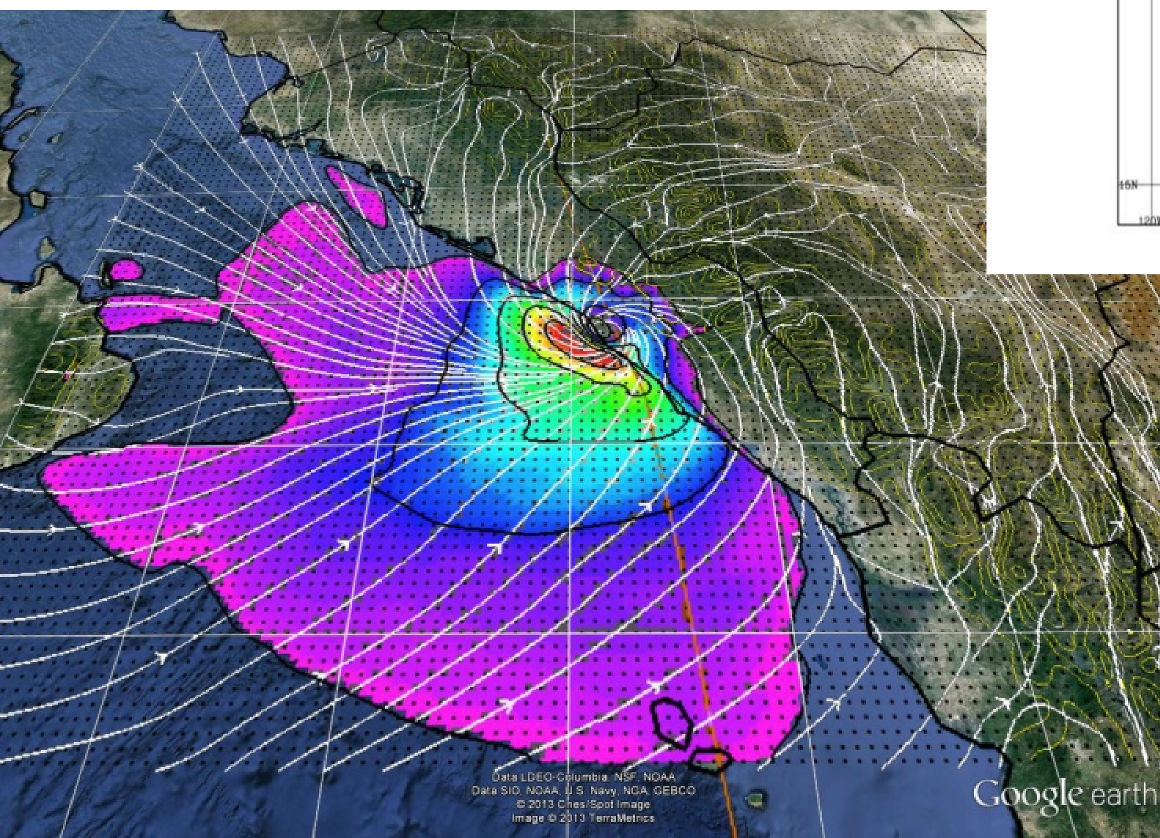
→

-National Hurricane Center

-Official forecast (National Hurricane Center)
-Operational models (NOAA)

-HWRF model applied at CICESE (black line/boxes)





← **HWRF model simulations**

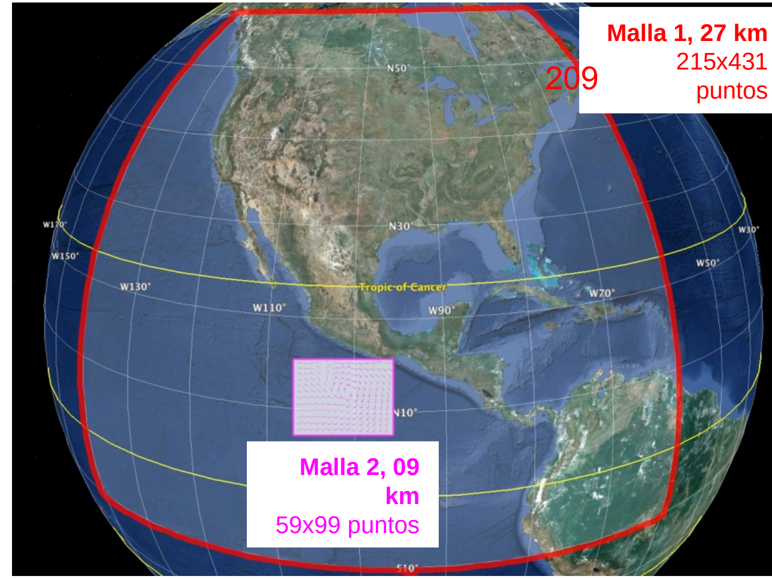
-Input data from global scale analysis

-Telemática/Cómputo resources

-5-day (120 hour) forecast

Figure shows 10-m streamlines

HWRF™
The Hurricane Weather Research And Forecast System



http://www.dtcenter.org/HurrWRF

- HWRF is a numerical model able to predict temporal changes of tropical cyclone positions and intensity. The air-sea interactions are considered.

- It has been applied, in an operational environment, since the storm season of 2007.

- Outer grid: 80x80^0, spatial resolution 27 km
  2 nested, moving grids with 9- and 3-km resolutions
  42 vertical levels from the surface to 20 km.

- For each active tropical cyclone, HWRF (at NOAA) performs 126-hour forecasts every 6 hours
  http://www.emc.ncep.noaa.gov/?branch=HWRF

HWRF™
The Hurricane Weather Research And Forecast System

http://www.dtcenter.org/HurrWRF

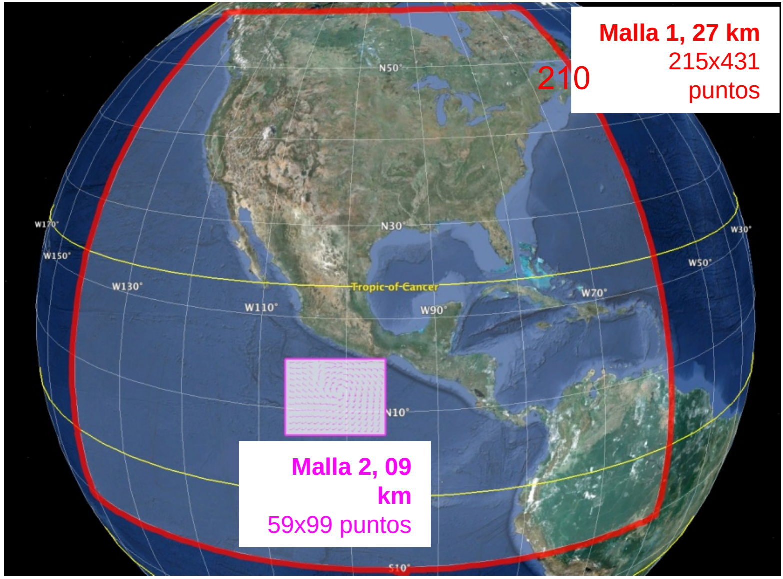
- HWRF is a numerical model able to predict temporal changes of tropical cyclone positions and intensity. The air-sea interactions are considered.

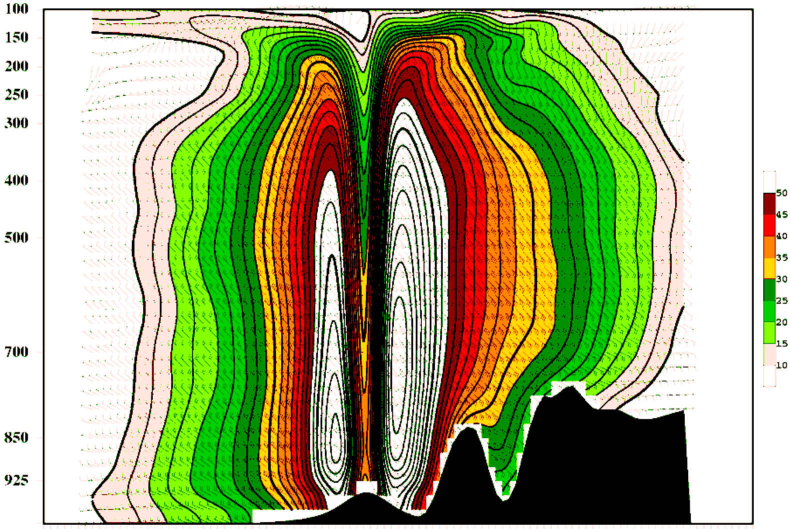
- It has been applied, in an operational environment, since the storm season of 2007.

- Outer grid: 80x80^0, spatial resolution 27 km
  2 nested, moving grids with 9- and 3-km resolutions
  42 vertical levels from the surface to 20 km.

- For each active tropical cyclone, HWRF (at NOAA) performs 126-hour forecasts every 6 hours
  http://www.emc.ncep.noaa.gov/?branch=HWRF

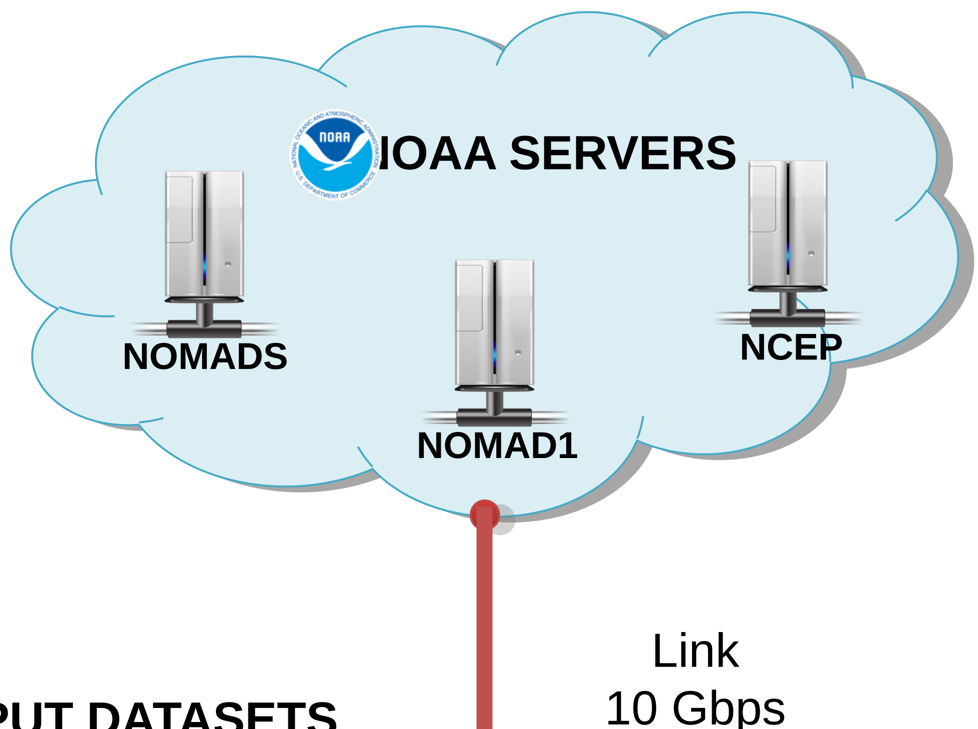
- At CICESE, we are using the HWRF model since 2012 in the study of recent landfall events over northwestern Mexico
  http://hwrf2.cicese.mx/





Vertical cross section of wind field
From surface up to 100 mb (15 km)
Black filling represents inland topography

CURRENT INFRASTRUCTURE



CICESE NODE HARDWARE SPECIFICATIONS

2 Processors Intel(R) Xeon(R) CPU E5---26600 @ 2.20GHz, 8 cores/node
64 GB RAM Memory
1 Gbps Network Card
2 TB storage

OUTPUT DATASETS

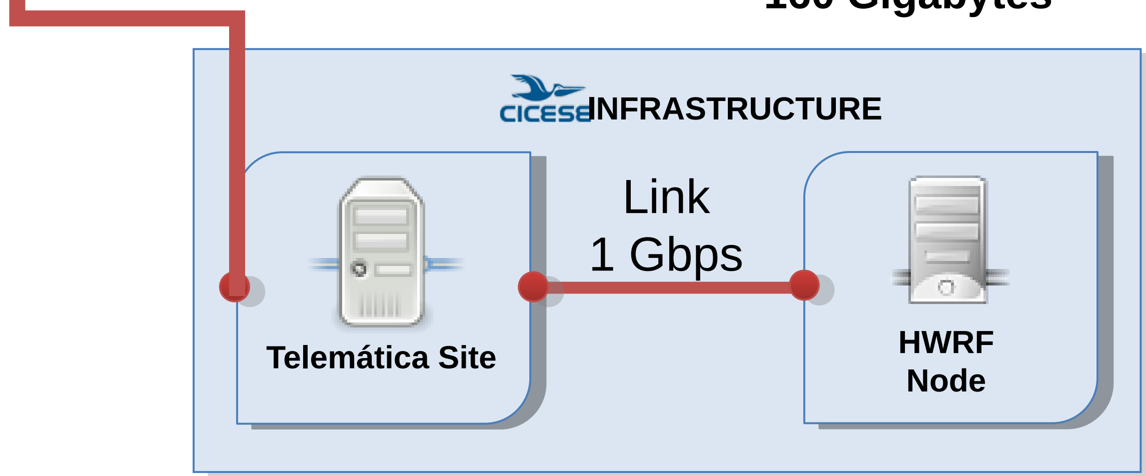
HWRF results include wind, temperature, humidity, rainfall, etc.

160 Gigabytes

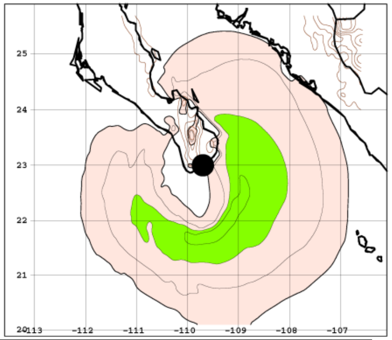
INPUT DATASETS

Meteorological, global information in gridded format

117 Gigabytes

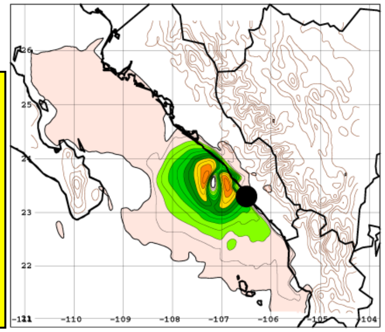


Georgette (2010)



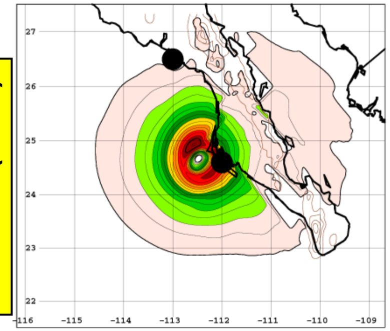
Initialization at 12UTC 21 September
Landfall
HWRF: Sep/21/09 (37 kt, 996mb)
NHC: Sep/21/18 (35 kt, 999mb)

Rick (2009)



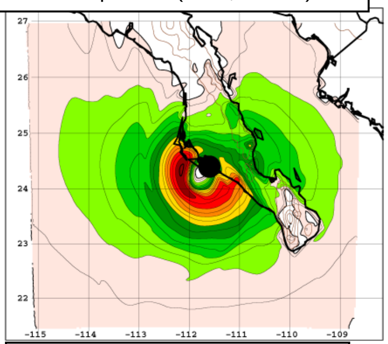
Initialization at 00UTC 20 October
Landfall
HWRF: Oct/22/00 (76 kt, 975mb)

Jimena (2009)



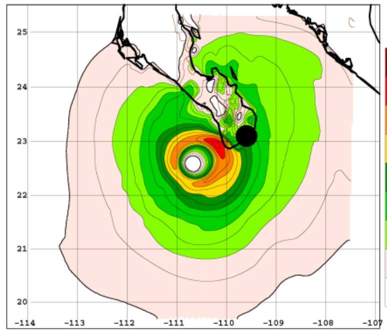
Initialization at 00UTC 01 September
Landfall
HWRF: Sep/02/15 (97 kt, 958mb)
NHC: Sep/02/12 (90 kt, 971mb)

Norbert (2008)



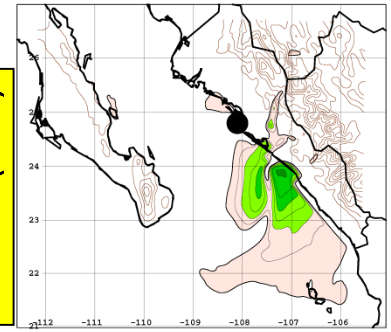
Initialization at 12UTC 09 October
Landfall
HWRF: Oct/11/18 (93 kt, 956mb)

Henriette (2007)



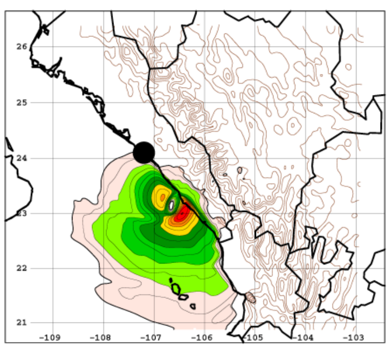
Initialization at 12UTC 03 September
Landfall
HWRF: Sep/04/15 (83 kt, 970mb)
NHC: Sep/04/21 (70 kt, 973mb)

Paul (2006)



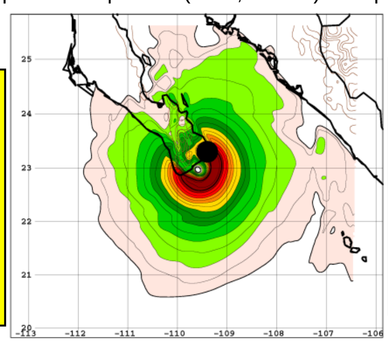
Initialization at 12UTC 23 October
Landfall
HWRF: Oct/25/15 (50 kt, 999mb)
NHC: Oct/25/04 (25 kt, 1007mb)

Lane (2006)



Initialization at 18UTC 15 September
Landfall
HWRF: Sep/16/06 (88 kt, 974mb)

John (2006)

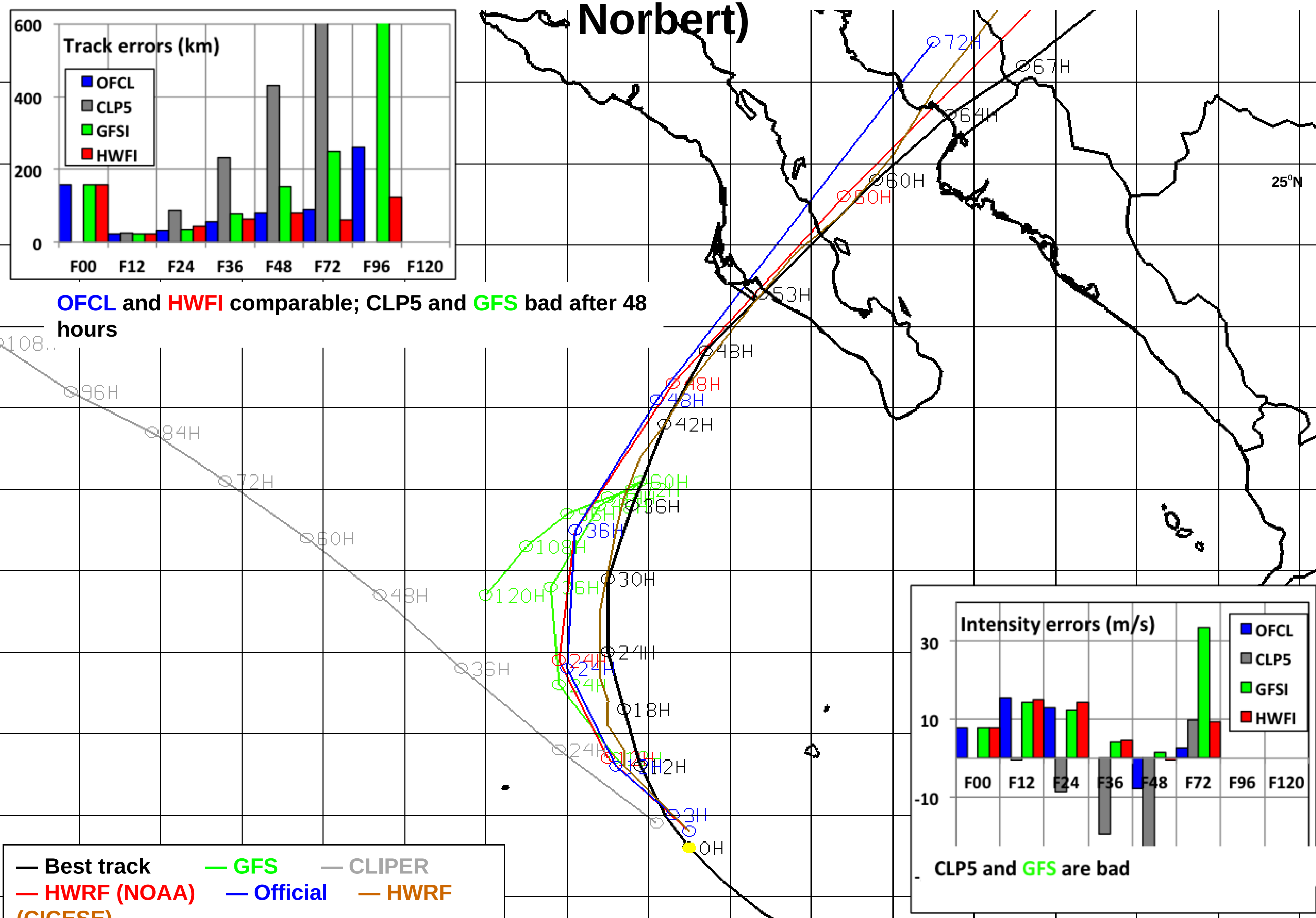


Initialization at 00UTC 01 September
Landfall
HWRF: Sep/01/15 (106 kt, 949mb)

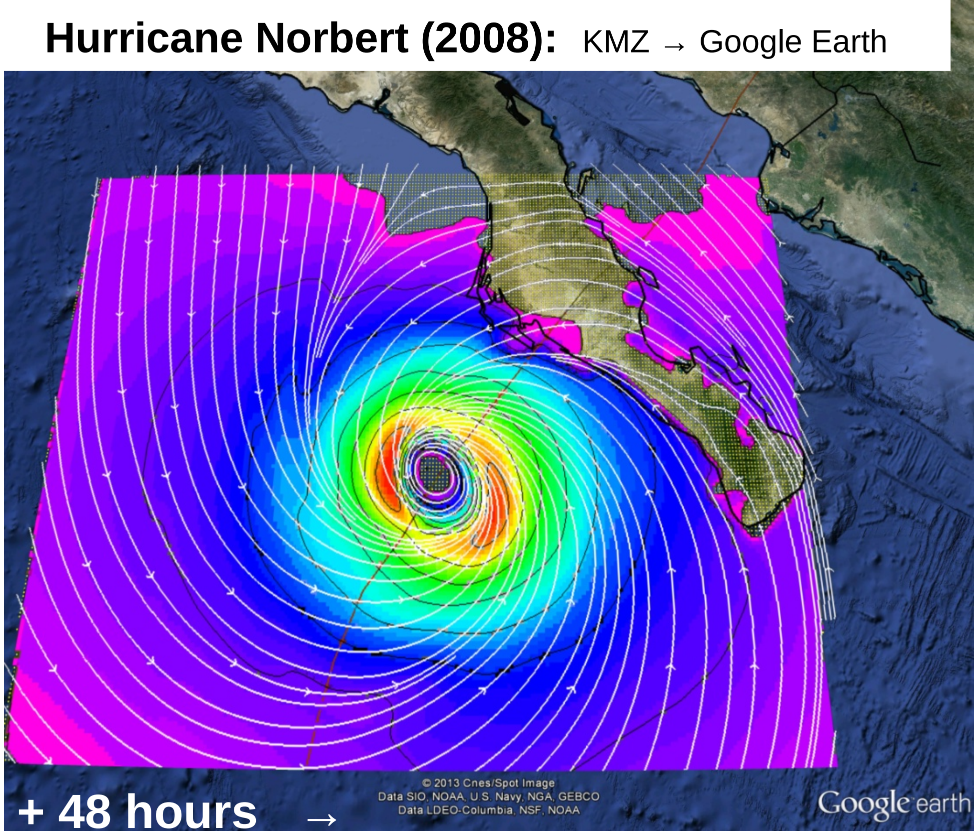
This is a sample of wind field displays at landfall for 8 case-studies performed at CICESE.

They made landfall, between 2006 and 2010, over Baja California or the mainland (Sinaloa, Sonora, Nayarit)

Forecasts initialized at 12UTC 09 October, 2008 (Hurricane Norbert)

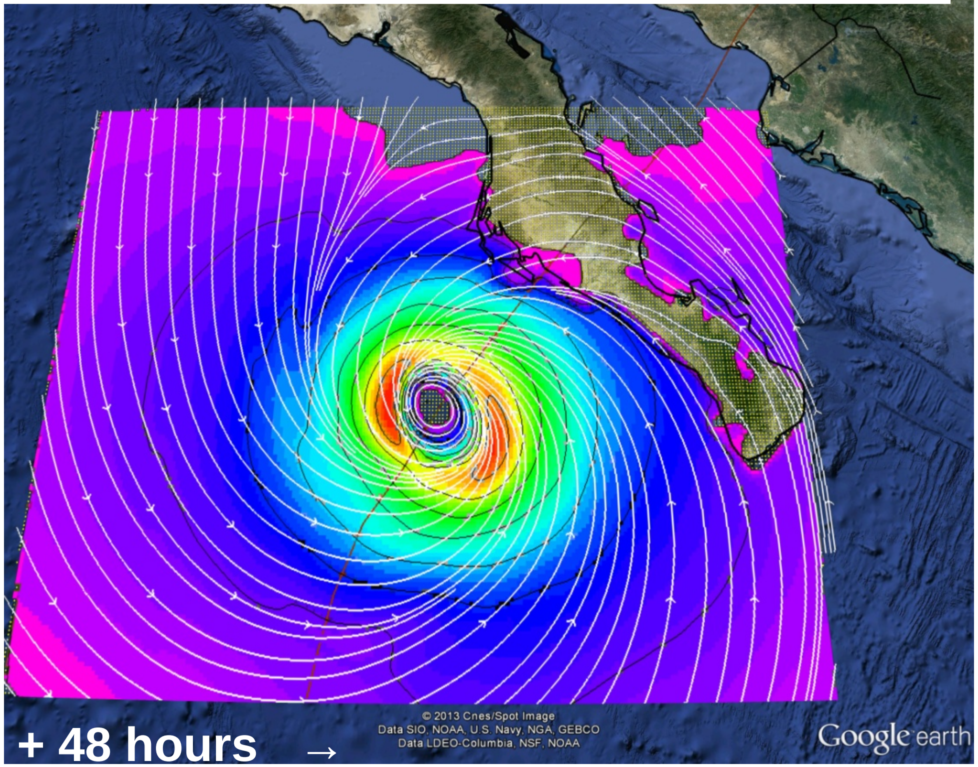


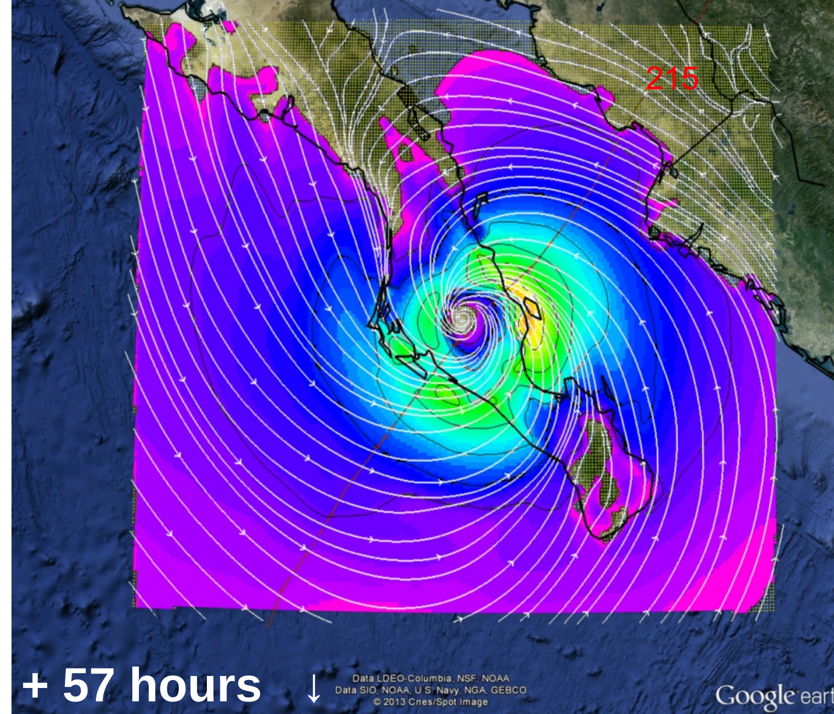
Hurricane Norbert (2008): KMZ → Google Earth

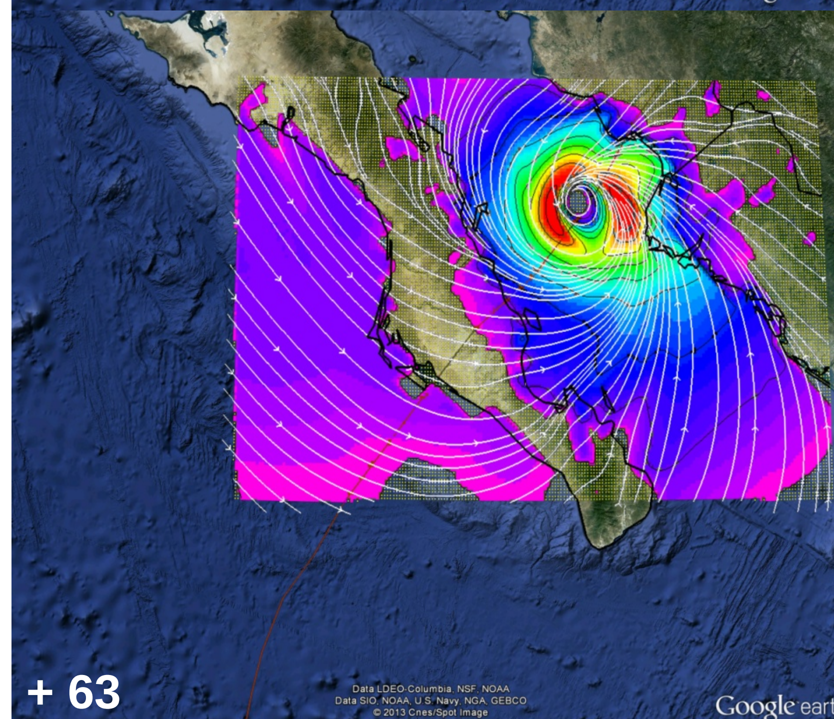


- The (model) vortex approaches Baja California, makes landfall, and emerges over the Gulf of California
- Note weakening at landfall and (re)intensification over the gulf
- White represents streamlines and shading are isotachs (wind speed) at the model lowest level (10 meters).

Hurricane Norbert (2008): KMZ → Google Earth

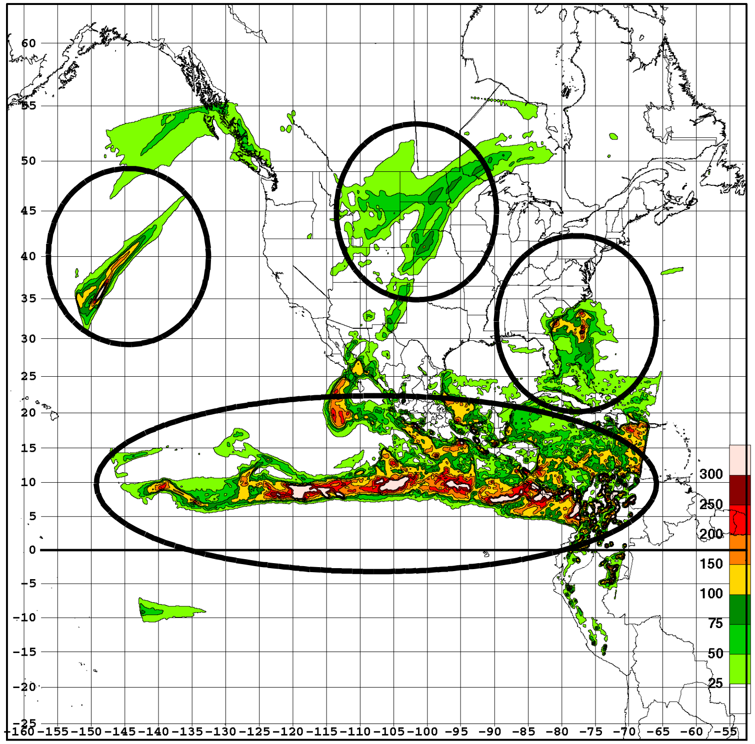


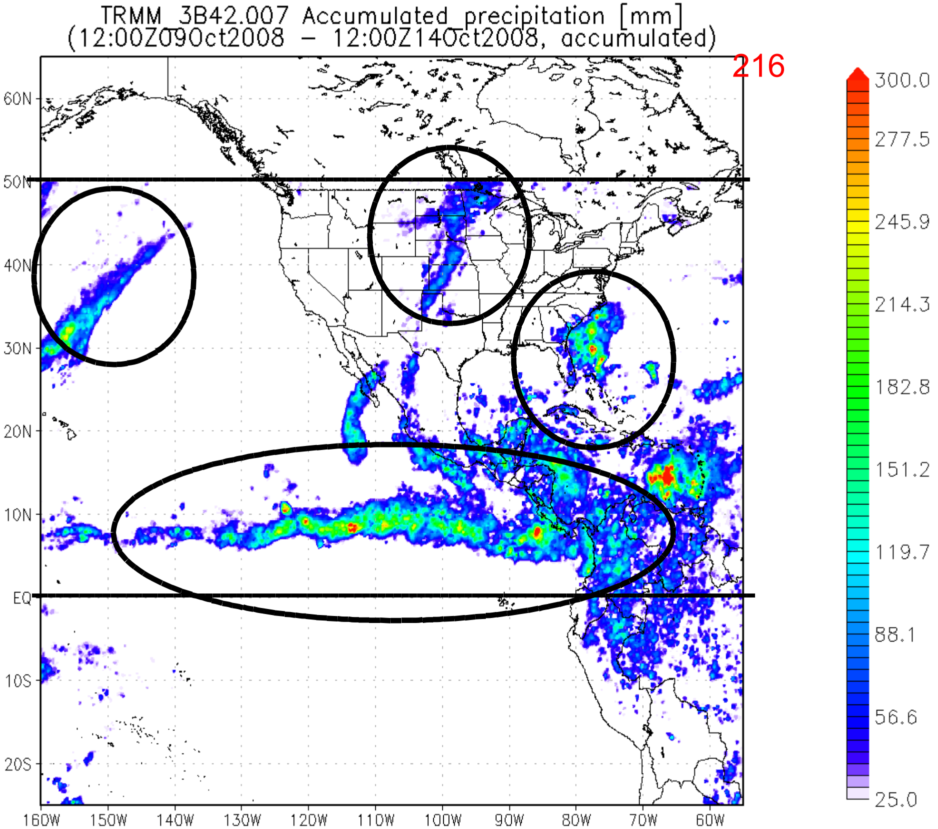




- The (model) vortex approaches Baja California, makes landfall, and emerges over the Gulf of California
- Note weakening at landfall and (re)intensification over the gulf
- White represents streamlines and shading are isotachs (wind speed) at the model lowest level (10 meters).

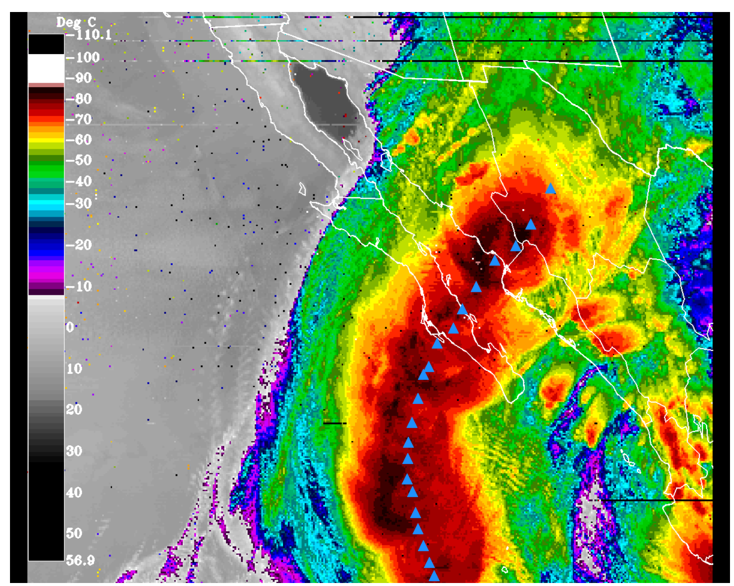
HWRF, acummulated rain 120h
12GMT 9 oct – 12GMT 14 oct





Rainfall rate is an important component of the model output.

It is a difficult field to predict and one of the most important elements to be used by emergency managers (and general public) in the affected regions



GOES-11 infrared imagery Oct/9-11

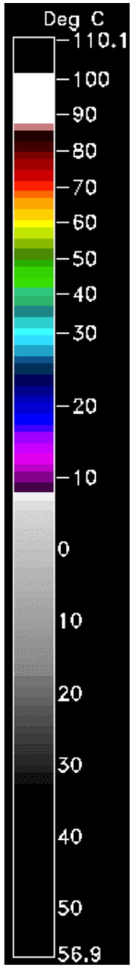
CICESE provides technical support to local emergency managers ←

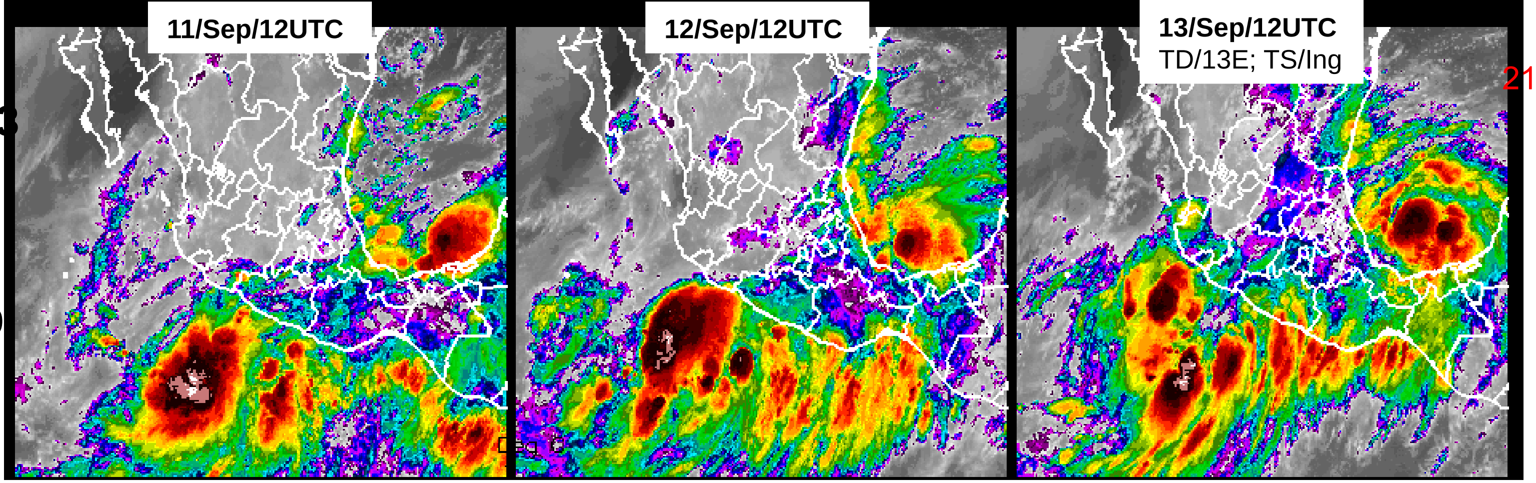
Hurricane Norbert (Oct, 2008) ↑

Tropical storm Lorena (Sep, 2013)

NOAA GOES-13 infrared imagery

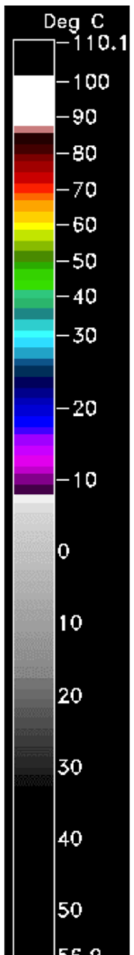
Sep/11-19 2013

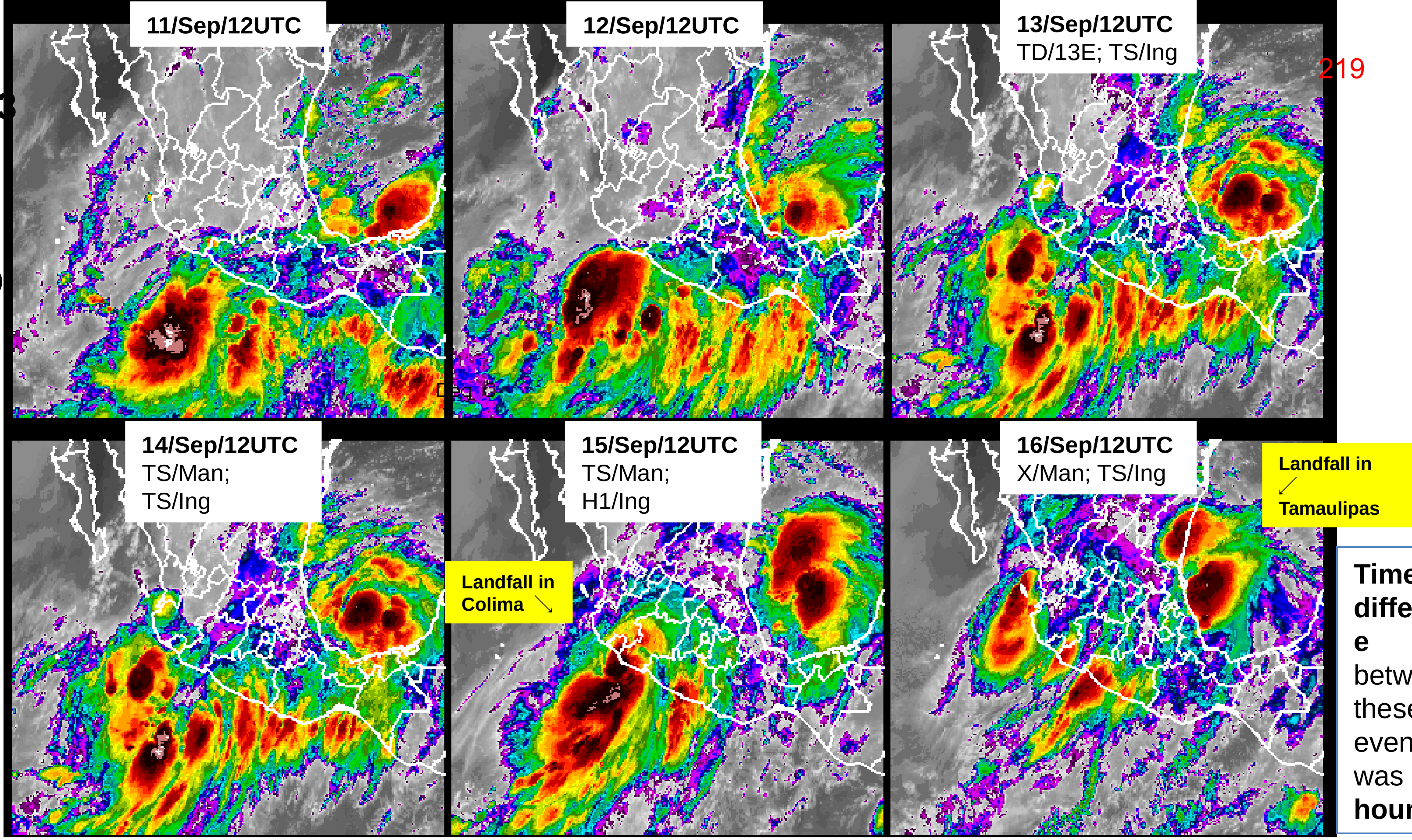




NOAA GOES-13 infrared imagery

Sep/11-19 2013

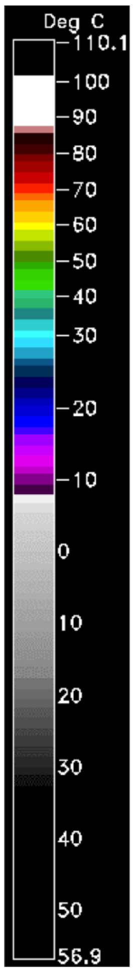


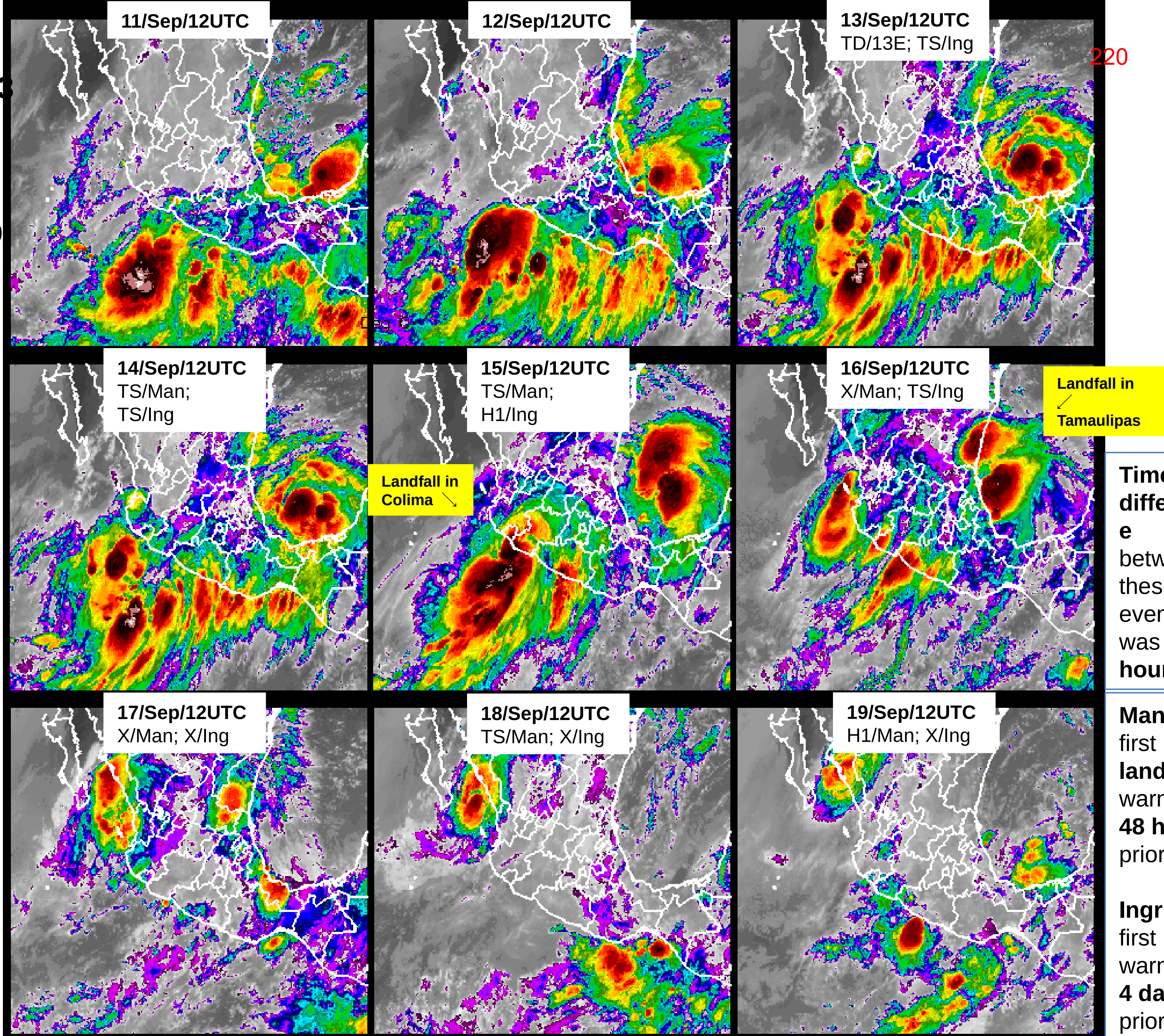


Time difference between these events was **24 hours**

NOAA GOES-13 infrared imagery

Sep/11-19 2013





Time difference between these events was **24 hours**

Manuel, first **landfall** warning, **48 hours** prior to…

Ingrid, first warning **4 days** prior to

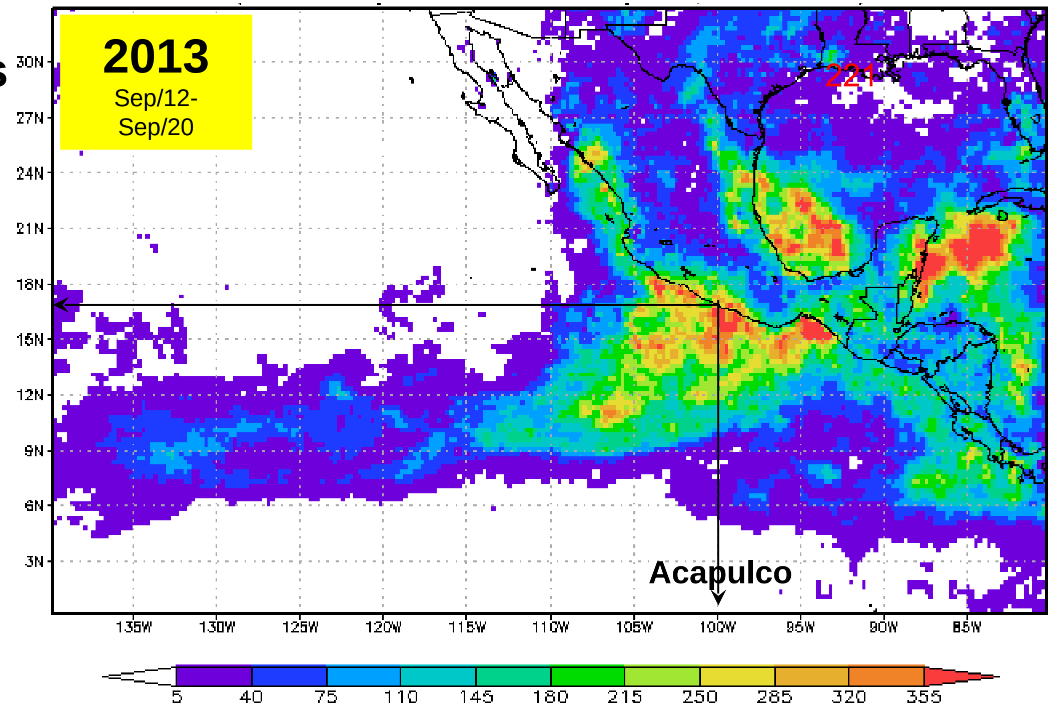
Accumulated rainfall (mm) estimates from TRMM

•In general, the accumulation pattern from September 12-20 shows:

1) distinct maxima along and off the Pacific coast (T.C. Manuel)

2) A secondary maximum in the eastern Gulf of Mexico (T.C. Ingrid)

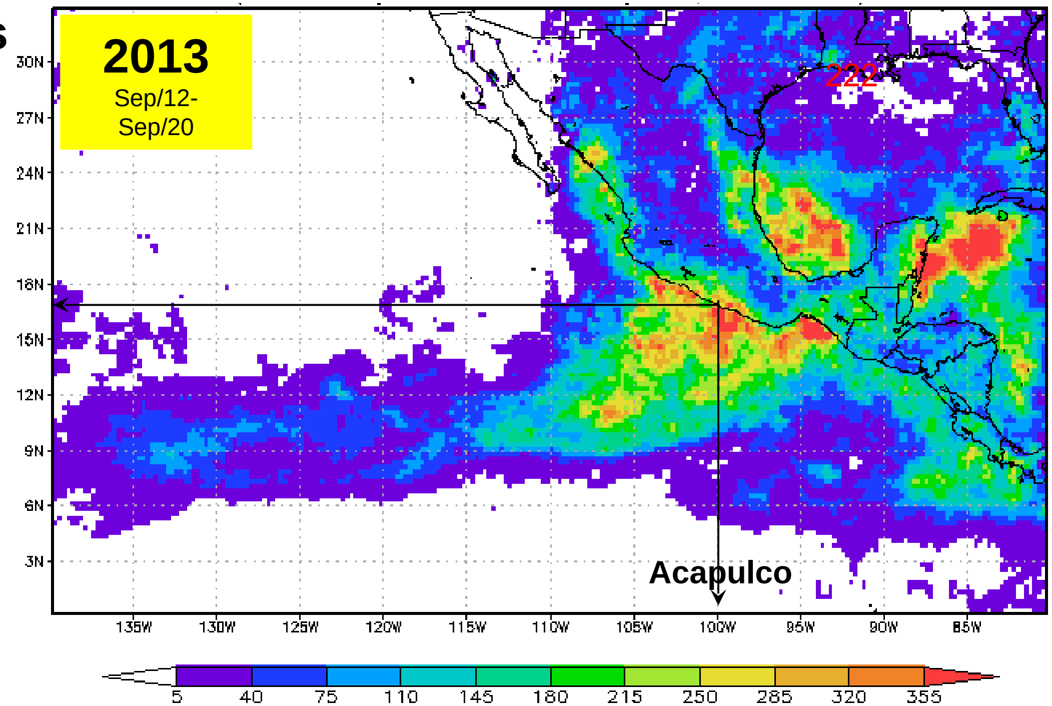
3) Another maximum over the southern Gulf of California and in Sinaloa (T.C. Manuel)



Maximum, actual accumulations close to 1,000 mm (40 inches) at weather stations in Guerrero.

Also, there are estimates of > 150 deaths and over 1.5 million affected.

Accumulated rainfall (mm) estimates from TRMM



•In general, the accumulation pattern from September 12-20 shows:

1) distinct maxima along and off the Pacific coast (T.C. Manuel)

2) A secondary maximum in the eastern Gulf of Mexico (T.C. Ingrid)

3) Another maximum over the southern Gulf of California and in Sinaloa (T.C. Manuel)

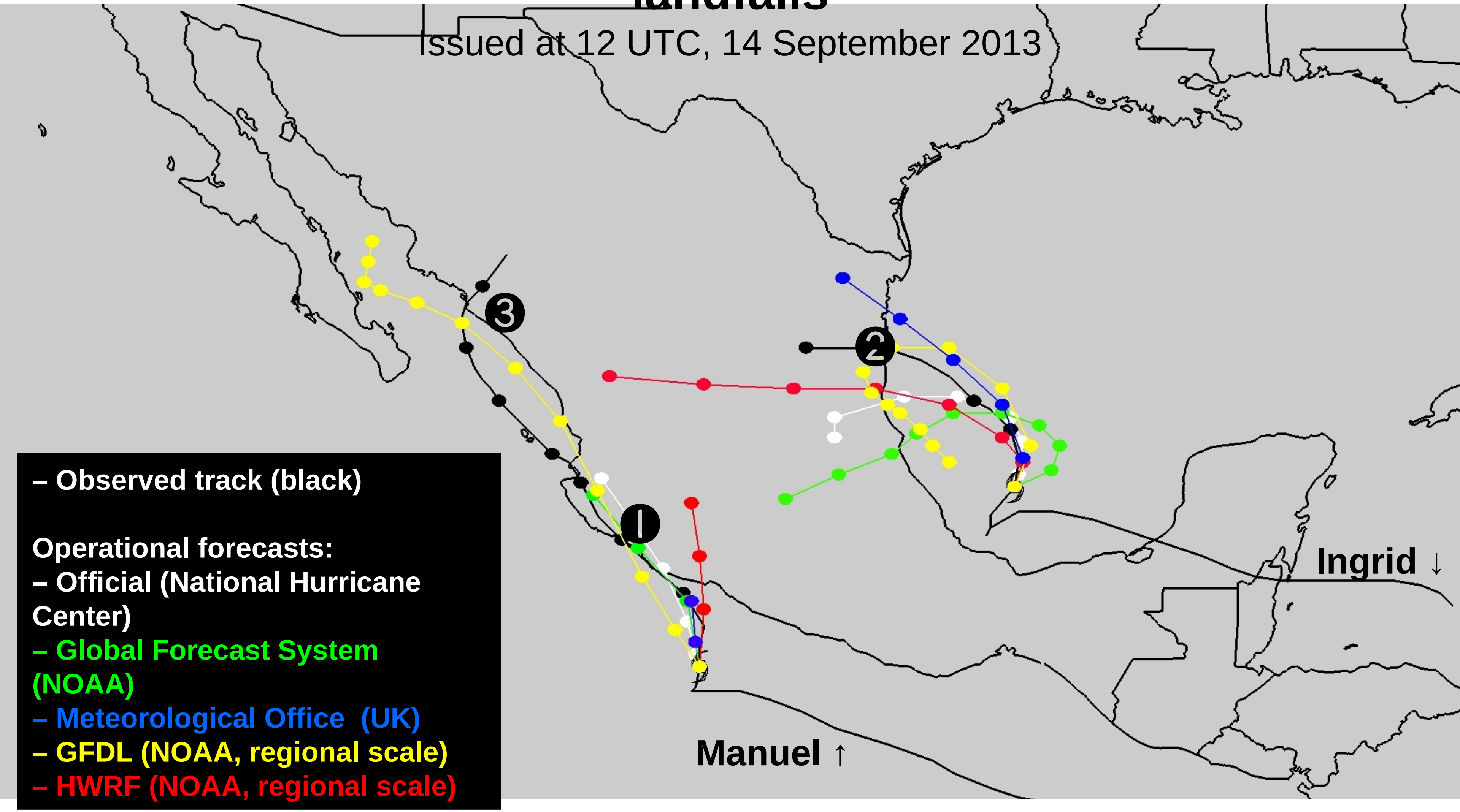
Maximum, actual accumulations close to 1,000 mm (40 inches) at weather stations in Guerrero.

Also, there are estimates of > 150 deaths and over 1.5 million affected.

COORDINACIÓN NACIONAL DE PROTECCIÓN CIVIL MÉXICO

SEGOB
SECRETARÍA DE GOBERNACIÓN

Track forecasts issued prior to Manuel (❶,❸) and Ingrid (❷) landfalls

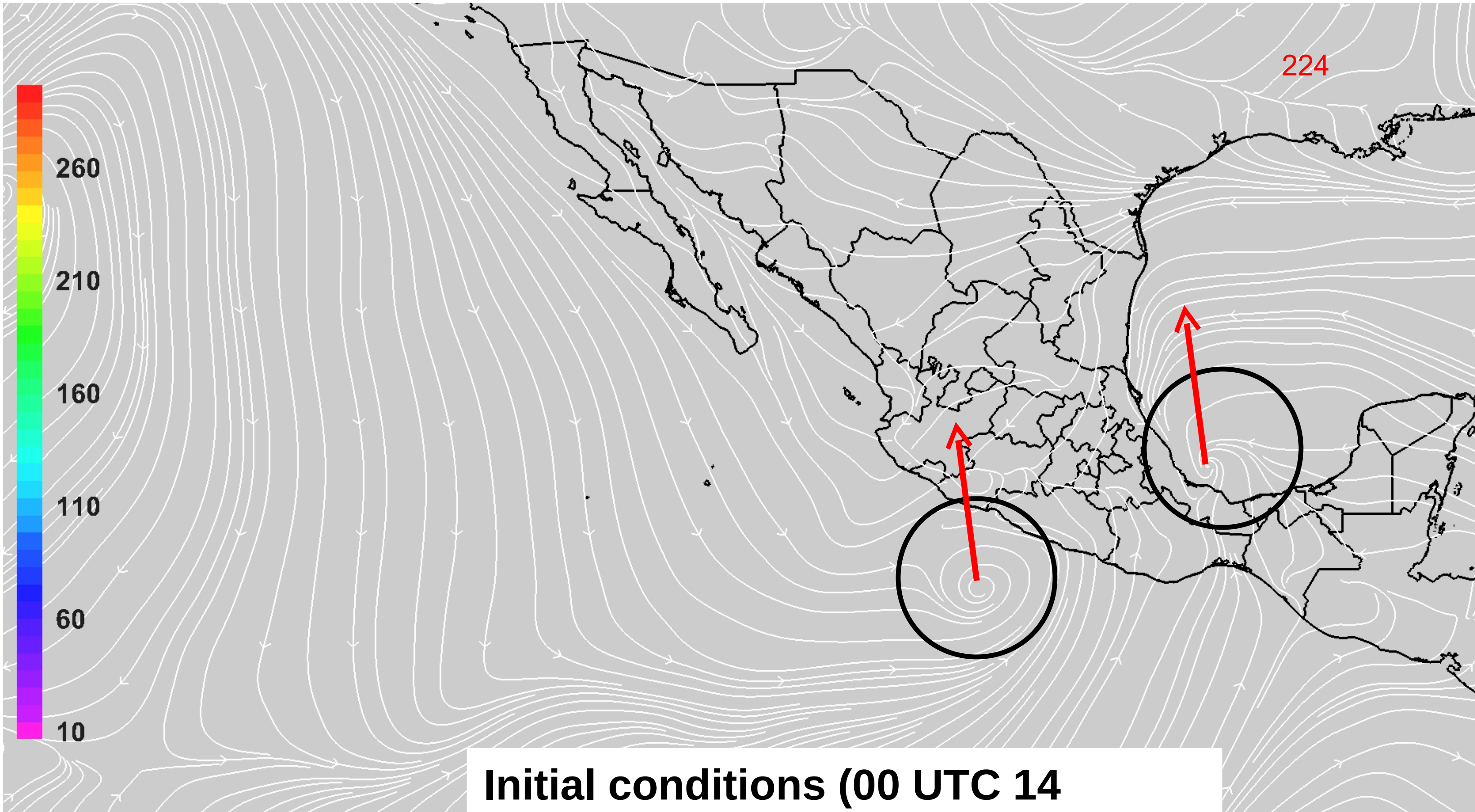


- The National Hurricane Center issued 23 official forecasts for Manuel and 20 for Ingrid
- Ingrid: all forecasts indicated landfall in northern Veracruz or Tamaulipas
- Manuel: first forecast period had landfall predictions in Guerrero/Michoacán
  second forecast period over northwestern Mexico

HWRF model at CICESE

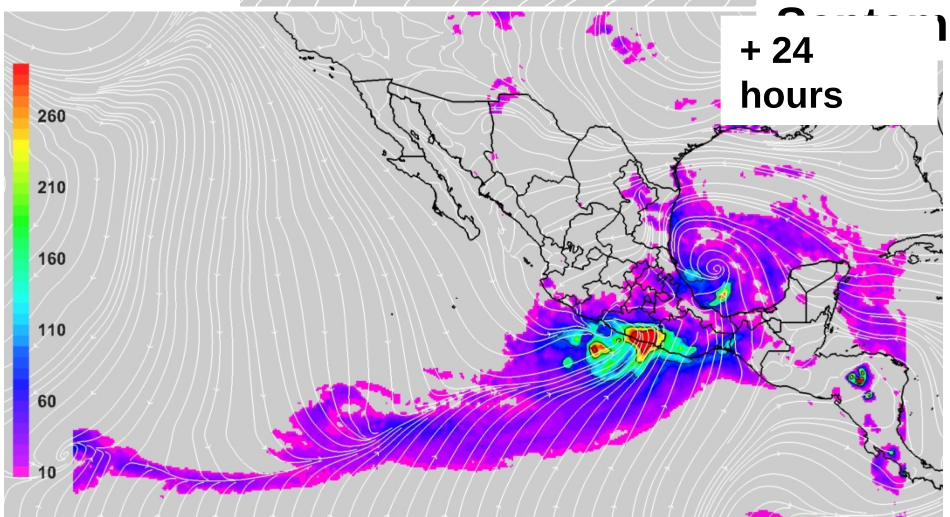
10-m streamlines

Accumulated rainfall (mm) from the initial time.

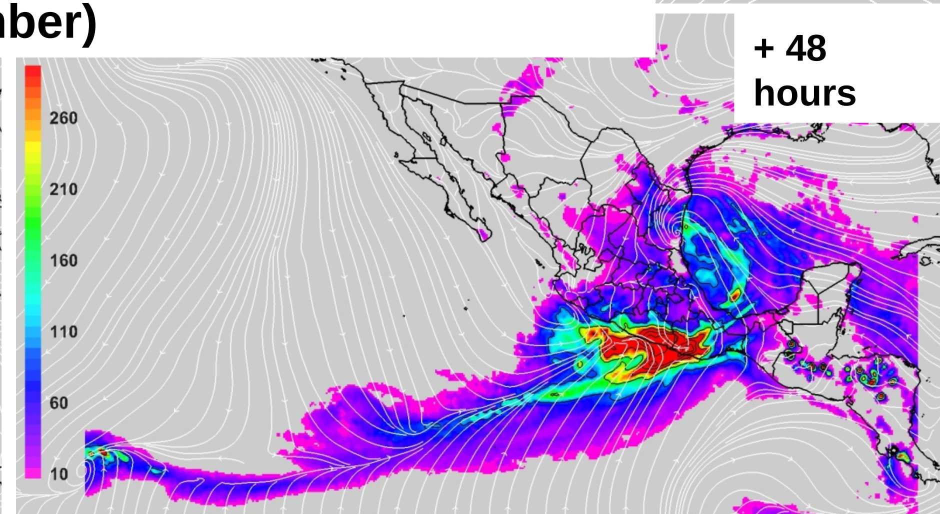


Initial conditions (00 UTC 14 September)

+ 24 hours

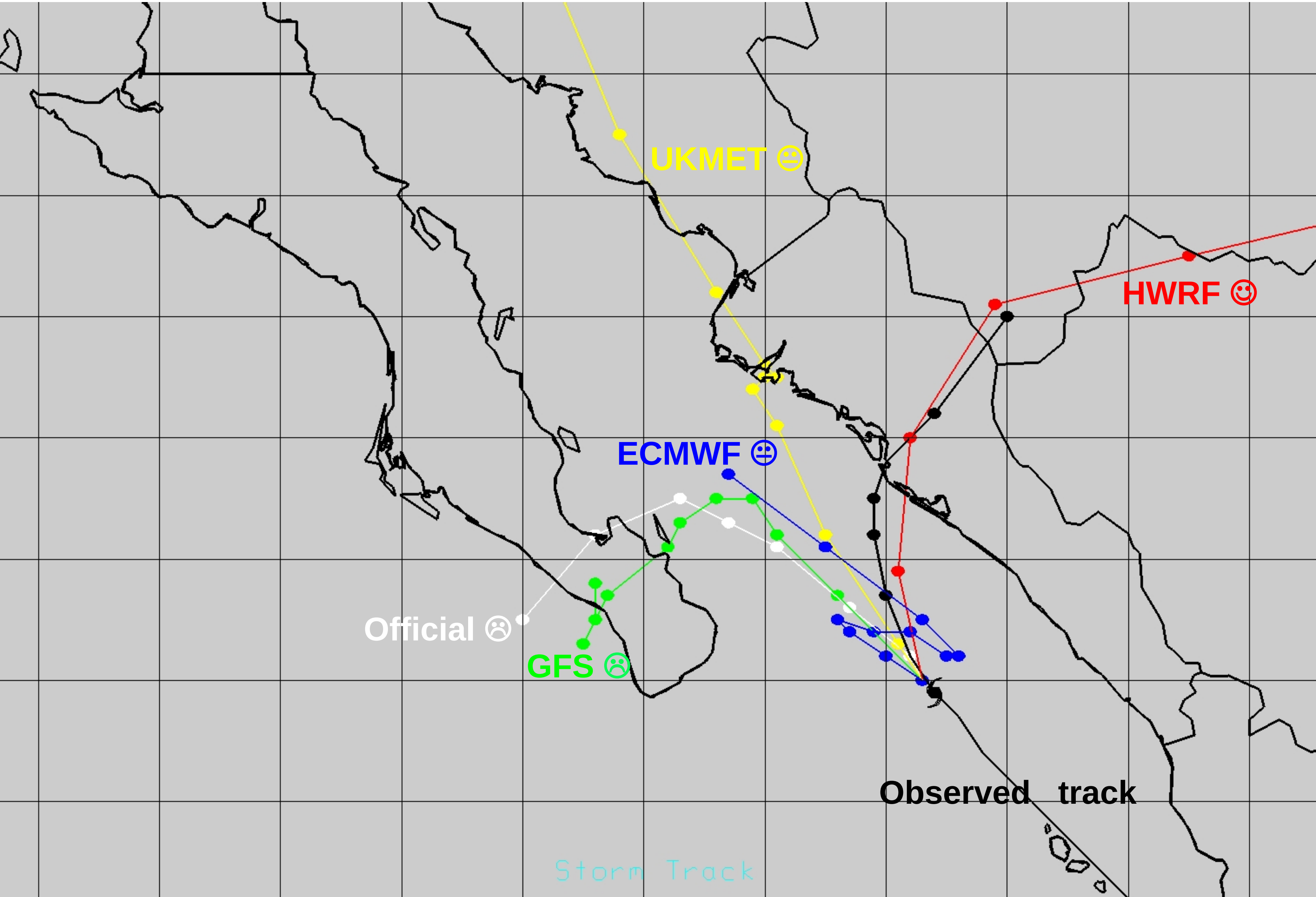


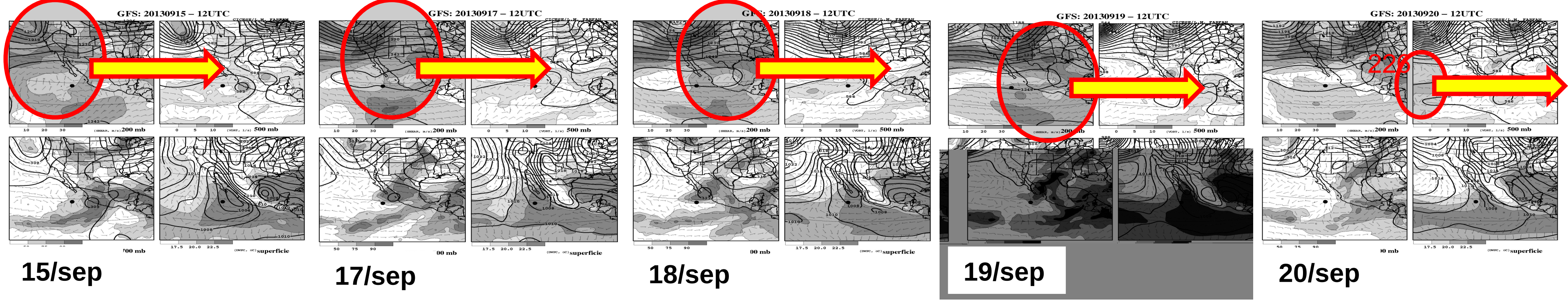
+ 48 hours



Track forecasts for Tropical Cyclone Manuel, 0600 UTC 18 September 2013

The official forecast (white line) suggested westward motion toward Baja California ☹



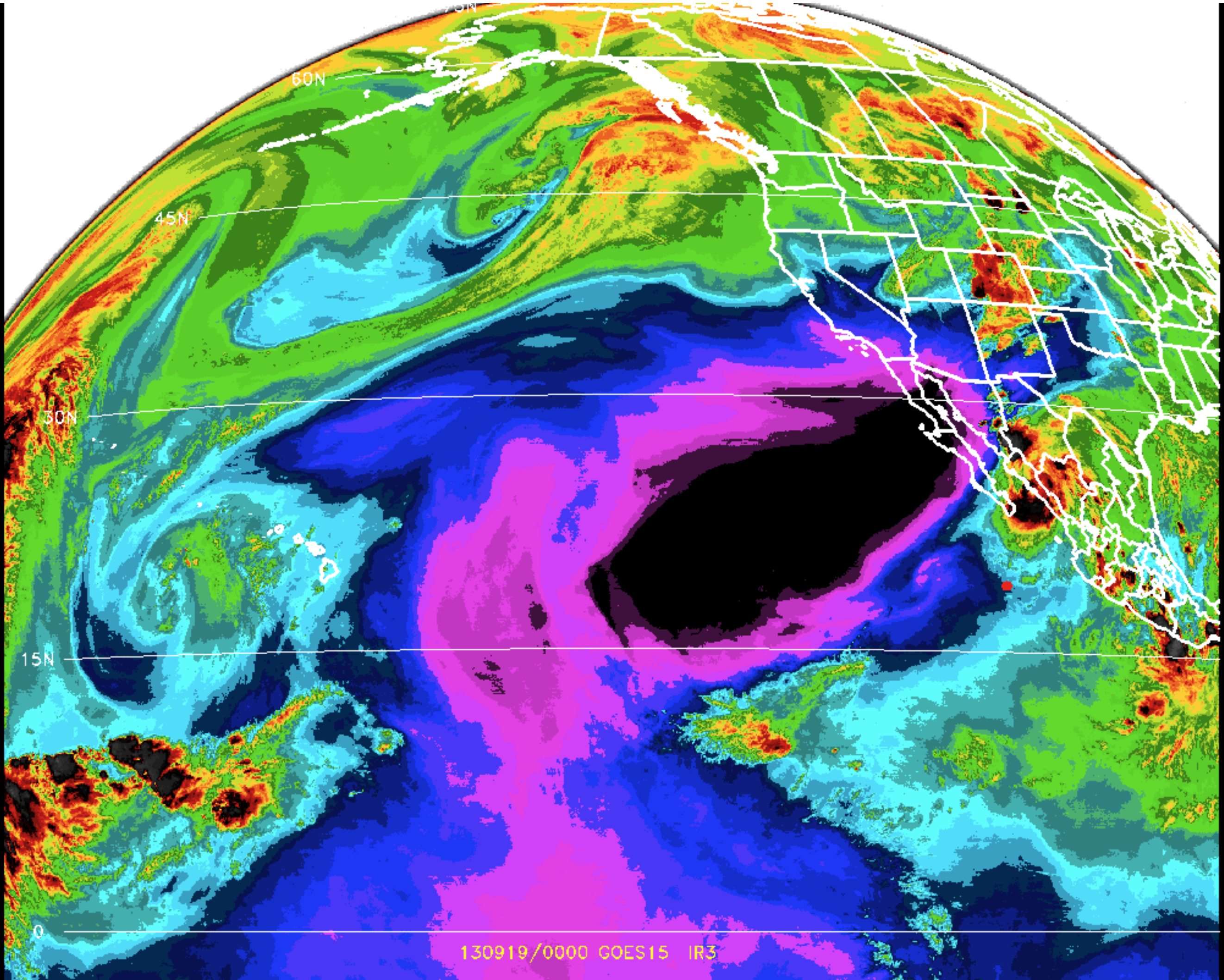


19/Sep 2013

NOAA's GOES-15 water vapor imagery

Black/pink= dry
Green= humid
Red= deep clouds

The development of mid-latitude systems may be the reason for Manuel's



Numerical Modeling for Tropical Cyclone Landfall in Mexico

Luis M. Farfán, Ismael Villanueva and Julián J. Delgado

- Tropical cyclones are atmospheric systems that may become a source of damage to the population, property and environment

- The application of numerical models has served to issue advance warnings of landfall along coastal areas

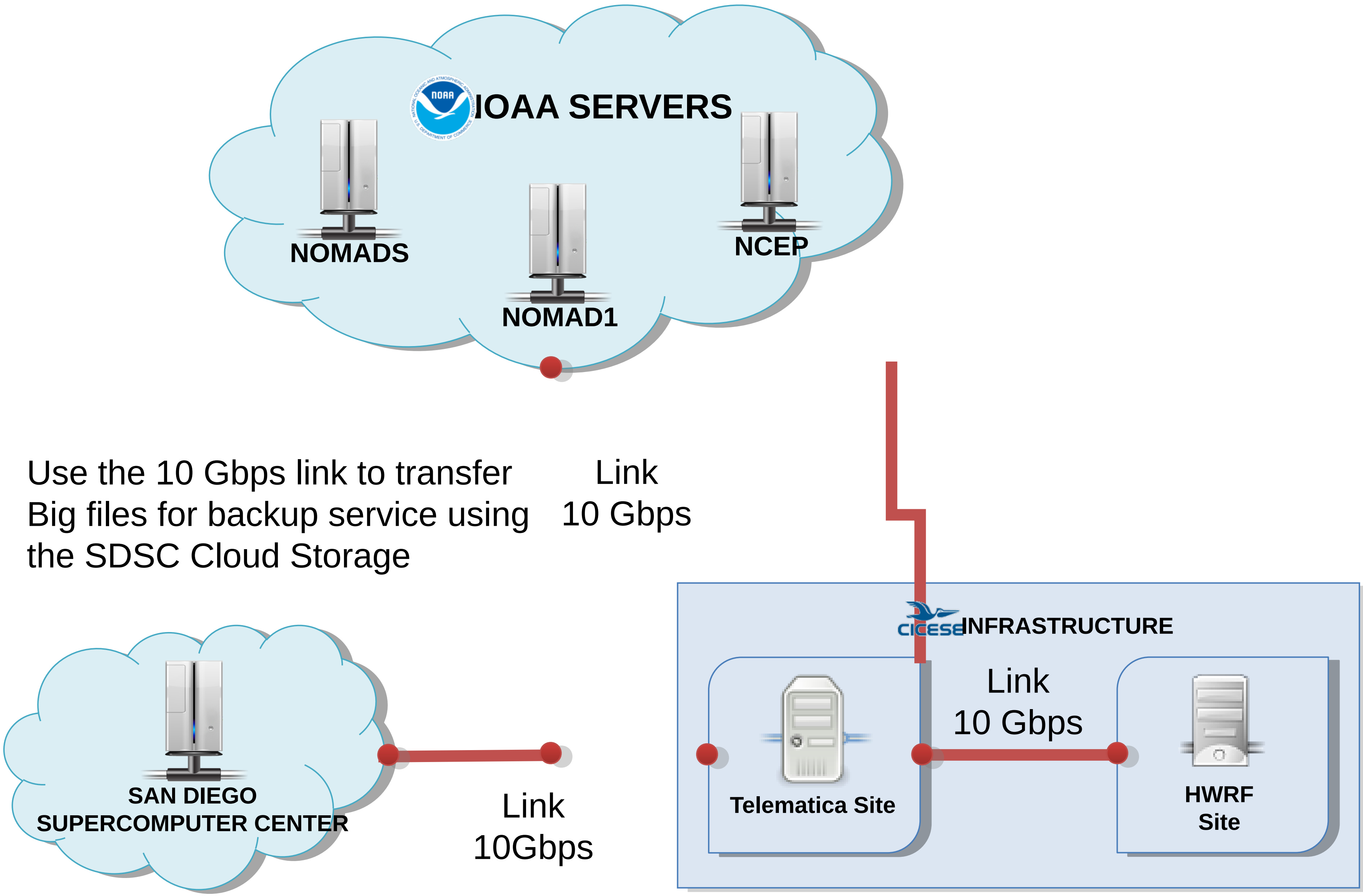
- HWRF is a model installed at CICESE’s Computing Department and is able to perform high-resolution forecasts for cyclones in the eastern Pacific Ocean

- By using HWRF, We are studying tropical cyclones that made landfall in northwestern Mexico during the period 2006-2010

- Future work includes HWRF automated applications on a real-time basis and use the WRF model for 24-72 hr predictions for along Baja California.

Acknowledgements:
- Partial funding from CONACYT, REDESClim, and IAI
- Technical support from CICESE’s Dirección de Telemática.

FUTURE INFRASTRUCTURE



Transport and dispersion of hydrocarbons in estuaries of the Gulf of Mexico:

an example of a proposed data-intensive application

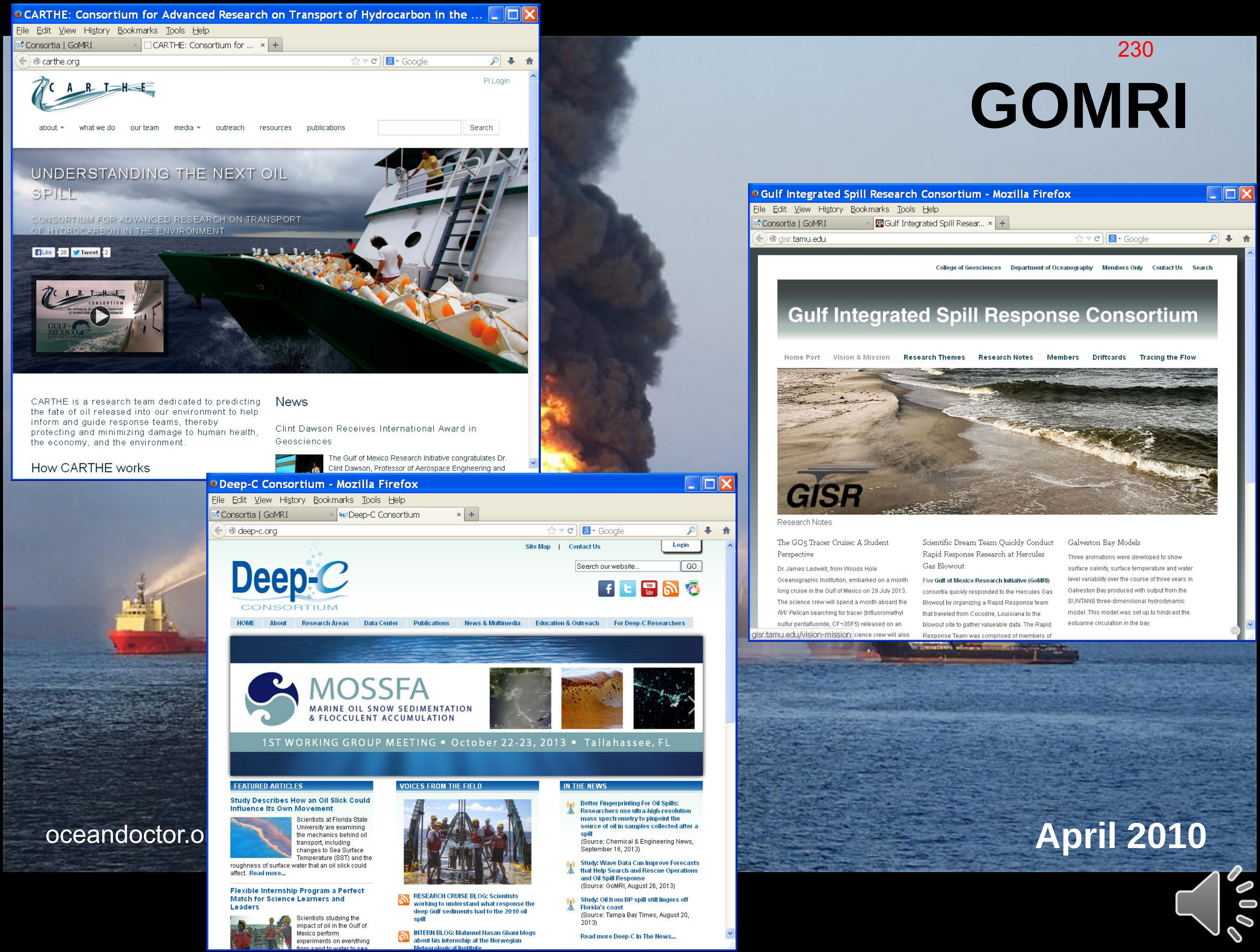
Arnoldo Valle-Levinson

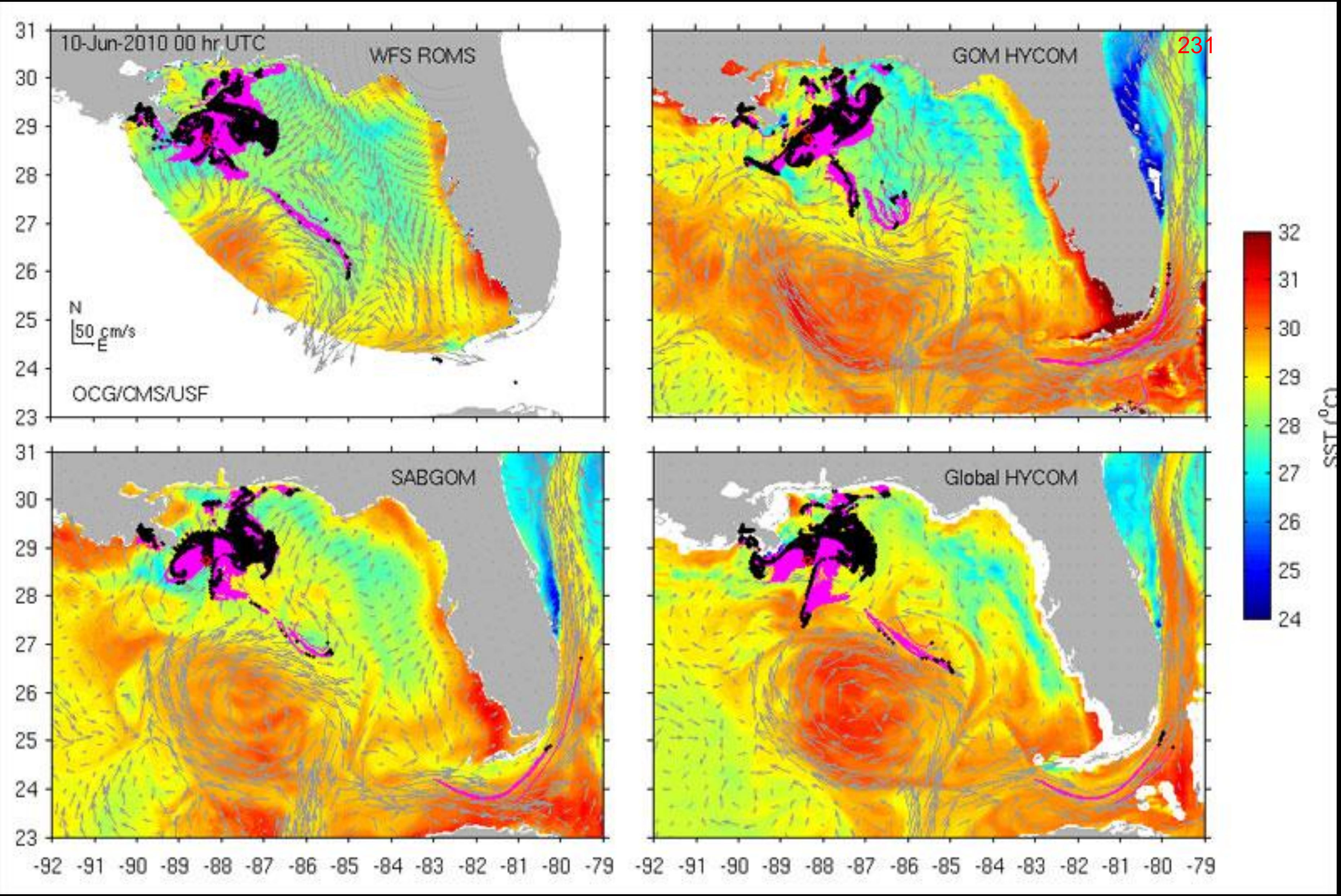
Professor

University of Florida

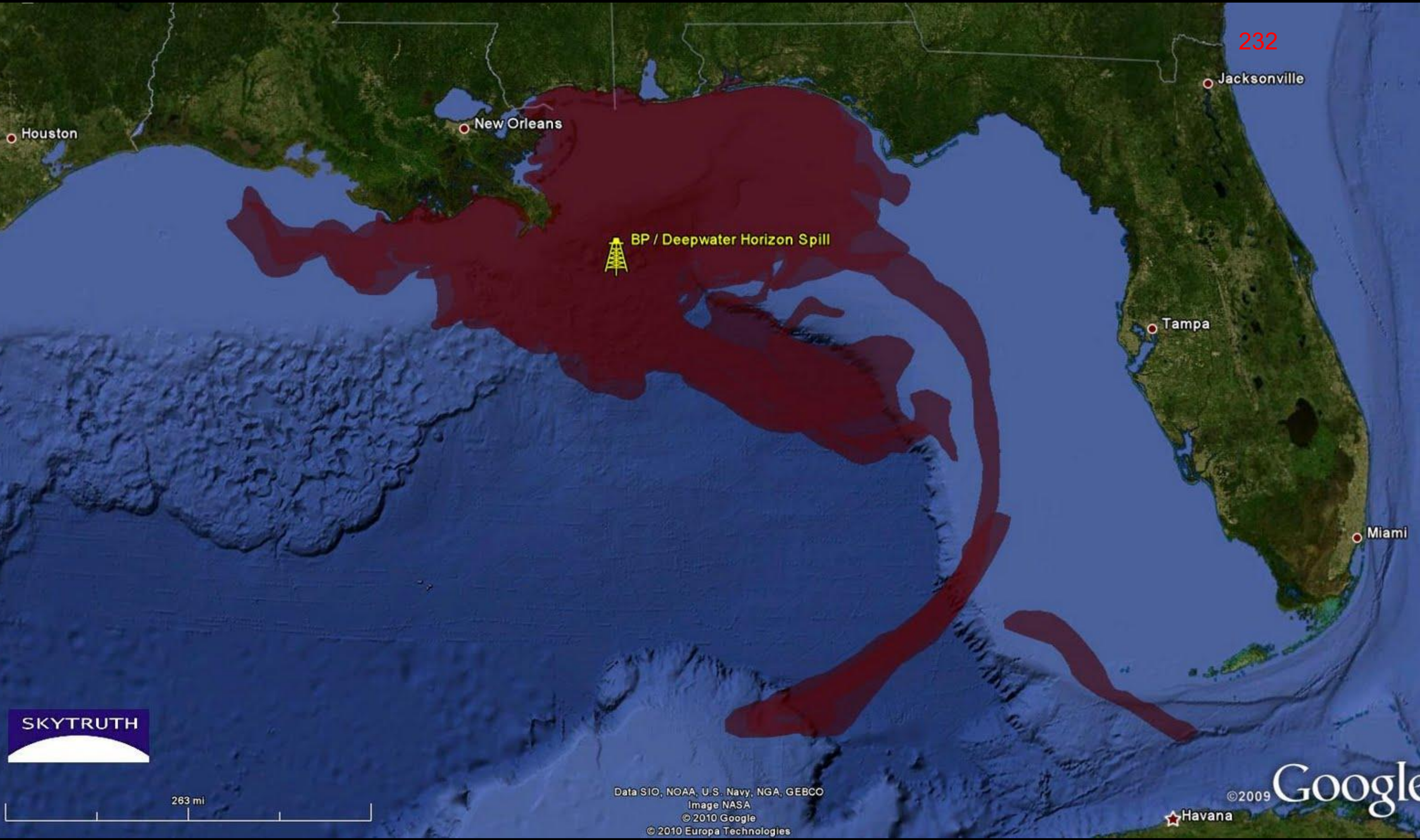
arnoldo@ufl.edu

Tel. 352-392-9537 x 1479


GOMRI
CARTHE: Consortium for Advanced Research on Transport of Hydrocarbon in the ...
carthe.org
UNDERSTANDING THE NEXT OIL SPILL
CONSORTIUM FOR ADVANCED RESEARCH ON TRANSPORT OF HYDROCARBON IN THE ENVIRONMENT
CARTHE is a research team dedicated to predicting the fate of oil released into our environment to help inform and guide response teams, thereby protecting and minimizing damage to human health, the economy, and the environment.
How CARTHE works
News
Clint Dawson Receives International Award in Geosciences
Gulf Integrated Spill Research Consortium - Mozilla Firefox
gisr.tamu.edu
Gulf Integrated Spill Response Consortium
Home Port
Vision & Mission
Research Themes
Research Notes
Members
Driftcards
Tracing the Flow
GISR
Research Notes
The GO5 Tracer Cruise: A Student Perspective
Scientific Dream Team Quickly Conduct Rapid Response Research at Hercules Gas Blowout
Galveston Bay Models
Deep-C Consortium - Mozilla Firefox
deep-c.org
Deep-C CONSORTIUM
MOSSFA
MARINE OIL SNOW SEDIMENTATION & FLOCCULENT ACCUMULATION
1ST WORKING GROUP MEETING • October 22-23, 2013 • Tallahassee, FL
FEATURED ARTICLES
Study Describes How an Oil Slick Could Influence Its Own Movement
Flexible Internship Program a Perfect Match for Science Learners and Leaders
VOICES FROM THE FIELD
IN THE NEWS
Better Fingerprinting For Oil Spills: Researchers use ultra-high-resolution mass spectrometry to pinpoint the source of oil in samples collected after a spill
Study: Wave Data Can Improve Forecasts that Help Search and Rescue Operations and Oil Spill Response
Study: Oil from BP spill still lingers off Florida's coast
oceandoctor.o
April 2010

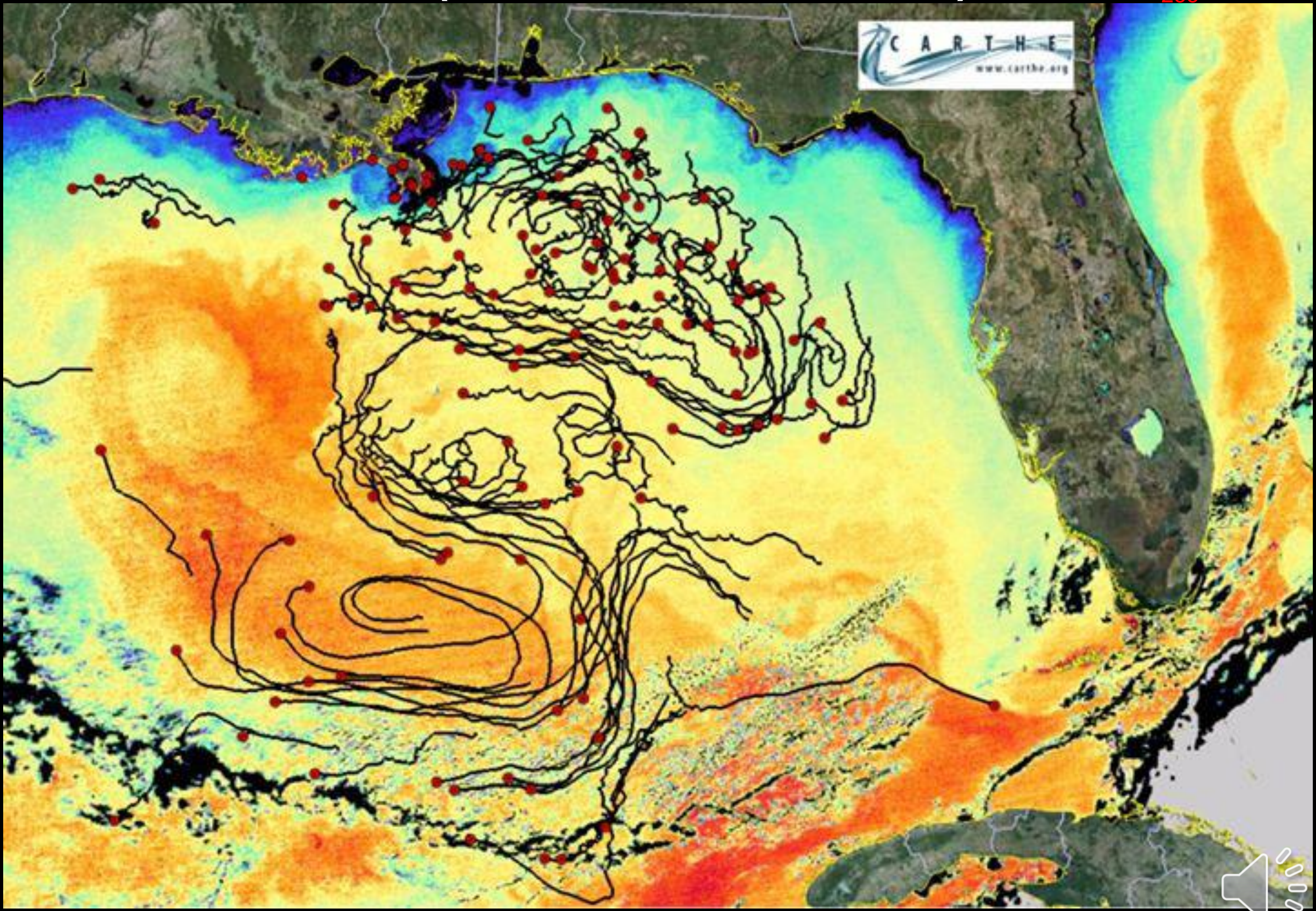


Yonggang Liu, University of South Florida



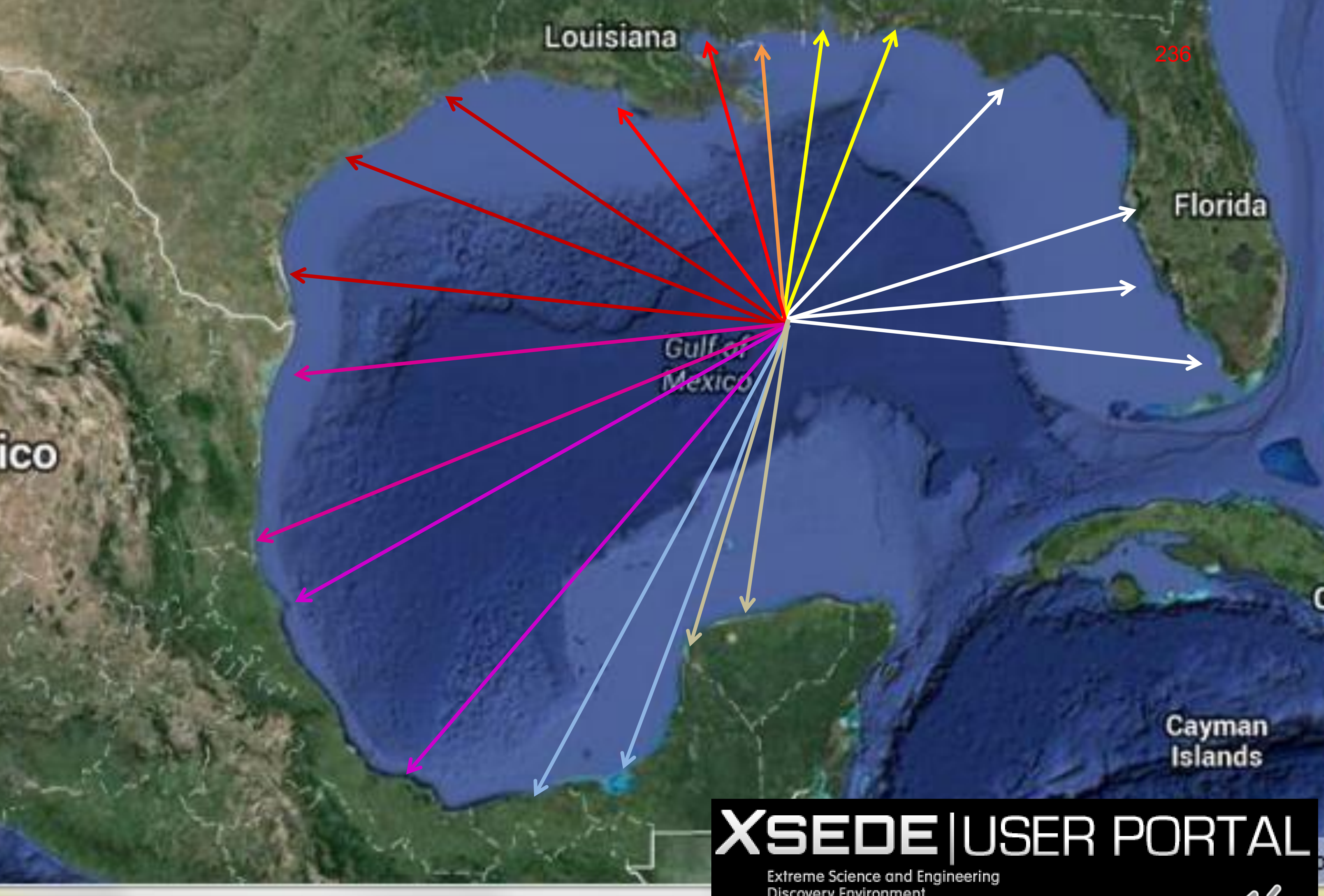
Mostly Open Ocean
Mainly US

Submesoscale (few km to m; hrs to seconds)motions



ocean.si.edu

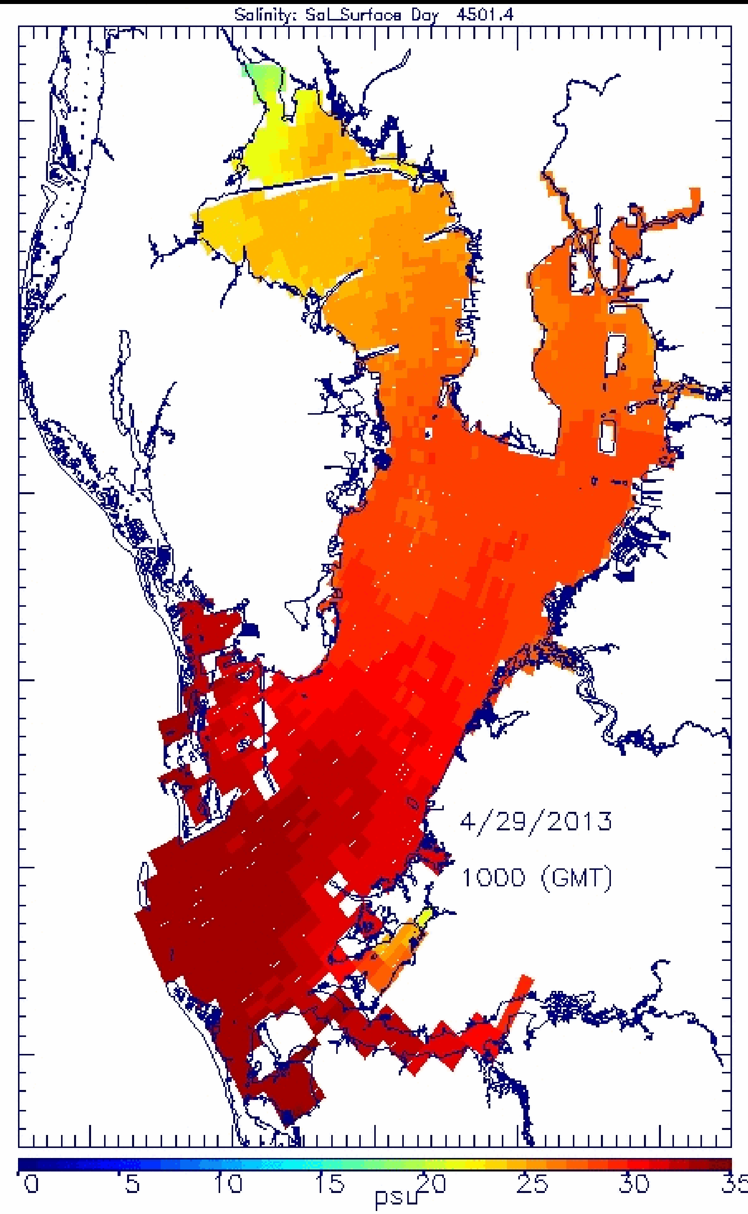
knowledge.allianz.com



XSEDE | USER PORTAL
Extreme Science and Engineering
Discovery Environment

Nancy Wilkins-Diehr – CICESE - CUDI

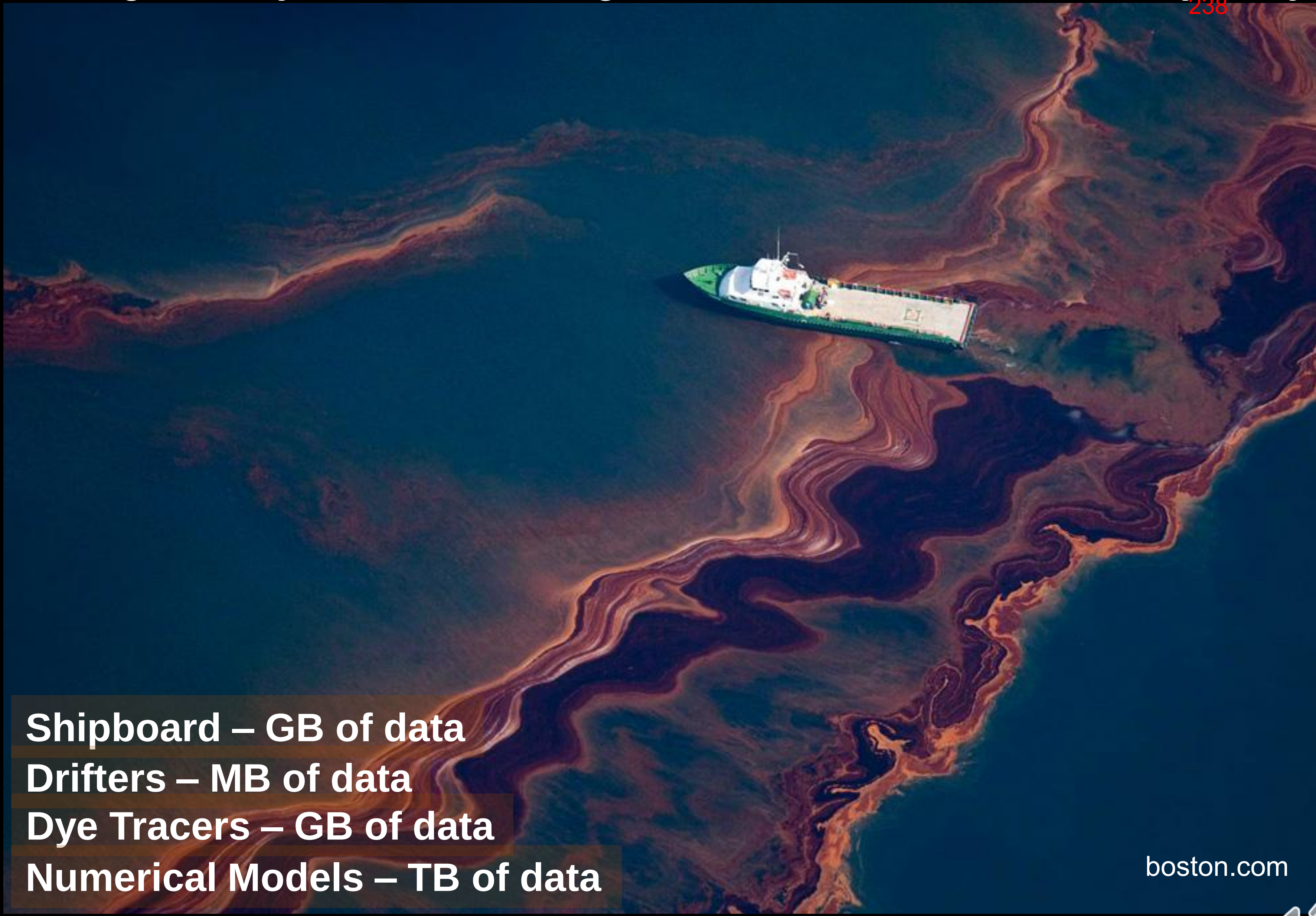
1. Plume Fronts – Freshwater Discharge



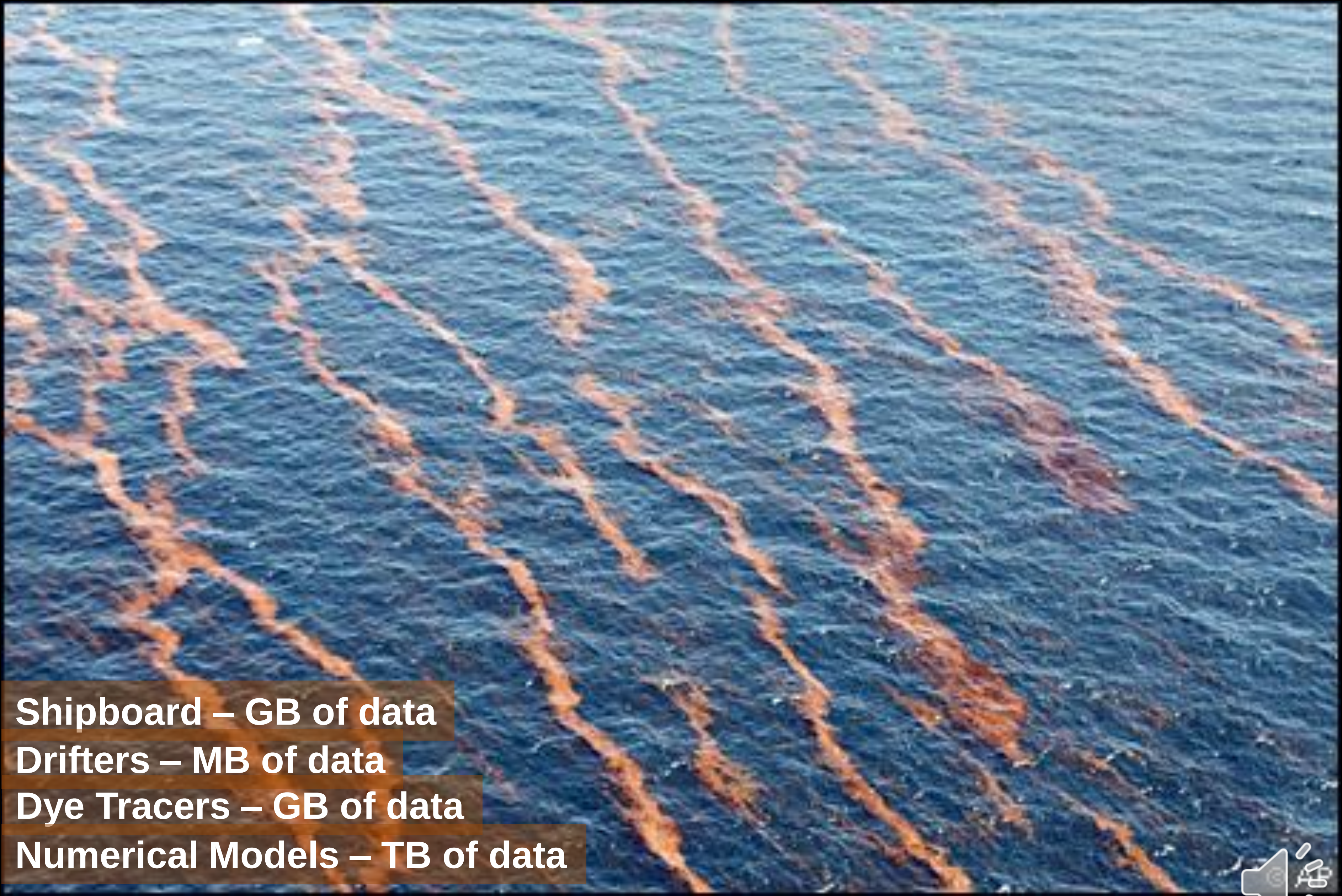
http://ompl.marine.usf.edu

Shipboard – GB of data
Drifters – MB of data
Dye Tracers – GB of data
Numerical Models – TB of data

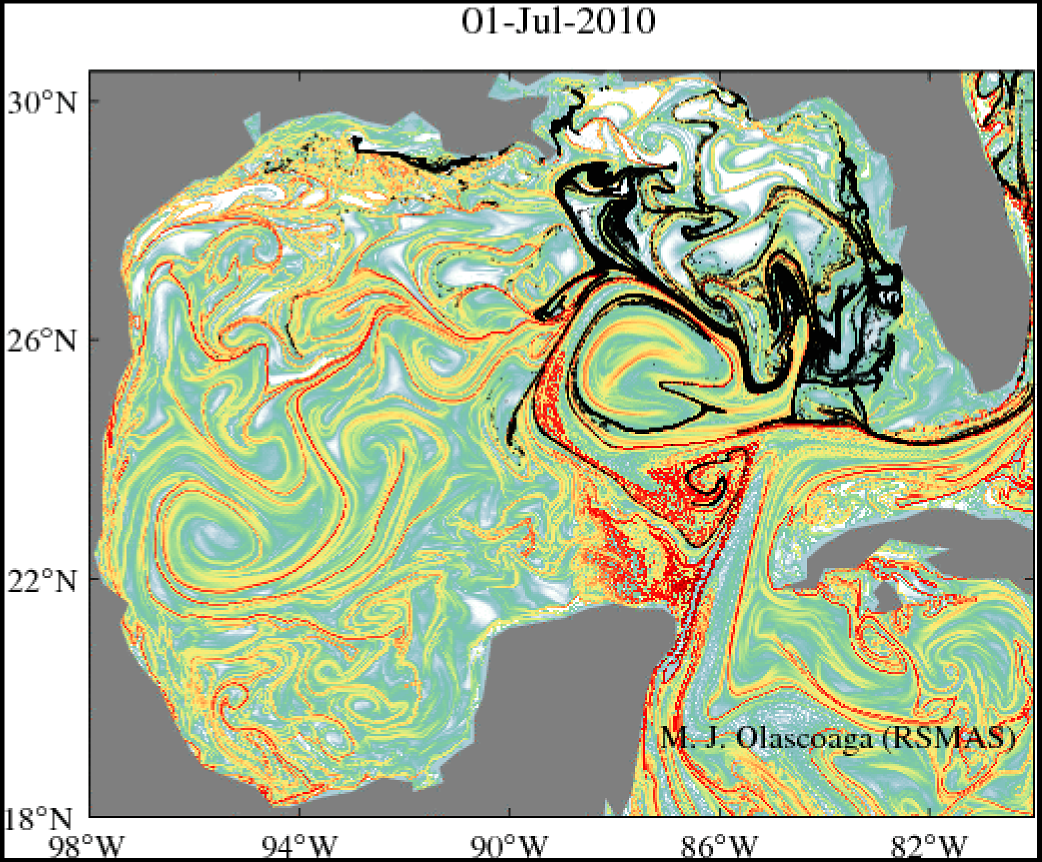
2. Along-estuary and Tidal Mixing Fronts – Tidal currents - Bathymetry



3. Langmuir Cells – Wind Driven



FOUR APPROACHES WILL PROVIDE RELIABLE VELOCTY FIELDS FOR LAGRANGIAN COHERENT STRUCTURES



1. Challenges for network enabled cross-border research and education collaborations

Transfer of Information

Data repository/storage

Computational efficiency/expediency

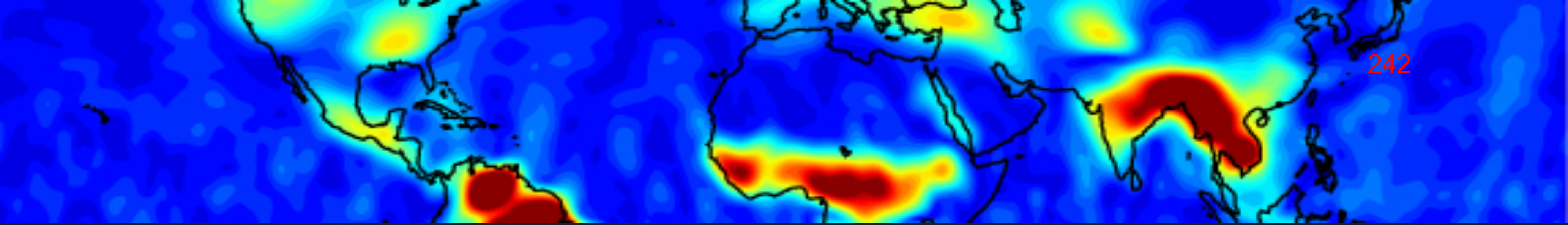
Same as above

Plus educational activities

Arnoldo Valle-Levinson

arnoldo@ufl.edu

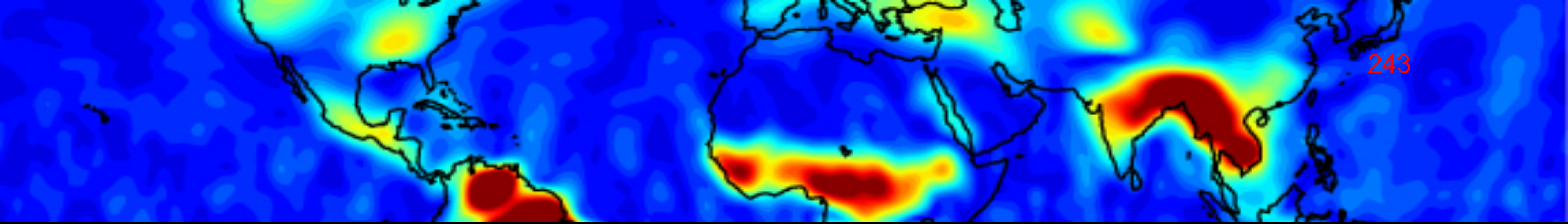
It has already been successful by providing
a platform to present and exchange
ideas on international collaborations



AMLIGHT, Simulation Datasets, and Global Data Sharing

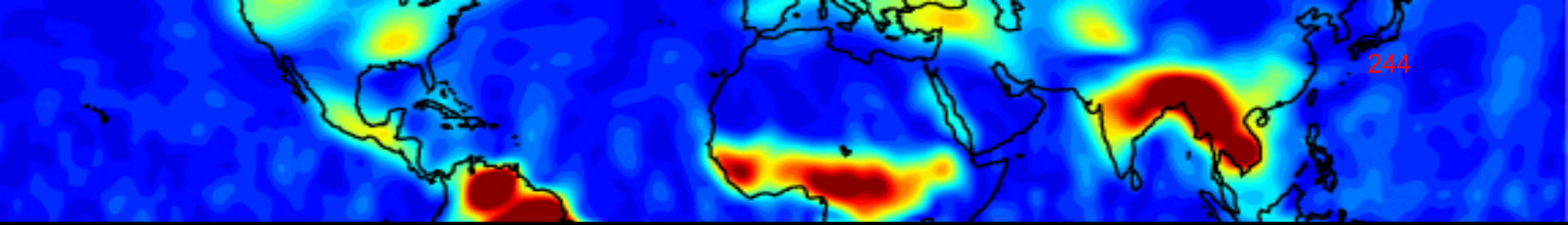
Jean-Bernard Minster[1,2,4,6], John J. Helly[1,2], Steven M. Day[3,4], Raul Castro Escamilla[5], Philip Maechling[4], Thomas H. Jordan[4], Amit Chourasia[2,4], Mustapha Mokrane[6]

[1] SIO, [2] SDSC, [3] SDSU, [4] SCEC, [5] CICESE, [6] ICSU-WDS



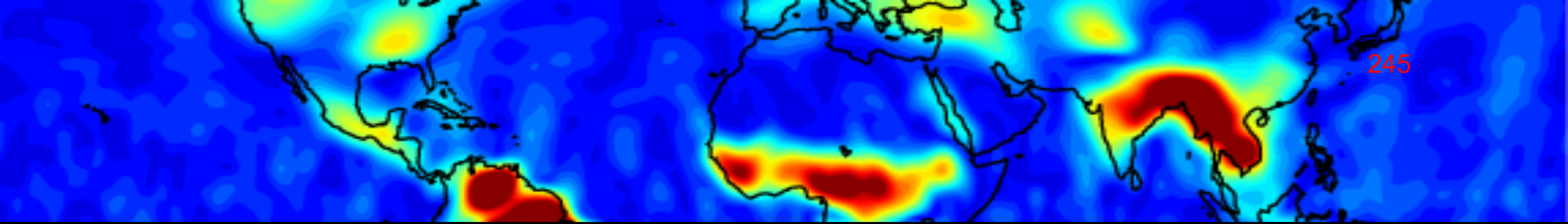
"Open data"

- Many countries have adopted an **open data policy**, at least for research and education (e.g. US, France, UK, ZA, etc.)
- This often includes the output of numerical models and simulations.
- But, because of different laws, large international organizations discuss "principles" instead of "policy".



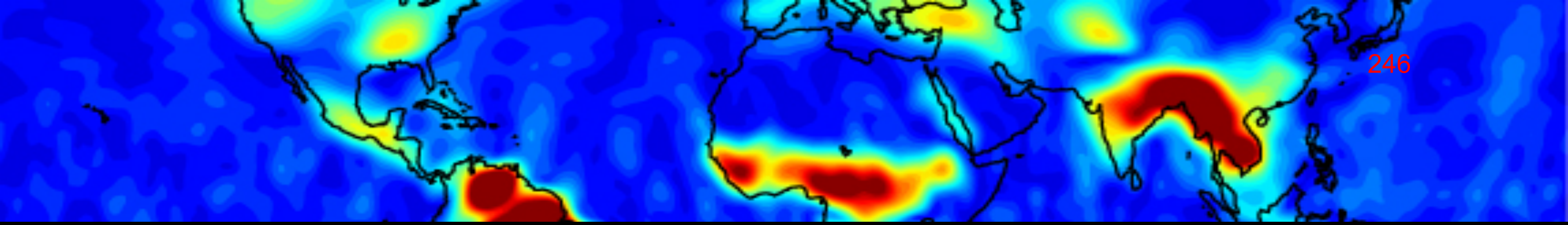
Data Sharing Policy

- ICSU World Data Centers (1958-2007)
- Federation of Astronomical and Geophysical Data Analysis Services (1958-2007)
- "Full and Open access to data"
- "Long-term data Stewardship and curation"



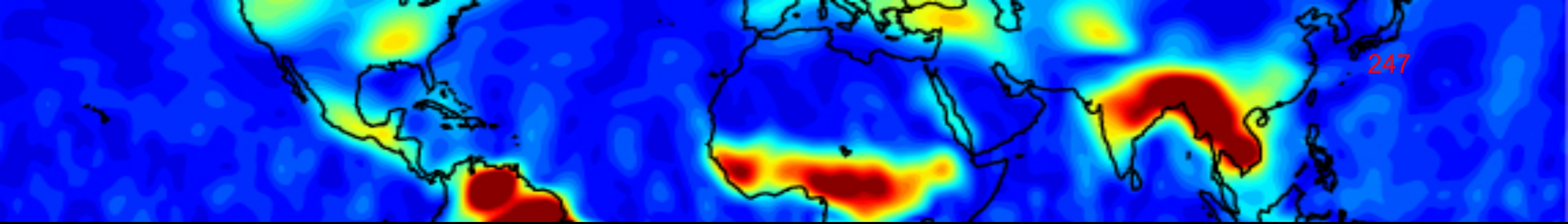
Data Sharing Principles

- Group on Earth Observations (GEO, 130+ nations) / Global Earth Observation System of Systems (GEOSS). 2010-present.
- Equitable, unimpeded access to data for research and education
- Long-term data preservation
- Many exceptions (National security, privacy laws, commercial protection, ecological protection)



Data Sharing Policy

- ICSU World Data System Data Policy (2008-present)
- “Full and Open access to data”
- “Long-term data Stewardship”



WDS Data Policy

ICSU

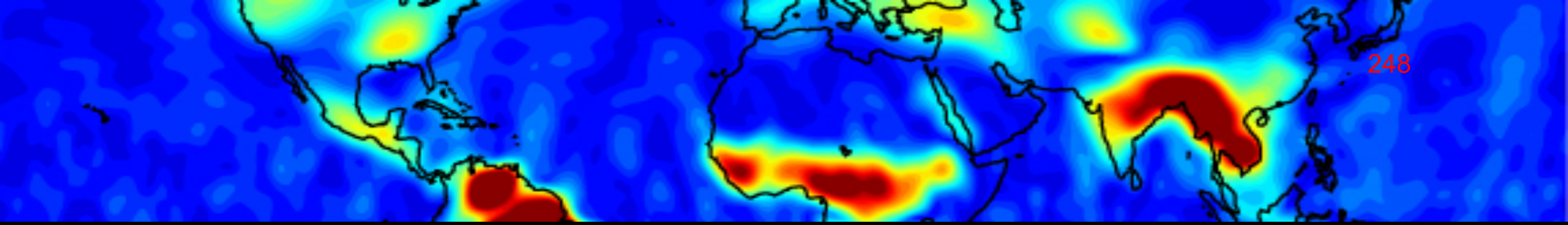
International Council for Science

ICSU WORLD DATA SYSTEM

WDS Data Policy Final Statement

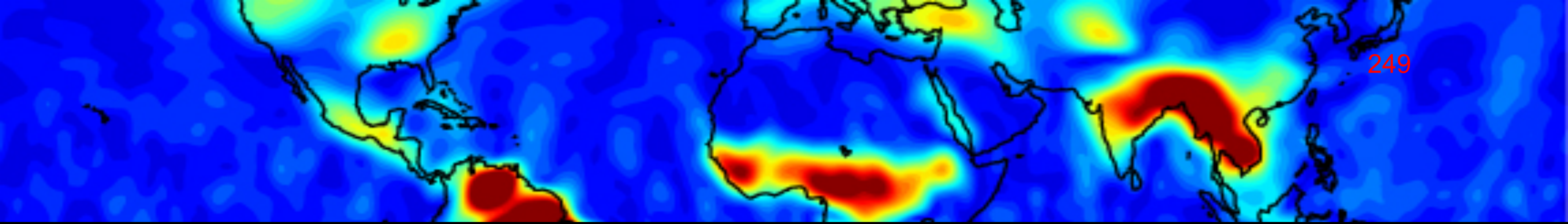
The International Council for Science World Data System (ICSU WDS), recognizing the benefits and importance of contributing to the growing international efforts of data sharing, has adopted the same principles from GEO/GEOSS data sharing principles as follow:

- There will be full and open exchange of data, metadata and products shared within WDS, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.



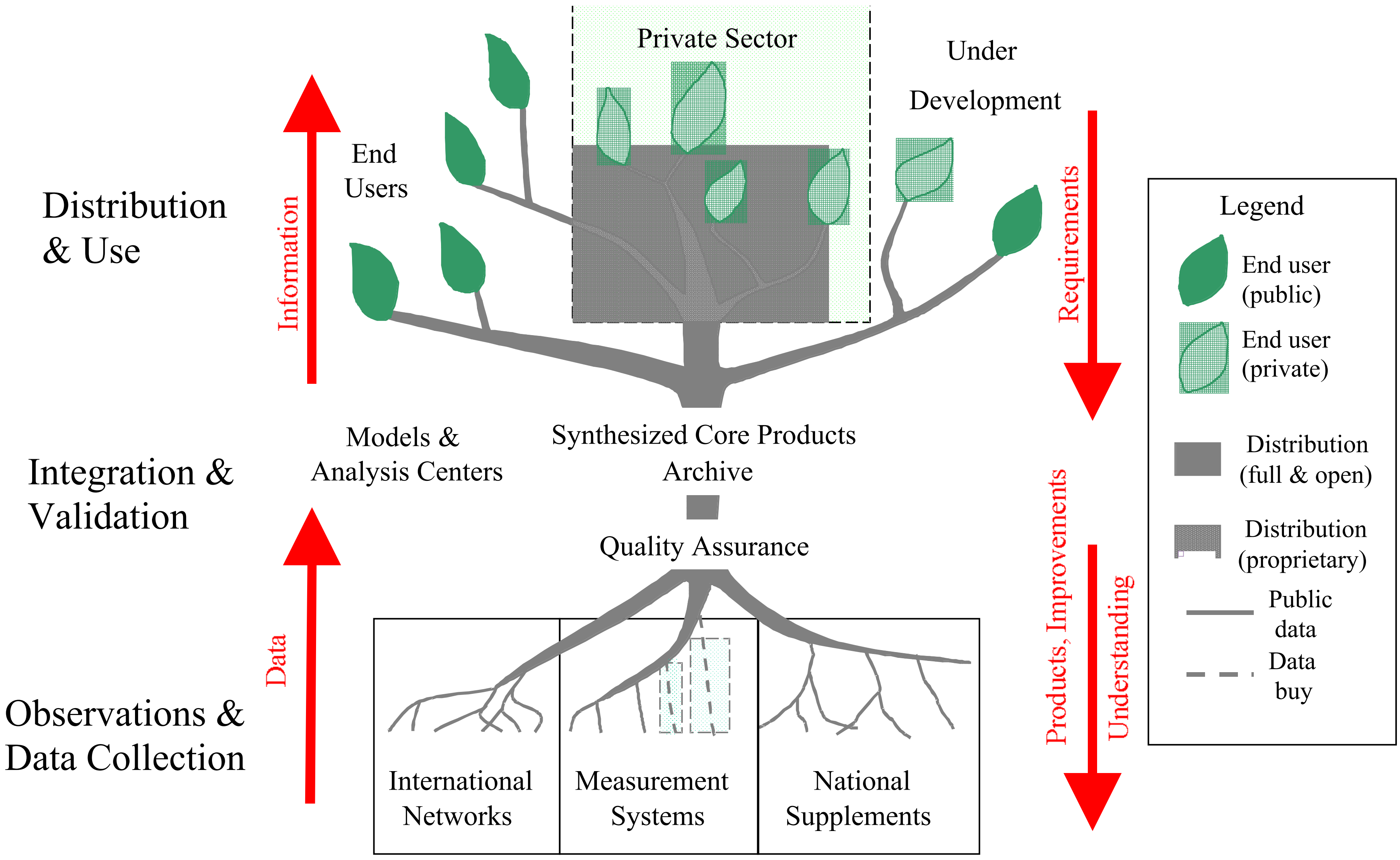
Research Data Alliance and WDS (RDA/WDS, 2013)

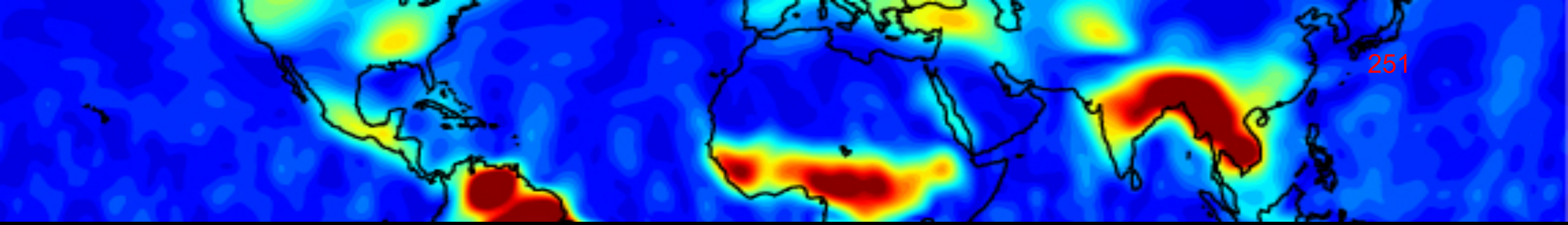
- Include socio-economic, health, and other data in policy discussions
- Explore data publishing concepts and issues
- Collaboration with publishers



- This works for observational data in the natural sciences, especially environmental data, that can never be acquired again…
- Perhaps also for socio-economic, and human health data sets (with caveats, so as aggregation)…

The Environmental Information System Tree





What about numerical simulation outputs?

- Issues are many, and difficult, e.g.:
  - Volume (can be enormous)
  - Quality (how is it measured and controlled?)
  - Metadata (what should be included?)
  - Costs (is it cheaper to re-compute?)
  - Needs (longitudinal studies, vs. punctual studies)
  - Requirements for data assimilation
- Examples: weather prediction, climate simulations, earthquake simulations, earthquake prediction algorithms
- This calls for a broad discussion

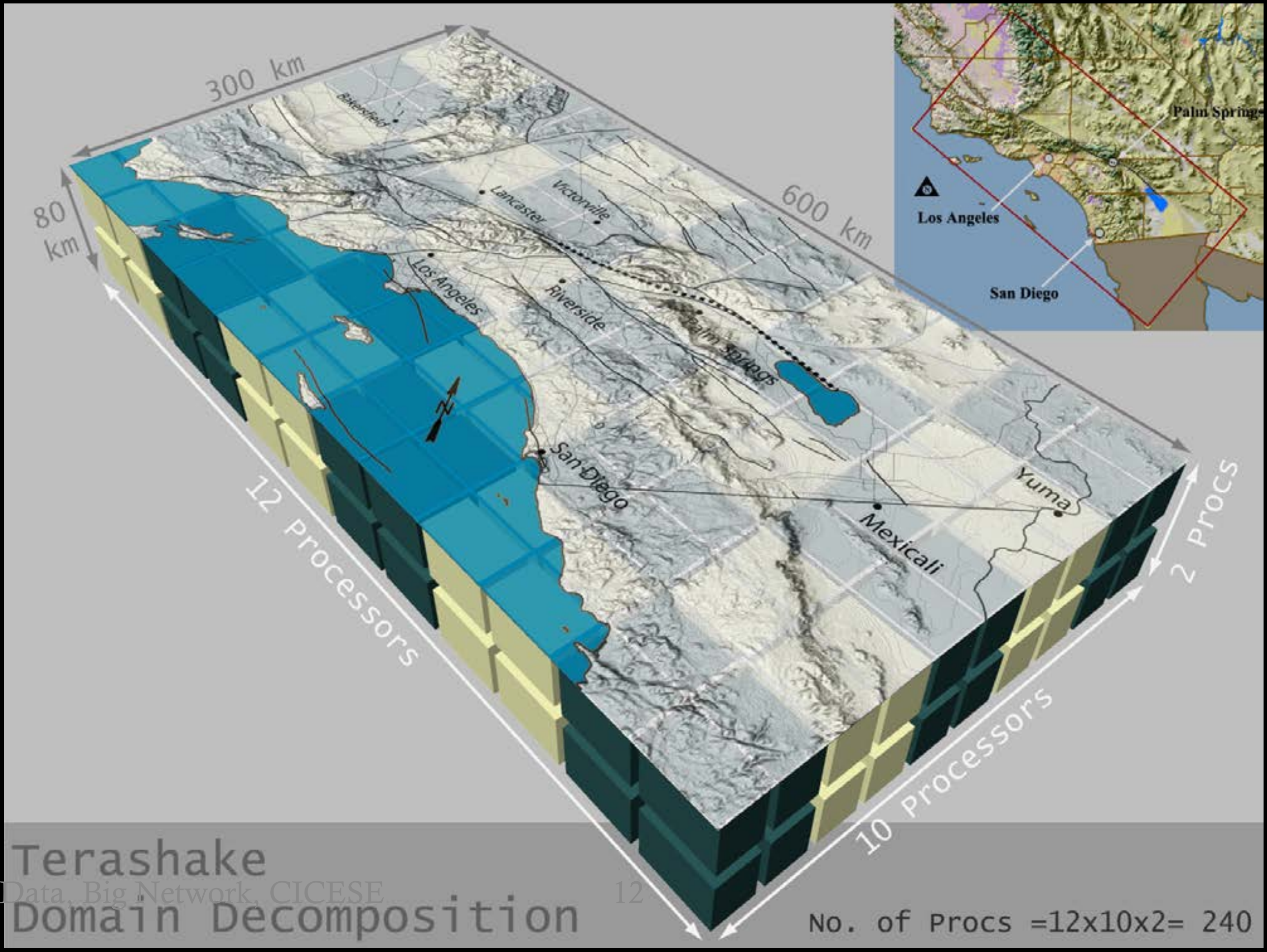
Minimalist Metadata (automatic capture)

- Code version
- HW platform (e.g. CPU, GPU, word length, etc)
- SW Platform (e.g compiler, options)
- Input and runtime options (workflow?)
- Other (Author, etc, Dublin core)

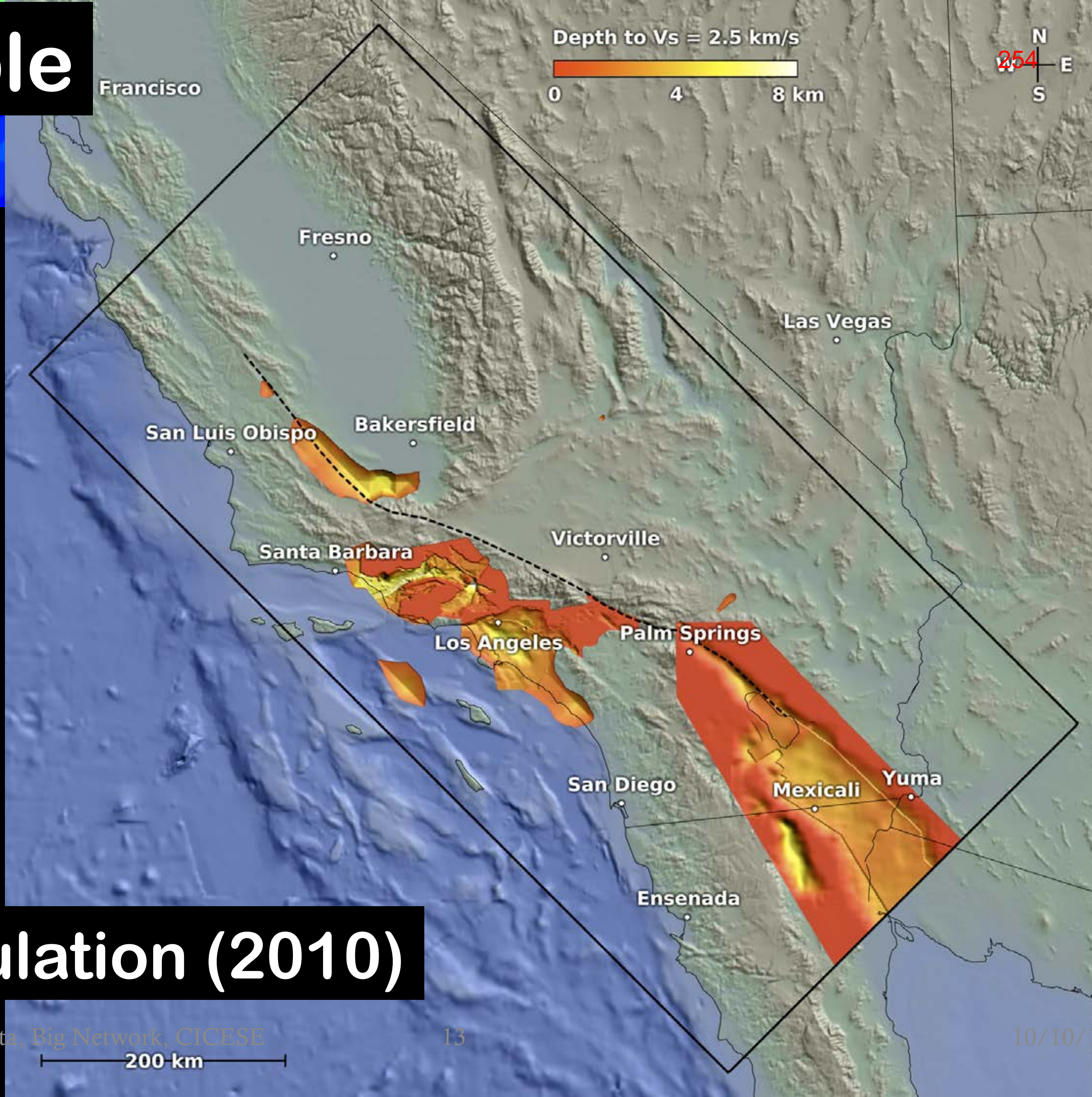
Even then, output might not be duplicated in future re-run. Many numerical outputs become obsolete.

Example

TeraShake Simulation (2004)



Example



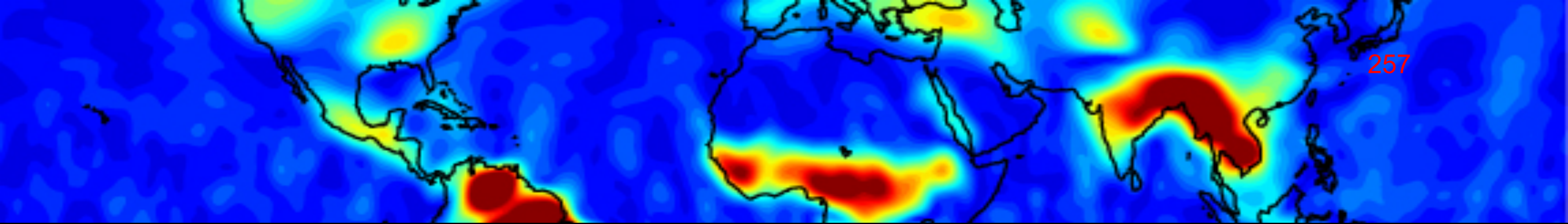
M8 Simulation (2010)

TeraShake vs. M8 comparison

	Terashake	M8	Notes
Dimensions	600x300x80 km	810x405x85 km	
# cells	$2\ 10^9$	$436\ 10^9$	
Time step	0.011 sec.	0.0023 sec.	
# steps (Duration)	20,000 180 sec.	160,000 368 sec.	
# cores	240 (Datastar)	223,074 (CPU) 16,600 (GPU)	
Wall clock	5 days	24 hours (CPU)* 5 hours (GPU)**	*220 Tflop/s **2.3 Pflop/s
Checkpoints	Every 1,000th step	Every 20,000th step	
Checkpoints, each	150 Gbytes	32 Tbytes	Cannot transfer
Checkpoints, total	3 Tbytes	192 Tbytes*	*Every 4 hrs

TeraShake vs. M8 comparison

	Terashake	M8	Notes
Surface Velocity vector field	All nodes, every step: 1.1 TB	Every other node, every 20th step: 4.4 TB (out of 352 TB)	Resolution OK for visualization
Total volume velocity field, all nodes, all steps	432 Tbytes	384 Pbytes	
Volume velocity field, decimated	All nodes, every 10th step: 45 Tbytes **	Every other node, Every 20th step 4,8 Pbytes	**No longer usefully readable, because of tape read errors
Typical Viz. movie	<100 Gbytes	< 100 Gbytes	Interactive Viz. possible



So…what to save?

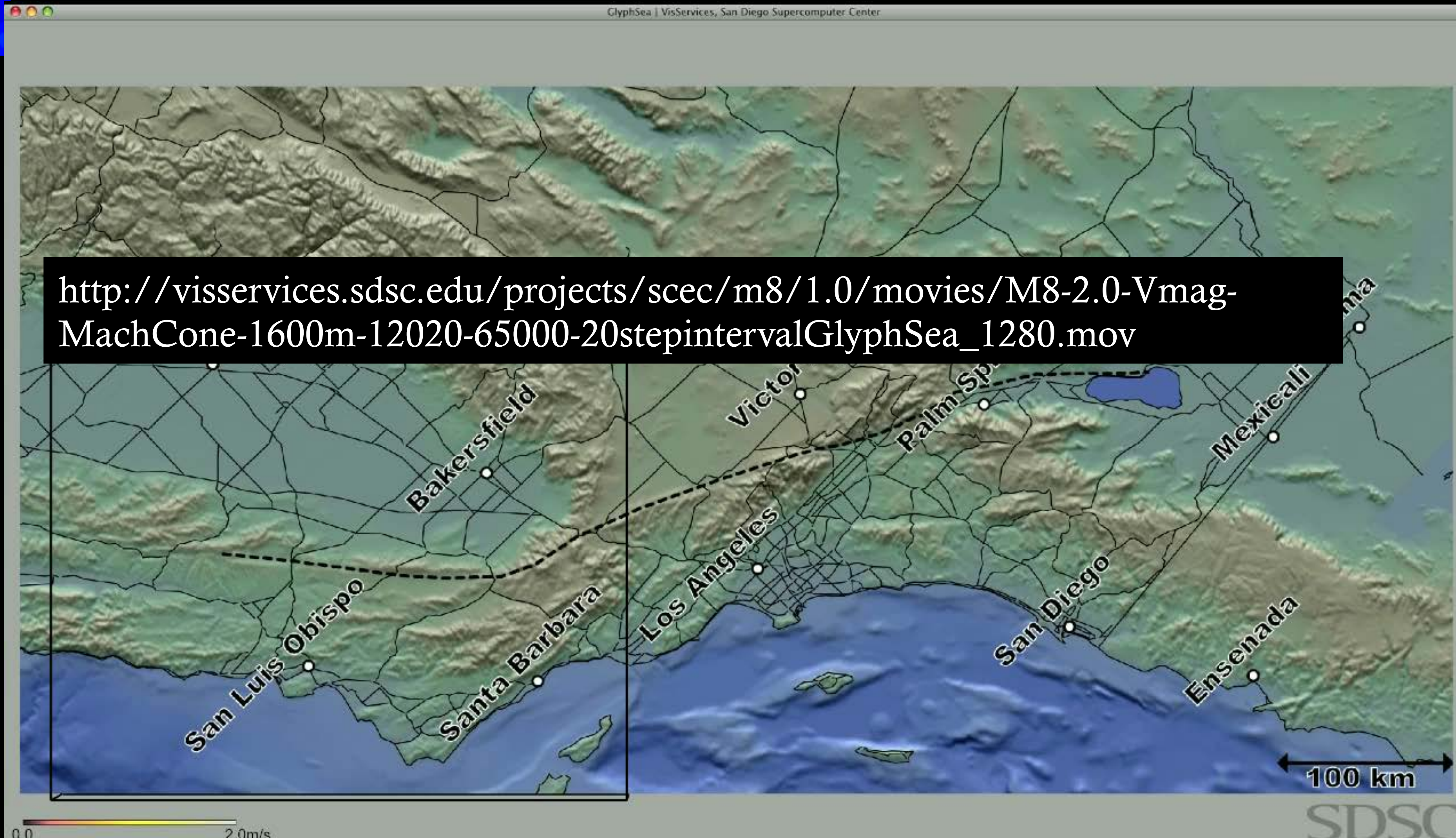
- Possible strategy: Only save enough to allow interactive (user or purpose-specific) visualization, and use checkpoints to restart partial calculation. This works for punctual simulations (e.g. 1-day weather, single earthquake). AMLIGHT permits that.
- Save selected individual visualizations that characterize the run (small size data sets). AMLIGHT makes it easy.
- For long-term longitudinal research, such as climate research or earthquake prediction algorithms, some output may require long-term curation by a trusted repository… This must be discussed on a case-by-case basis. AMLIGHT makes the data repository look proximal.

TeraShake Visualization

http://visservices.sdsc.edu/projects/scec/vectorviz/glyphsea/movies/GlyphSea-720p-cbr6.mp4

SDSC
VisServices

M8 Visualization



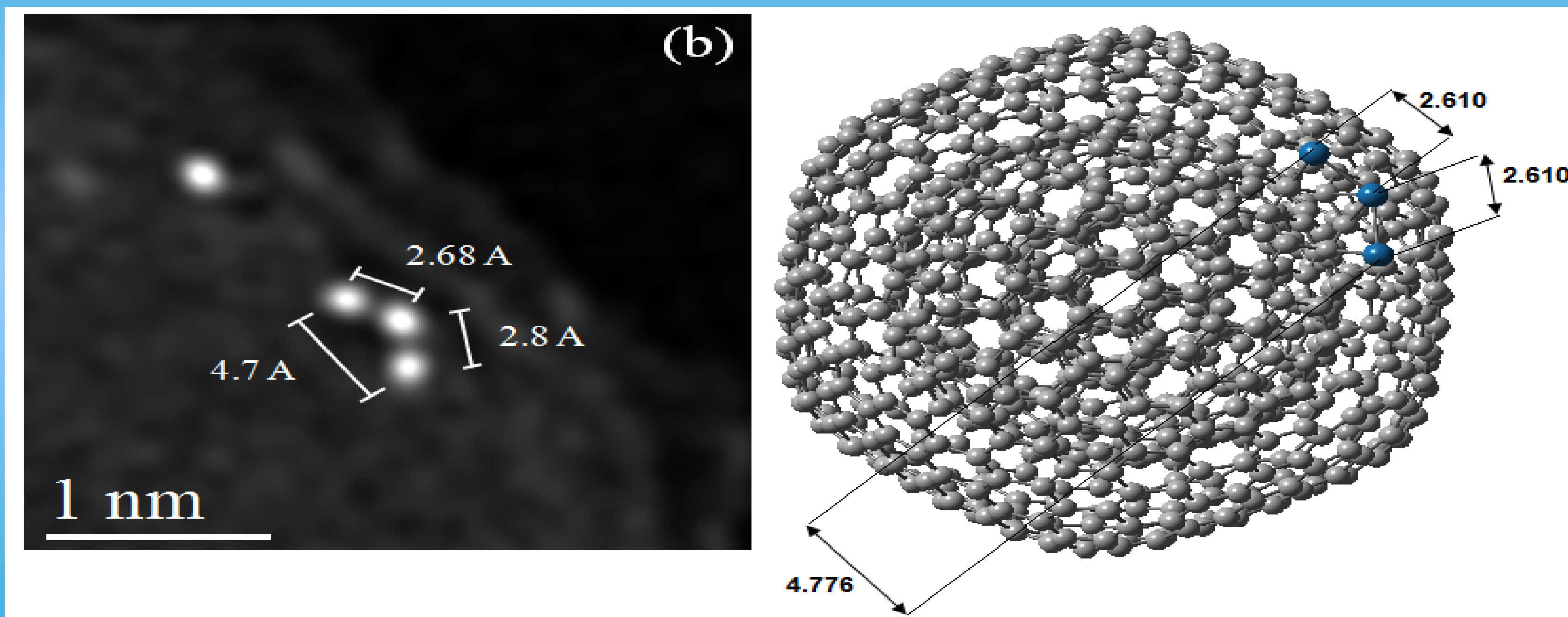
http://visservices.sdsc.edu/projects/scec/m8/1.0/movies/M8-2.0-Vmag-MachCone-1600m-12020-65000-20stepintervalGlyphSea_1280.mov

1H-MoS_2 nanoparticles grown on graphene and on 1H-BN monolayers

Donald H. Galvan[1], Gabriel Alonso[1], and S. Fuentes[1]

[1] Centro de Nanociencias y Nanotecnología-UNAM, Apartado Postal 2681, C. P. 22800, Ensenada, B. C., México.

Platinum Atom and Cluster Electrodeposition on Unsupported Carbon Nano-Onions



Diana Santiago, G. G. Rodríguez-Calero, A. Palkar, Diana Barraza-Jiménez, ***Donald H. Galván,*** G. Casillas, A. Mayoral, Miguel José-Yacamán , L. Echegoyen, and C. R. Cabrera

Langmuir 28 (2012)17202-17210.

Outline:

- Motivation
- Inquire about the electronic properties of 1H-MoS_2 nanoparticles grown on graphene and on 1H-BN monolayers
- Summary

The Nobel Prize in Physics 2010
Andre Geim, Konstantin Novoselov


Carbon is the *materia prima* for life and the basis for all organic chemistry. Due to the flexibility of its bonding carbon based systems show an unlimited number of different structures with an equally large variety of physical properties.

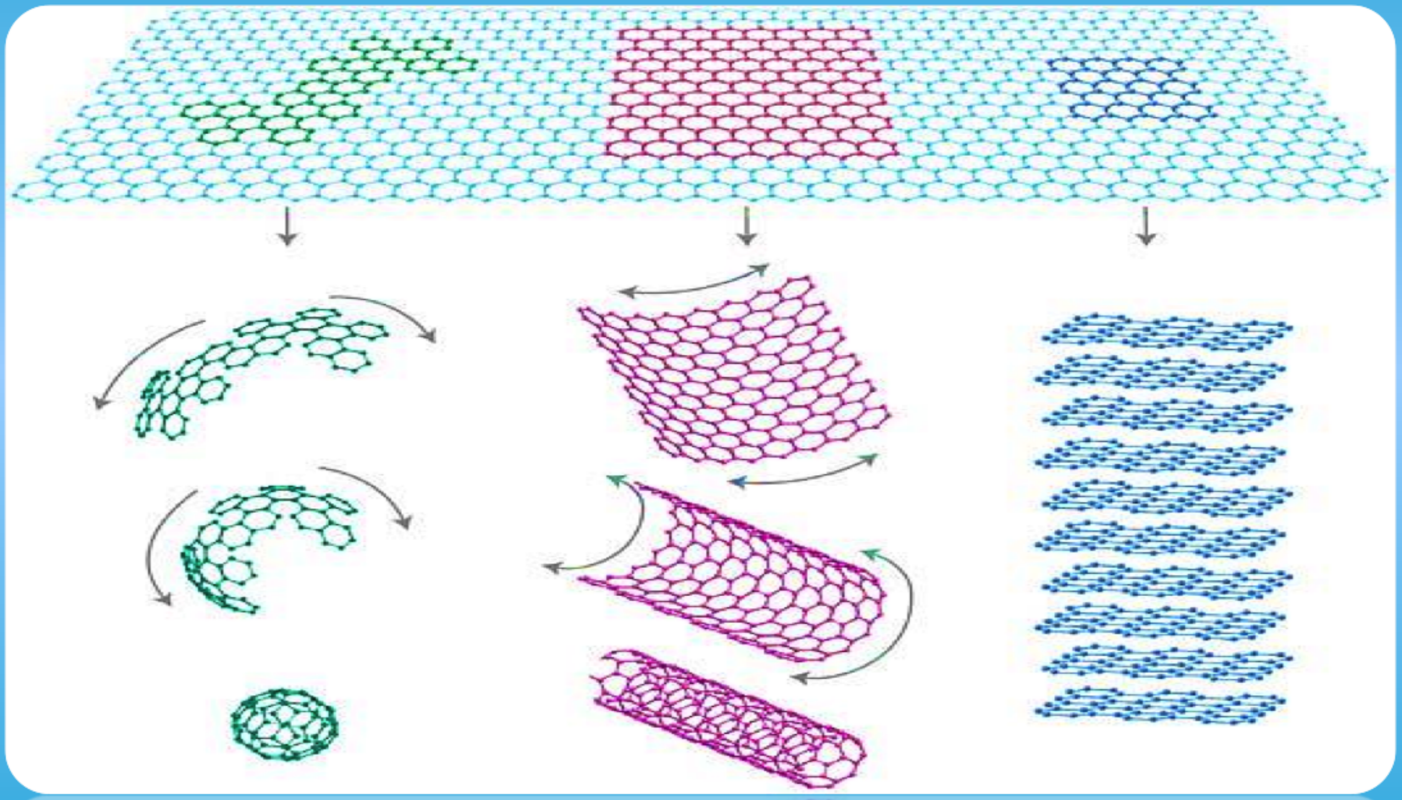


Figure 1 Graphene is a 2D building material for carbon based materials of all dimensions. It can be wrapped up into 0D buckyballs, rolled into 1D nanotubes or stacked into 3D graphite.

Introduction

•**Boron Nitride** (BN) the hexagonal form corresponding to graphite is the most stable among its polymorphs, and used in lubricants and additive to cosmetic products. BN has a great potential in nanotechnology. BN is an insulator with a forbidden energy gap Eg ~5.5 eV (same as diamond) independent of tube chiriality and morphology. It can be used in paints, dental cement and pencil leads.

•**Transition Metal Dichalcogenides (TMD)** were the first genuine quasi two dimensional compounds where charge density waves (CDW) instabilities was found. TMD have been extensively reviewed by Wilson and Yoffe. 2H-MoS2 crystallizes as an hexagonal structure. It has been reported that 2H-MoS2 crystalline bands are split into three sub bands separated by an energy gap Eg~ 1-1.9 eV.

•***MoS2*** is used as catalyst in the desulfurization in petroleum refineries.

• Solid lubricant.

•Photocatalytic hydrogen production.

•Nanotubes and Buckyballs-like molecules composed of MoS2 exhibit unusual tribology electronic properties.

•Recently, **graphene** layers have produced high interest because of their unique electronic properties. Graphene is the name given to a monolayer of carbon atoms packed in a two-dimensional honeycomb lattice. This system shows a peculiar effect, reminiscent of Relativistic Quantum Hall effect, which trigger the scientific curiosity for the fundamental physics behind its behavior. Graphene showed unusual properties derived from its reduced dimensionality as well as its small size, which yielded enormous potential to be used in ultrafast electronic transistors and switches. Graphene also shows a very high conductivity and Hall Effect. Graphene is considered a *zero gap semiconductor.*

J. Greim, K. A. Schwetz (2005) “Boron Carbide, Boron Nitride and Metal Borides”. Ullmann’s Encyclopedia of Industrial Chemistry. Weinheim Wiley-VCH, doi: 10.1002/14356007.ao4-295 pub. 2.
A.Wilson, A. D. Yoffe, Adv. Phys. 18(1969) 193.
F. Mattheiss, Phys. Rev. B8 (1973) 3719.
A. J. Grand, T. M. Griffiths, G. D. Pitts, A. D. Yoffe, J. Phys. C.: Solid State Phys. 18(1975) L17.

Theorical Calculations

The calculations reported in this work, have been carried out by means of the tight-binding method within the Extended Huckel framework using YAeHMOP (Yet Another extended Huckel Molecular Orbital Program) computer package with f-orbitals. The Extended Huckel method is a semi empirical approach for solving Schrödinger equation for a system of electrons, based on the variational theorem.

Theoretical calculations were performed on a system selected as a repeated cluster originated from a super cell. The super cells were generated from an inf nite single sheet of carbon atoms which aroused from crystalline graphite using the following primitive vectors: a = 2.456 Å, c = 6.696 Å and space group P63/mmc (194), while for 2H-MoS_2 the following primitive vectors were used:
a = 2.456 Å, c = 6.696 Å and space group P63/mmc (194). For BN layer the atomic position were obtained from Nanotube Modeler by JcrystalSoft (www.jcrystal.com). Atomic parameters for C, Mo, S, O, B and N atoms used in our calculation were obtained from Alvarez et al. and provided in Tables 1 and 2. Experimental lattice parameters instead of optimized values were used searching for a best matching of our theoretical results with the available experimental information.
To make the calculations simpler, a graphite single sheet made of 24 carbon atoms in a single layer was built with the construction of Graphene Oxide, while for 1H-MoS_2 we started from 2H-MoS_2 crystalline structure.

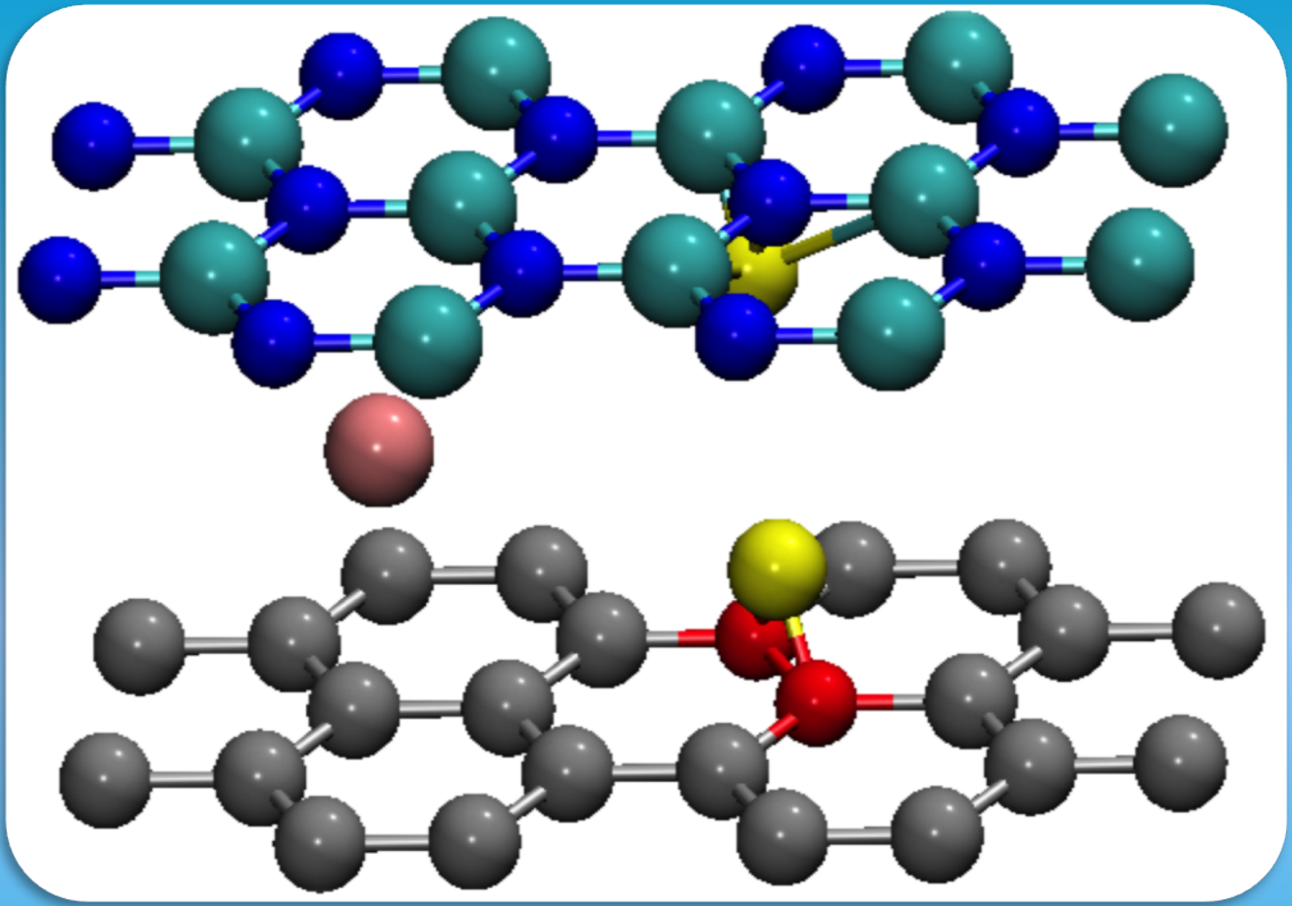


Figure 2 Hexagonal BN is made of B (blue) and N (sky blue). For 1H-MoS2 is formed by S (yellow) and Mollybdenum (pink). For graphene a cell made out of 24 Carbon (grey) atoms. Carbon 14 and 19 were removed and replaced by Oxygen to create vacancies. In addition O14 and O19 were replaced by C in order to create a defect.

Y. Li, H. Wang, L. Xie, Y. Liang, G. Hong, H. Dai, JACS in Press.

Atom	Orbitals	H_{ii} (eV)	$_{1}$	C_1	$_{2}$	C_2
H	1s	-13.6	1.30			
O	2s	-32.30	2.27			
	2p	-14.80	2.27			
C	2s	-21.40	1.62			
	2p	-11.40	1.65			

Table 1 Extended Huckel parameters H_{ii} for C, O and H.

C^6 – $[He]^2\ 2s^2\ 2p^2$
$O^8 2s^2 2p^4$
H 1s

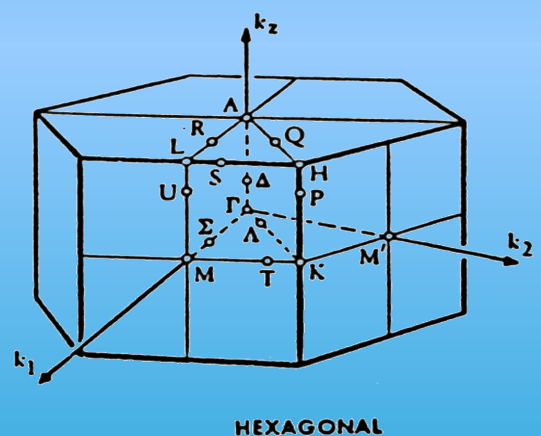


Figure 3 (Inset)

Atom	Orbitals	H_{ii} (eV)	$_1$	$_2$	C_1	C_2
Mo	5s	-8.34	1.96			
	5p	-5.24	1.90			
	4d	-10.50	4.54	1.90	0.6097	0.6097
S	3s	-20.0	2.12			
	3p	-11.0	1.83			
B	2s	-15.2	1.30			
	2p	-8.50	1.30			
N	2s	-26.00	1.95			
	2p	-13.40	1.95			

Table 2 Extended Huckel parameters H_{ii} for Mo , S, B and N.

Mo ^{42} $[Kr]^{36}$ $4d^{9}5s$
$S^{16}[Ne]^{10}3s^{2}3p^{4}$

B^{5} $[He]^{2}$ $2s^{2}$ $2p^{1}$
N^{7} $[He]^{2}$ $2s^{2}$ $2p^{3}$

[9] M. -H. Whangbo, R. Hoffmann, J. Am. Chem.. Soc. 100 (1978) 6093.
[10] R. Hoffmann, J. Chem. Phys. 39 (1963) 1397.
[11] G. A. Landrum, The YAeHMOP package is freely available on www at: http;//overlap.chem.Cornell.edu:8080/yaehmop.html. F orbitals are included in the calculations as version 3.0x, using W. V. Glassey's routine. W. V. Glassey, G. A. P [12] D. H. Galván, J. Mat. Sci. Lett. 17 (1998) 805apoian, R. Hoffmann, J. Chem. Phys. 111 (1999) 893.
[13] P. D. Fleischauer, J. R. Lence, P. A. Bertrand, R. Baner, Langmuir 5 (1985) 1009.

(a)

(b)

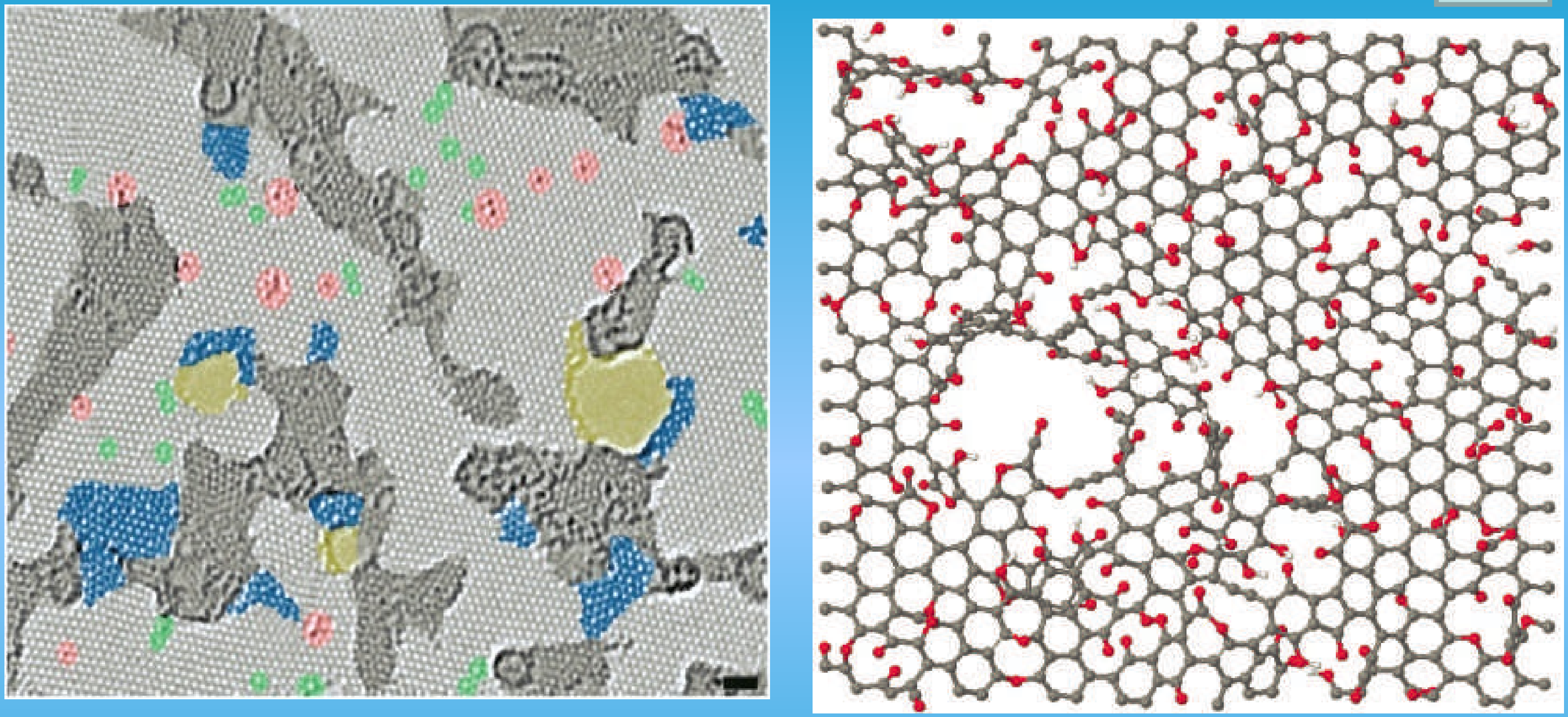
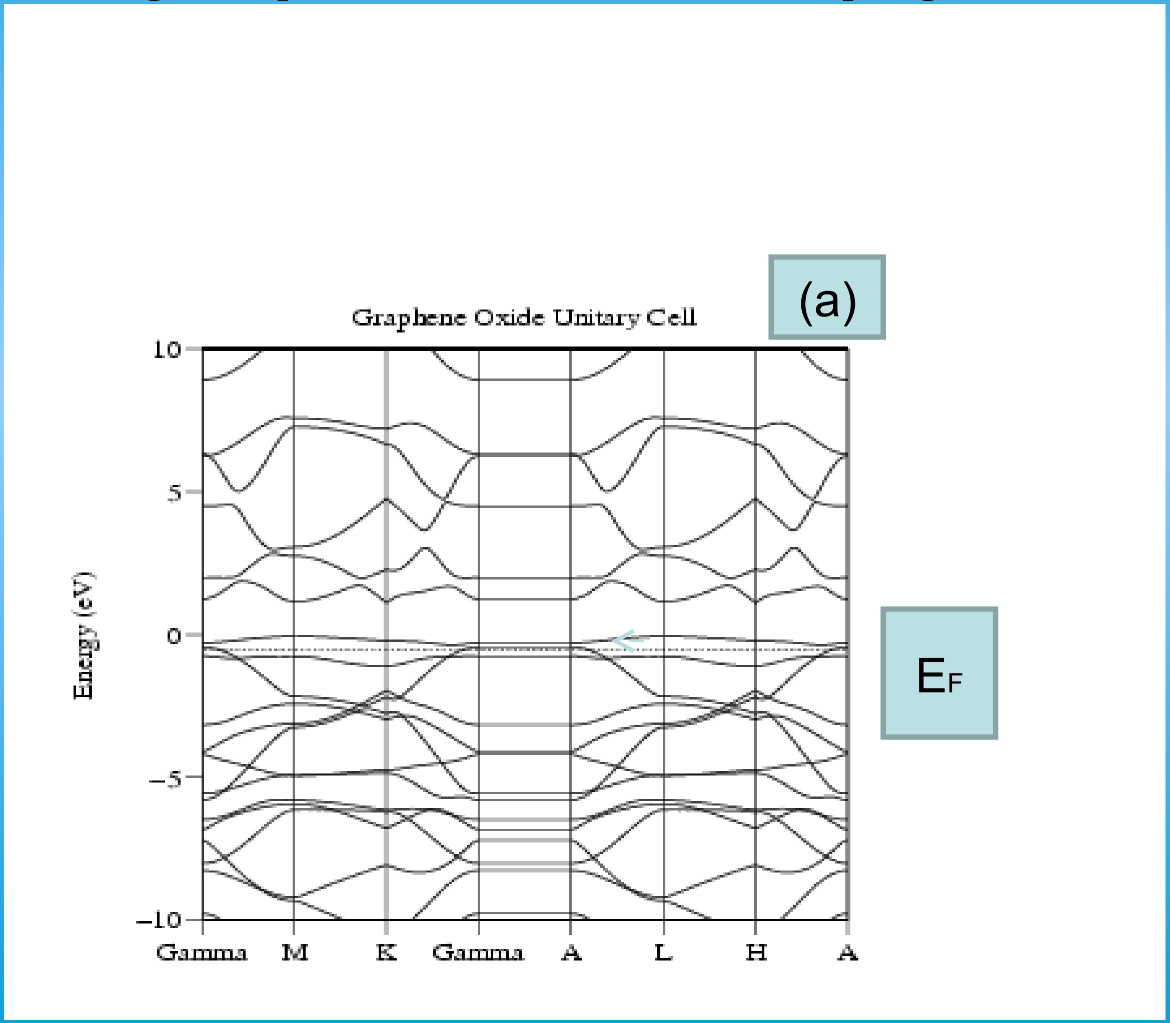


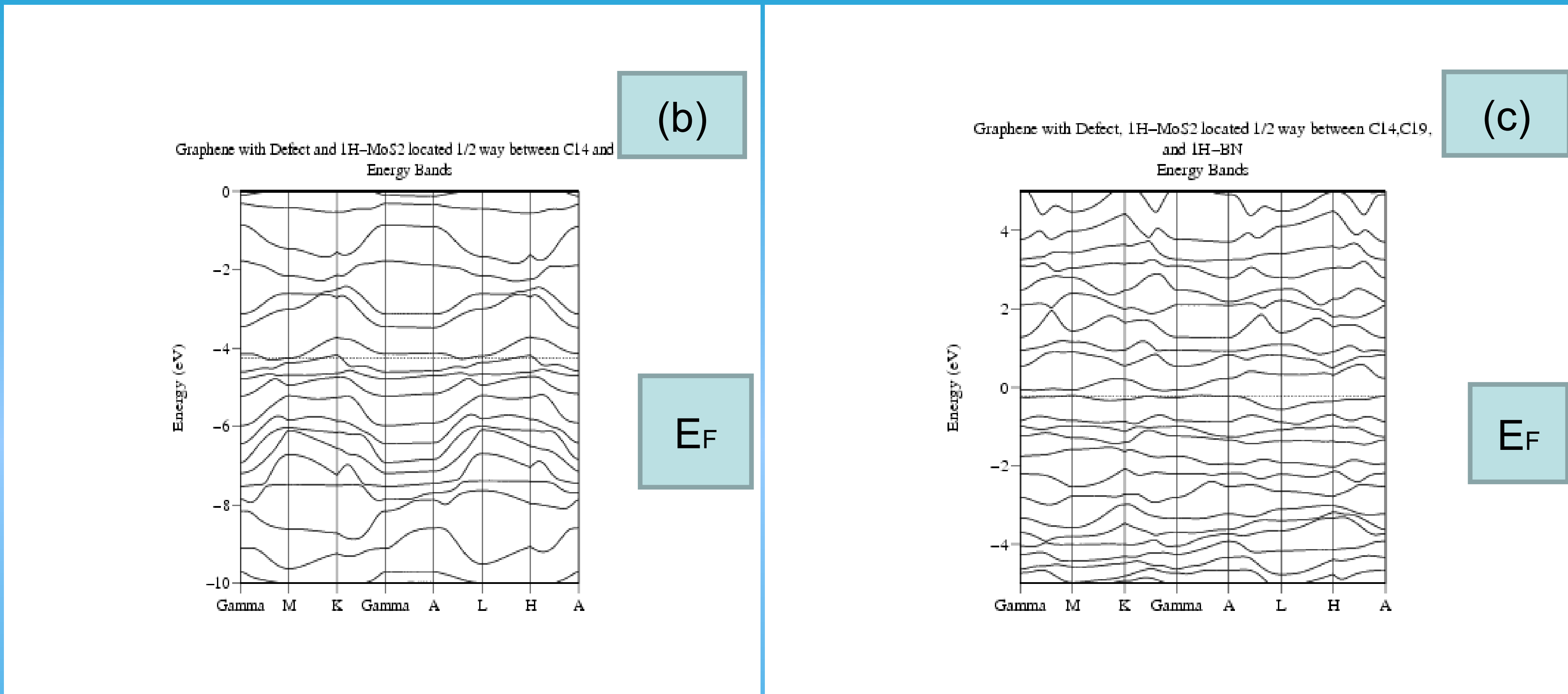
Figure 3 Atomic resolution, aberration-corrected high-resolution transmission electron micrograph of a single-layer GO membrane. (b) Atomic model schematically illustrating disordered GO basal plane consisting of holes, topological defects and remnants of oxygen groups.

K. P. Loh, Q. Bao, G. Eda, M. Chowalla, Nature Chem. 2 (2010) 1015.

Results and discussion

Band structure calculations for each case under investigation. They were calculated using 51 k points for each case, and sampling the First Brillouin zone





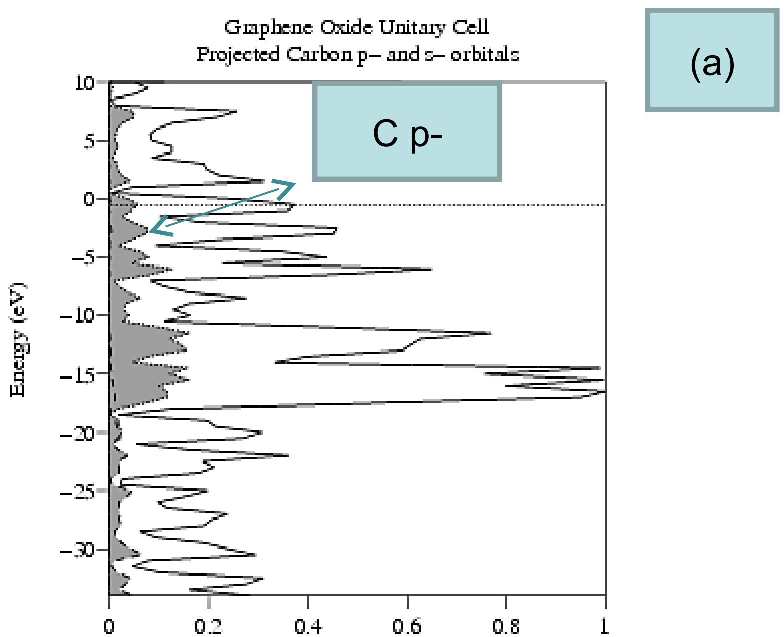
Figures 3 Energy Bands for (a) Graphene Oxide unitary cell, (b) Graphene Oxide with 1H-MoS_2 in defect, S-edge, (c) Graphene Oxide with 1H-MoS_2 located on the defect and then a BN layer is created above.

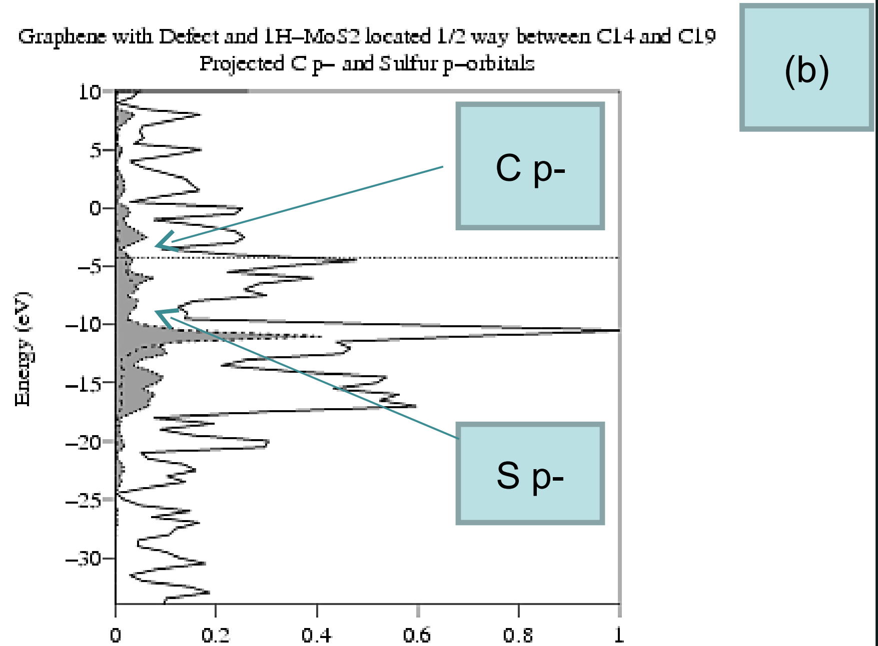
Structure	E_g (eV)	Fermi level (eV)	Behavior
Graphene cluster unitary cell (2D)	0.00	-0.5223	Semimetallic
Graphene with defect and 1H-MoS2	0.00	-4.2563	Semimetallic
Graphene with defect, the 1HMoS2 over the defect and then with 1H-BN	0.12	-0.2244	Semiconductor

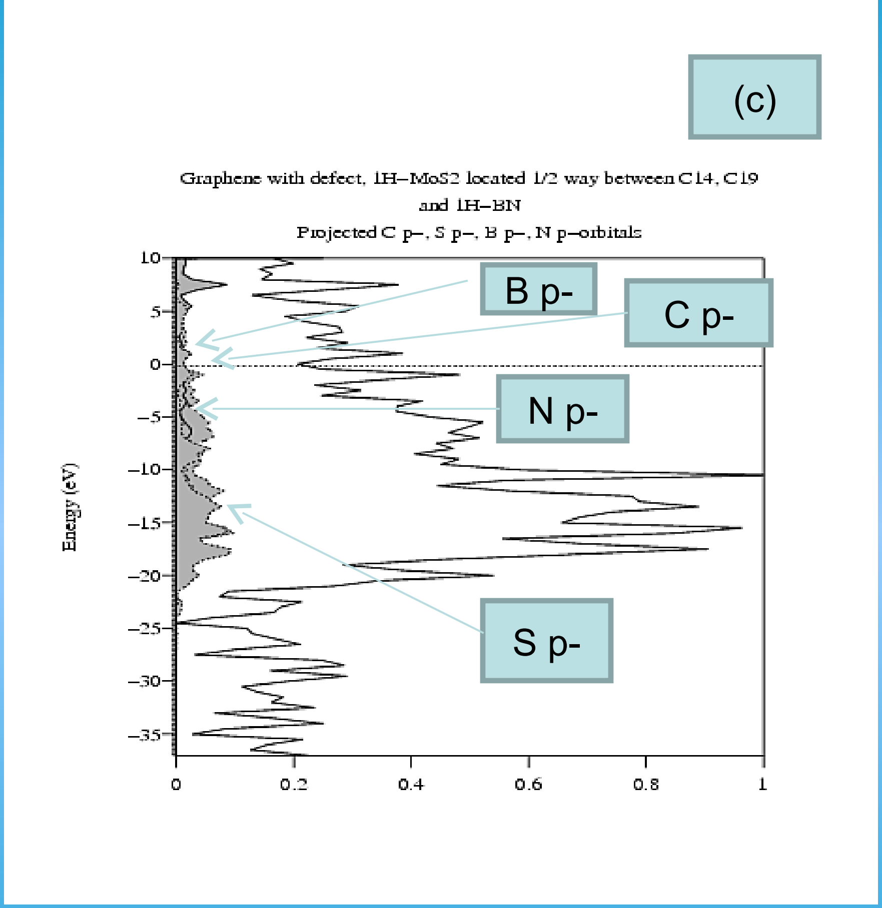
Table 3 Analysis for the forbidden gap E_g (eV), Femi level location and behavior for each configuration under investigation.

For the third configuration, the system behaves as a small gap semiconductor

Total and Projected DOS







Figures 5 (a) Graphene Oxide unitary cell, (b) Graphene with defect (C14, C19) with 1HMoS2 s-edge located ½ in-between C14 and C19. (c) Graphene with defect, 1H-MoS2 located ½ between C14, C19, then 1H-BN layer located above.

Table 4 % Contribution of each orbital to the Total DOS in the vicinity of the Fermi level.

	C p-	s-		C p-	S p-		C p-	S p-	B p-	N p-	
	%			%				%			
Graphene unitary cell	5	0	**(5)**								
Graphene with defect and 1H-MoS2				3	1	**(4)**					
Graphene with defect and 1H-MoS2 over the defect and with 1H-BN							2	<1	<1	2	(5)

Overlap occurred between atomic orbitals. The overlap form a hybridized band.

Average Net Charge

Structure	C_2	C_3	C_4	C_8	C_{14}	C_{18}	C_{19}	Mo_{25}	S_{26}	S_{27}
Graphene unitary cell	-0.1266	-1.1344	-0.6407	-0.6135	-0.6585	-0.1290	-1.1397			
Graphene with defect 1H-MoS_2	+0.3607	-0.5585	-0.2725	-0.3410	+0.3272	+0.3308	-0.0179	-9.5504	+0.6843	-1.9860
Graphene with defect, 1H-MoS_2 over defect and 1H-BN	-0.0188	-0.9130	-0.4933	-0.4591	+0.0665	+0.0888	-0.0474	-10.1750	+0.2429	+0.0829

B_{28}	B_{29}	B_{33}	B_{34}	B_{36}	N_{40}	N_{41}	N_{43}	N_{44}	N_{47}
+1.2509	+1.2853	+1.0916	+1.4335	+1.2043	-1.1024	-0.0259	-1.3354	-1.4114	-1.2587

Table 5 Average net charge in some selected atoms in an specif c conf guration.

Redistribution of charge depending on the conf guration

Conclusions

The results obtained yielded information regarding Graphene cluster, 1H-MoS_2 nanoparticles located on graphene layer when 1H-MoS_2 located on different places as oxygen vacancies and on defects, and 1H-MoS_2 nanoparticles grown on a single layer of graphene, then a 1H-BN monolayer was constructed on top to create a *sandwich* of these three structures. The f rst two structures are semimetallic while the third one is a small gap semiconductor. In addition, Total and Projected DOS analysis on each case enunciatated formerly, provided information on the contributions from each orbital from each atom to the Total DOS. This overlap of orbitals produce a very well localized hybrid band in the vecinity of Fermi level. Average Net charge provides information regarding charge distribution in the conf gurations analyzed.The experimental results are in progress hoping that in the near future we could be able to correlate experimental and theoretical results in the systems under investigation.

Acknowledgments

D. H. G. Acknowledge Y. Flores from Supercomputo-UNAM. Authors would like to acknowledge E. Aparicio, M. Sáinz and J. Peralta for technical support. Also, they acknowledge DGAPA-UNAM for providing computer time in New Supercomputer Facilities

Theoretical Background

The Extended Hückel Molecular Orbital calculation (EHMO) is the simplest approximation Linear Combination of Atomic Orbitals Molecular Orbital (LCAO-MO) approximation.[2] The c_{ij} are called the molecular orbitals coefficients. They may be either positive or negative and the magnitude of the coefficient is related to the weight of that AO in the MO. EHT uses only the valence orbitals of each atom in the molecule or solid, namely one s and three p orbitals for main elements, plus five d orbitals for the transition elements. These AO are assumed to be real functions and normalized such that the probability of finding an electron in $ø_i$ when integrated in all space is unity.[2a] The MOs are normalized and orthogonal (i. e. orthonormal).[2b]

$\Psi_\alpha = \sum c_\alpha \, ø_j \qquad \Psi_\alpha = \mathrm{MO}(\alpha) \qquad ø_j = \mathrm{AO}(j)$ [2]

$< ø_i \,|\, ø_i > = \int ø_i^* \, ø_i \, d\tau = 1$ [2a]

$< \Psi_\alpha \,|\, \Psi_\beta > = \int \Psi_\alpha^* \Psi_\beta \, d\tau = \delta_{\alpha\beta}$ [2b]

in bracket notation.

In the eigenvalue equation [3] the energy ε_α measures the effective potential exerted on an electron located in the αth molecular orbital (MO) Ψ_α.

$\mathbf{H} \, \Psi_\alpha = \varepsilon_\alpha \Psi_\alpha$ [3]

$\varepsilon_\alpha = < \Psi_\alpha \,|\, \mathbf{H} \,|\, \Psi_\alpha >$ [3a]

The coefficients are chosen such that the energy is minimized.[3a].
The system of linear equations obtained is express in a compact form in [4].
The theory of simultaneous equation requires that the determinant [5] vanishes.

$\sum [\, H_{ij} - \varepsilon_\alpha S_{ij}] \, c_{ij} = 0$ [4]

$|\, H_{ij} - \varepsilon_\alpha S_{ij} \,| = 0$ [5]

To solve [4] the EHMO method introduces some assumptions:

i) The diagonal elements of **H** are taken to be equal to the
ionization energy of an electron in the ith valence AO of the isolated atom in the appropriate state (Valence State Ionization Potential).

$H_{ii} = -VSIP(ø_j)$ [6]

This value can be derived from spectroscopy or computed from accurate methods (*Ab initio*).
ii)The off-diagonal elements of **H** are evaluated according to Wolfsberg-Helmholtz relation [6] where K is an adjustable paramenter.

$H_{ij} = K\ S_{ij}\ (H_{ii} + H_{ij})/2$ [7]

iii) Basis set: the atomic valence orbitals are approximated with Slater-type orbitals (STO), of single type for s and p, double zeta for d orbitals:

$ø_{s,p} = r^{n-l}\, e^{(-\ r)}\, Y(\theta,\omega)$ [8]

$ø_d\ = r^{n-l}\, [\, C_1\, e^{(-\ 1r)} + C_2\, e^{(-\ 2r)}]\, Y(\theta,\omega$ [9]

where n = principal quantum number
r = distance of the electron from the nucleus
$Y(\theta,\omega)$ = angular part of the wavefunction
, $_{1,}$ $_{2}$, $C_{1,}\ C_2$ are tabulated computer constants.

Future work:

The experimental part of this investigation is under progress at the moment.

Thank you very much for your atention

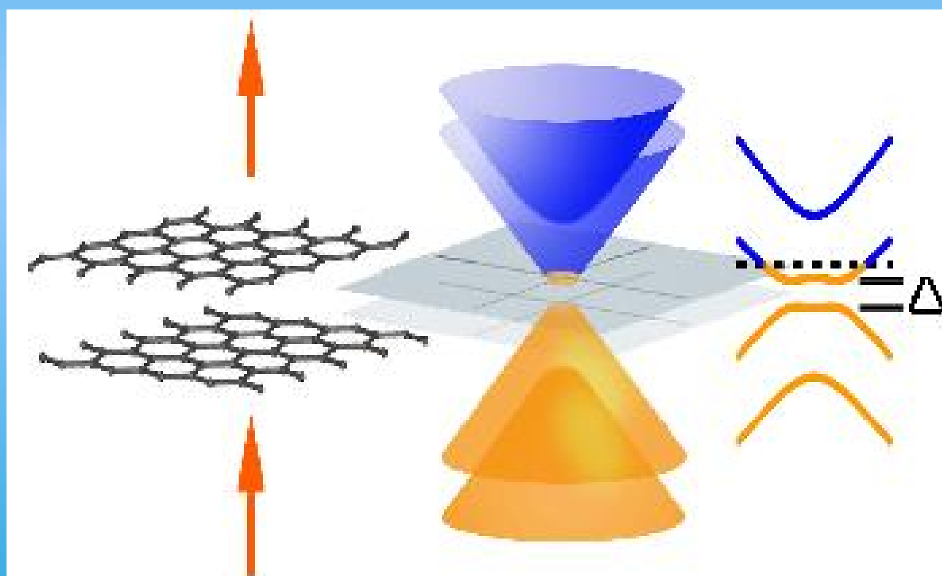
CIAD

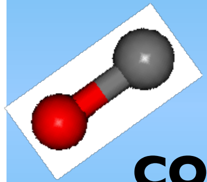
Centro de Investigación en Alimentación y Desarrollo A.C
UNIDAD DELICIAS

Centro de Investigación en Materiales Avanzados, S. C.

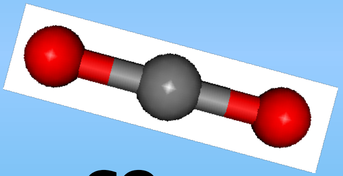
Universidad Nacional Autónoma de México

Bilayer Graphene as a Gas Detector

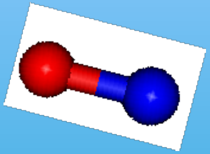




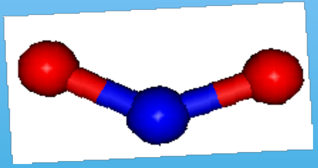
CO



CO_2



NO



NO_2

Diana Barraza-Jiménez, Manuel Alberto Flores-Hidalgo, Donald. H. Galvan

Carbon Nanomaterials

Graphene as a Gas Detector

Graphene, a two dimensional (2D) sheet of carbons in a honeycomb arrangement had attracted a huge scientific interest, this is because outstanding, transport properties, chemical and mechanical stability and to the scalability of graphene devices to nano-Dimensions.

Thanks to charge carrier oneness, the graphene-based electrical sensors have the potential to revolutionize sensor technology.

This is in agreement with the following characteristics:

• Graphene is a true 2D material, it offers its whole volume to the surface adsorbates, providing maximal sensitivity in terms of surface/volume ratio

Graphene as a Gas Detector

•It is highly conductive and exhibits low Johnson noise even in the limit of few charge carriers, thus several extra electrons can cause notable relative changes in charge carrier density

• Graphene is an easily accessible material for performing electrical four-probe measurements yielding contacts that are Ohmic and have low resistance.

Molecular Model of Graphene Bilayer Sheet

For understanding experimental results one needs to have a clear knowledge of the molecules' behavior on graphene surface: where is the minimum energy for a molecule to stick on the surface, what is the value of charge transfer in a graphene/molecule system.

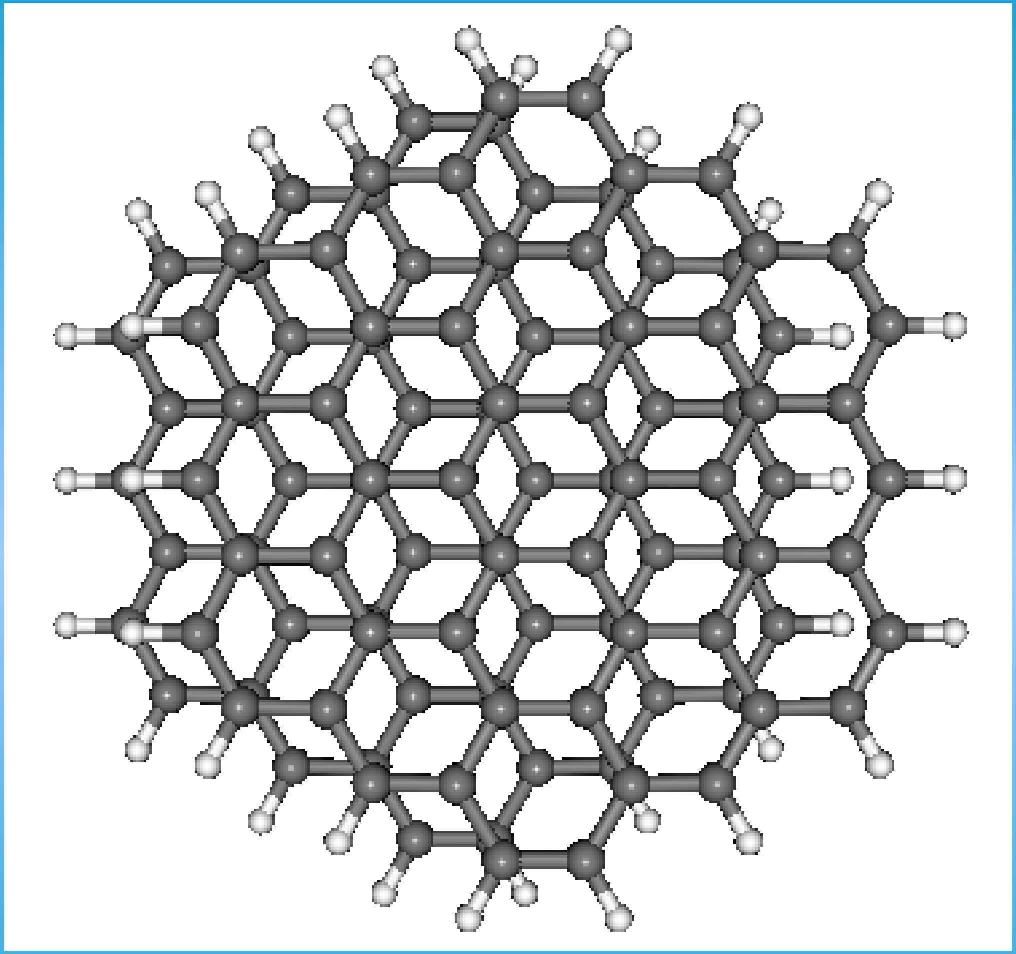
• This work is based in molecular simulation of graphitic/graphenic islands, which can provide a suitable molecular model of graphene sheets.

• Mimicing a graphene single sheet, we used a suggested model by Compagnini et al; 2011. This mode consist in a circumcoronene that is a planar polycyclic Aromatic hydrocarbons (PAH) with D_{6h} symmetry.

• Bi-layer graphene geometry was relaxed using density functional theory (DFT) with the B97-D functional and 631G(d) basis set both as implemented in the Gaussian09 program.

• Frequency and energy calculations were carried out for the graphene sheet in its ground state. All these calculations were performed in singlet state because stability presented this same state.

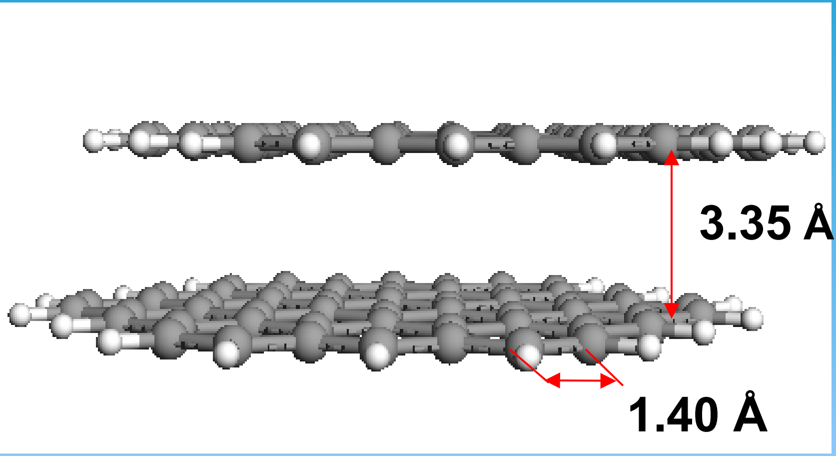
Molecular Model of Graphene Bilayer Sheet



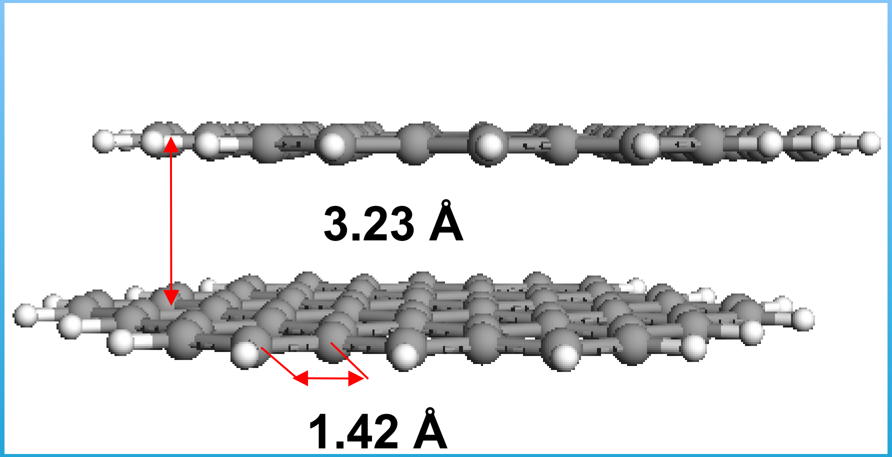
Each layer is a circumcoronene of 19 Benzene rings.

C+ ion Irradiation by 500 keV is posible to isolate graphitic domains in nanometer range Compagnini, et al; 2011.

Graphene before optimization

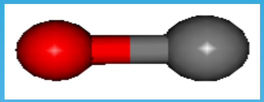


Graphene after optimization



Characteristics of the considered molecules

Carbon Monoxide



•Is an odorless, colorless, inflammable and highly toxic gas.

• If inhaled in highly levels, it could cause death.

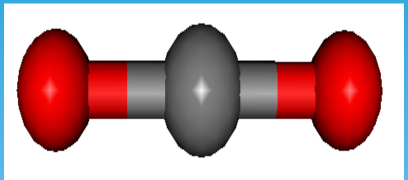
• Produced by incomplete combustion of substances like gas, gasoline, kerosene, carbon, petroleum, tobacco or wood.

• Stopped vehicles with the motor running can emit this gas as well.

• It could cause death by poisoning in few minutes, if inhaled even in moderate amounts, because it substitutes oxygen in the blood hemoglobin.

Gas Molecules Characteristics

Carbon Dioxide



- Molecule with lineal and symmetric geometry.

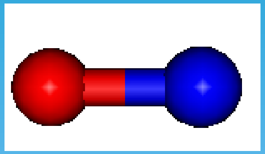
- One of the more abundant gasses in the atmosphere.

- The increase of carbon dioxide emissions are responsible for 50-60% of the global warming.

- The impact caused is related to the concentration level, 1000 times superior to any other product from the industry.

Gas Molecules Characteristics

Nitrogen monoxide



•It is a colorless with low water solubility which is present in low quantities in mammals.

• Extended in the atmosphere being produced by energy plants and automobiles. It is considered a toxic agent.

• It is a molecule highly unstable in air because it oxidizes quickly in the presence of oxygen becoming a nitrogen dioxide. For this reason it is also considered as a free radical.

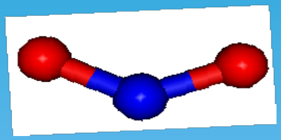
• At high temperatures molecular nitrogen (N_2) and oxygen (O_2) can combine to form nitric oxide.

• NO and NO_2 participate in the ozone layer depletion.

• In the high atmosphere it is a contributor to the global darkening.

Gas Molecules Characteristics

Nitrogen dioxide



•One of the main contaminants among the nitrogen oxide variants.

• Formed as a sub-product in high temperature combustion processes, like motorized vehicles and electrical plants.

• Exposition at high concentration levels in a short time can cause lung cell damage, meanwhile, long term exposition in lower levels of nitrogen dioxide can cause irreversible changes in the lung tissue, similar to the damage caused by emphysema.

• Nitrogen dioxide is paramagnetic due to a free electron. Sometimes is used as “radical trap”, because it reacts easily with radical organic substances.

Methodology

• Molecules were obtained from PubChem database.
http://pubchem.ncbi.nlm.nih.gov/

• Each one of the molecule geometries was optimized using Density Functional Theory (DFT) with B97-D funtional and 6-31G(d) basis set as implemented in the program Gaussian09.

• Frequency calculations allowed to verify that the structures found were really stationary points.

• Using the optimized geometry, ground state energy calculations were carried out, which enabled us to calculate values for the frontier orbitals HOMO and LUMO.

• To optimize the graphene molecule, we used the central hexagon of a circumcoronene and the atoms were located in the C-C bridge bond sites.

Methodology

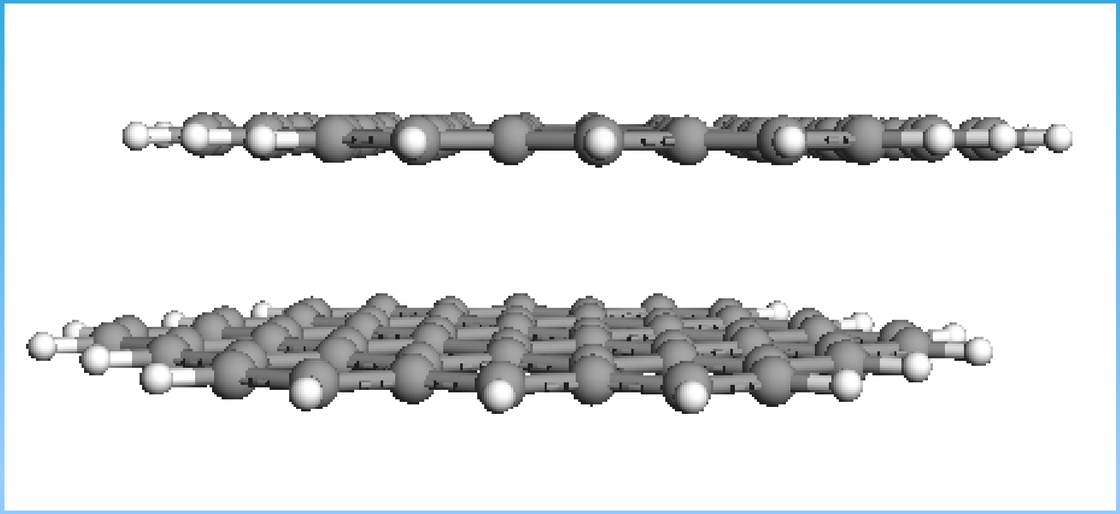
•The separation distance between the molecule and the graphene bilayer was 2.0 Å.

• Once placed the molecule over the graphene sheet, we performed the geometry optimization using the same theoretical level (with the same functional B97-D and the same basis set 631-G(d) as used in to treat the circumcoronene in prior calculation).

• Then, the next calculation was to obtain the frequency using the same functional and basis set used in previous calculation, to corroborate the stationary point was obtained (non-existence of imaginary frequencies).

• Finally the neutral energy is calculated, and from such calculation the frontier orbitals HOMO and LUMO were obtained.

Graphene Optimized Structure

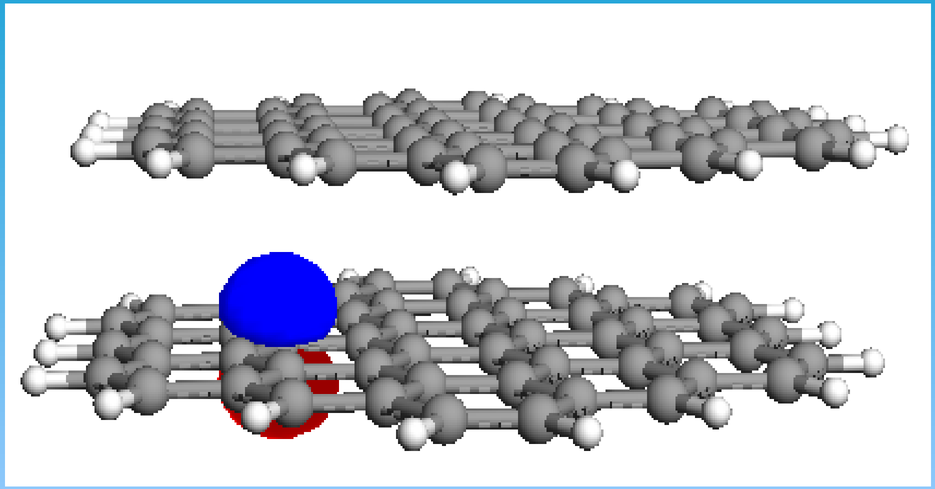


Distance Data

Distance Between layers: 3.23

Bond Length C-C: 1.42

Graphene Frontier Orbitals

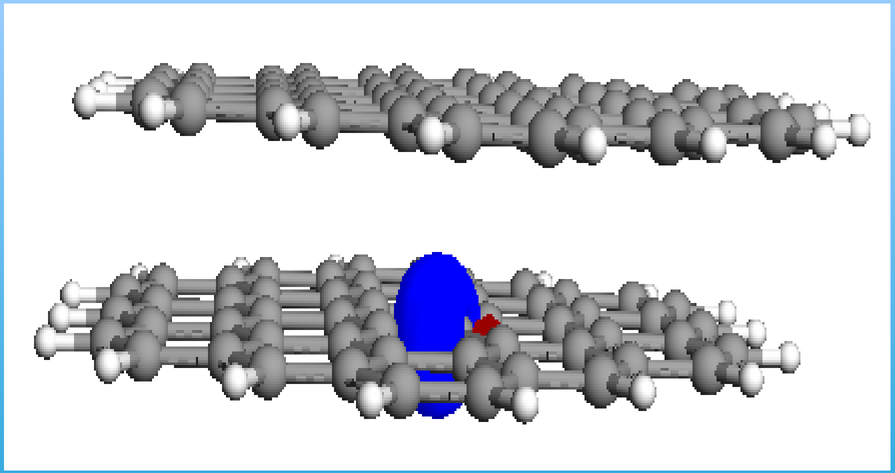


HOMO

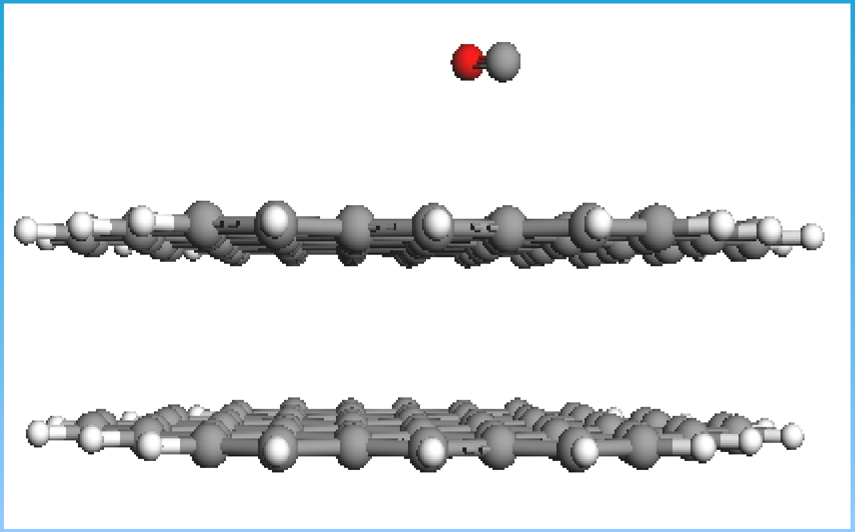
-0.15065 eV

LUMO

-0.09117 eV



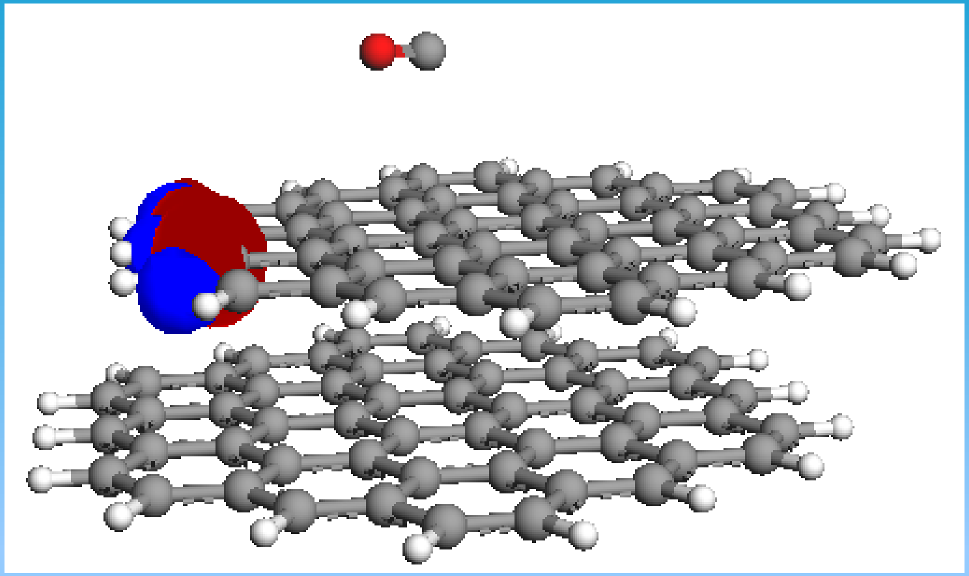
CO-Graphene Optimized Structure



Distance Data

Distance Between layers:	3.23 Å
Bond Length C-C:	1.42 Å
CO_Graphene Distance:	3.42 Å

CO-Graphene Frontier Orbitals



HOMO

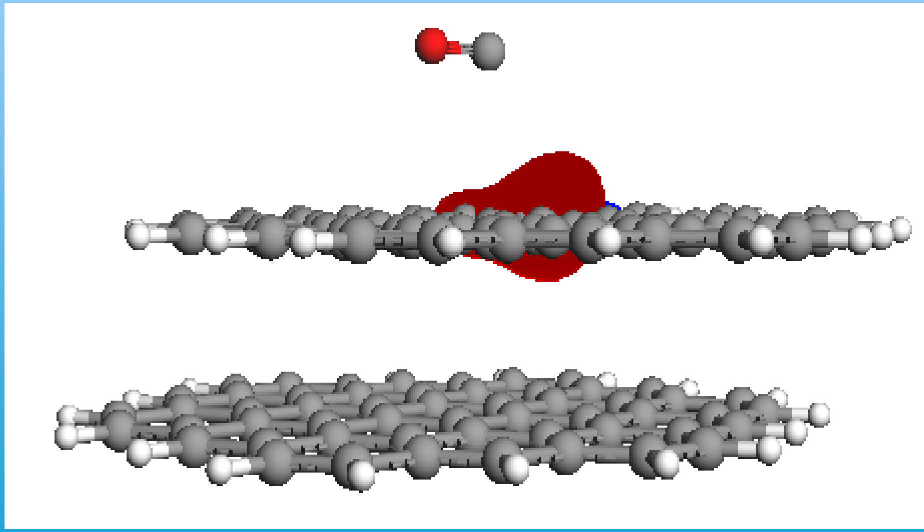
-0.15082 eV

LUMO

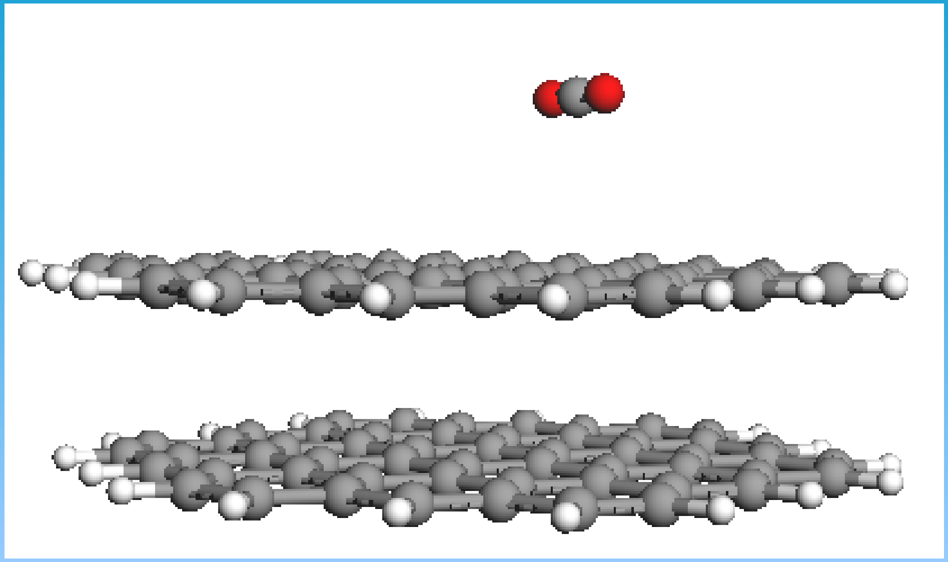
-0.09151 eV

E_a = 0.158 eV

HOMO-LUMO Gap = 1.614 eV



CO2-Graphene Optimized Structure



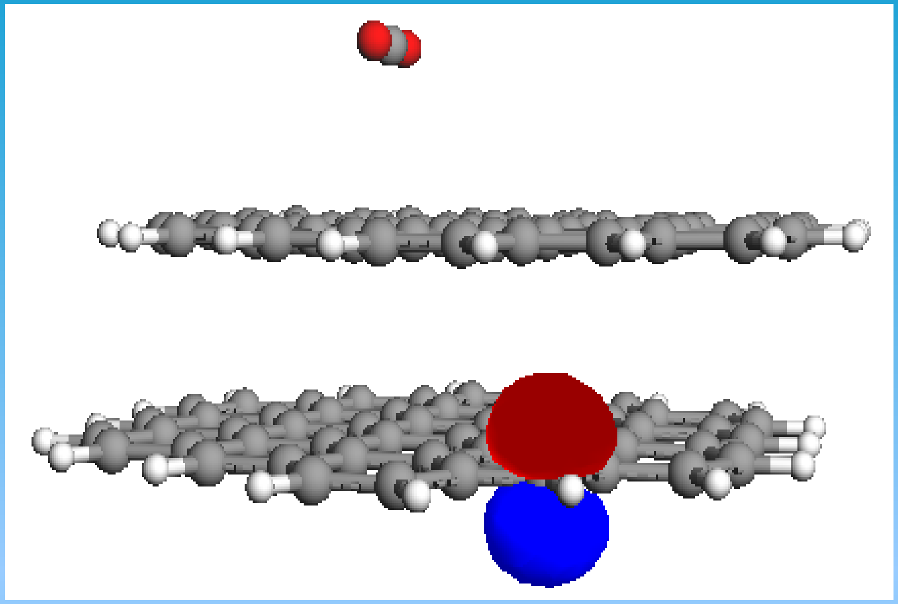
Distance Data

Distance Between layers: 3.23 Å

Bond Length C-C: 1.42 Å

CO_2_Graphene Distance: 3.14 Å

CO2-Graphene Frontier Orbitals

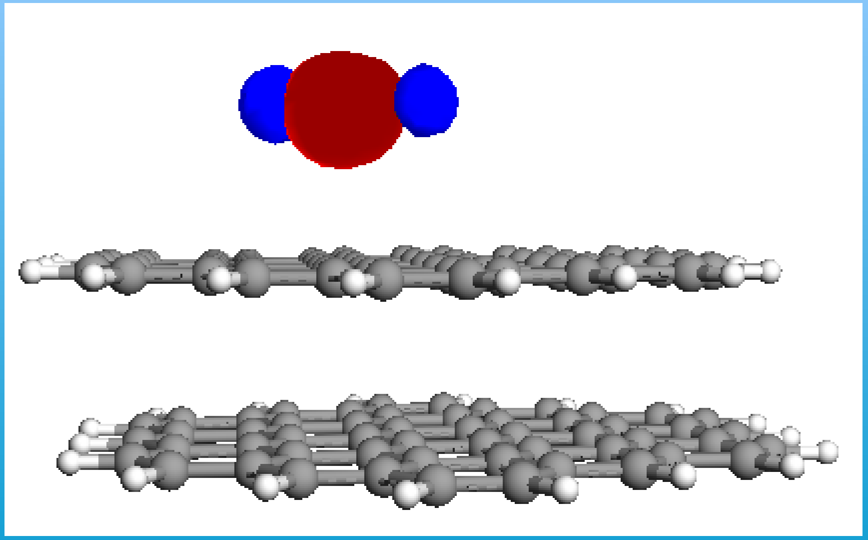


HOMO
-0.15082 eV

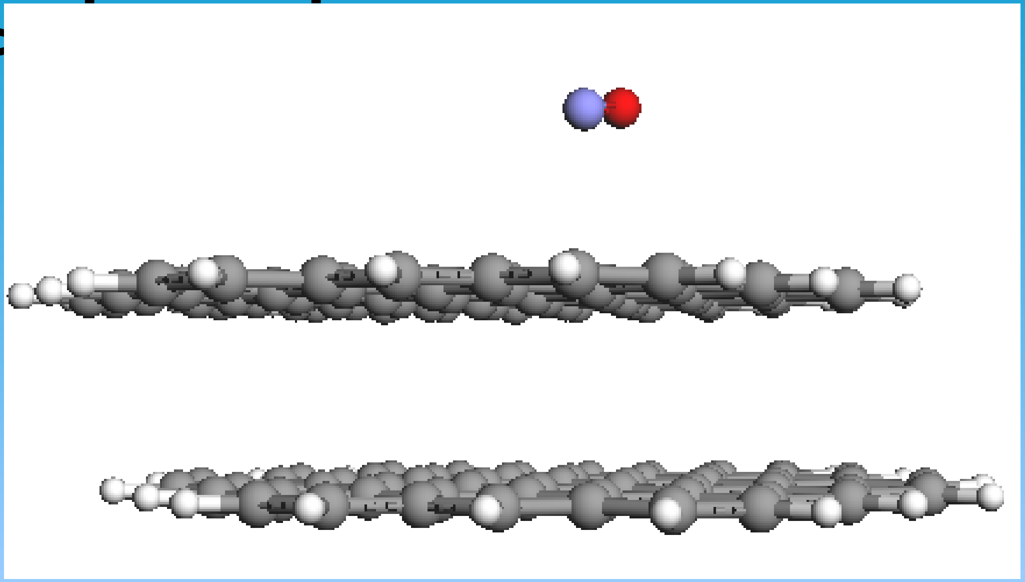
LUMO
-0.09146 eV

Ea = 0.240 eV

HOMO-LUMO Gap = 1.615 eV



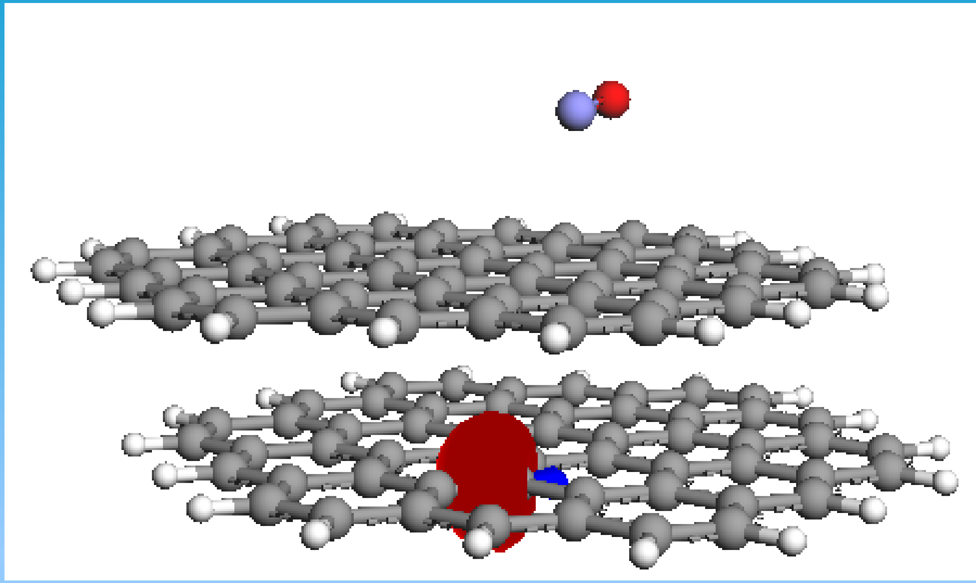
NO-Graphene Optimized Struc



Distance Data

Distance Between layers:	3.23 Å
Bond Length C-C:	1.42 Å
NO_Graphene Distance:	3.00 Å

NO-Graphene Frontier Orbitals



HOMO

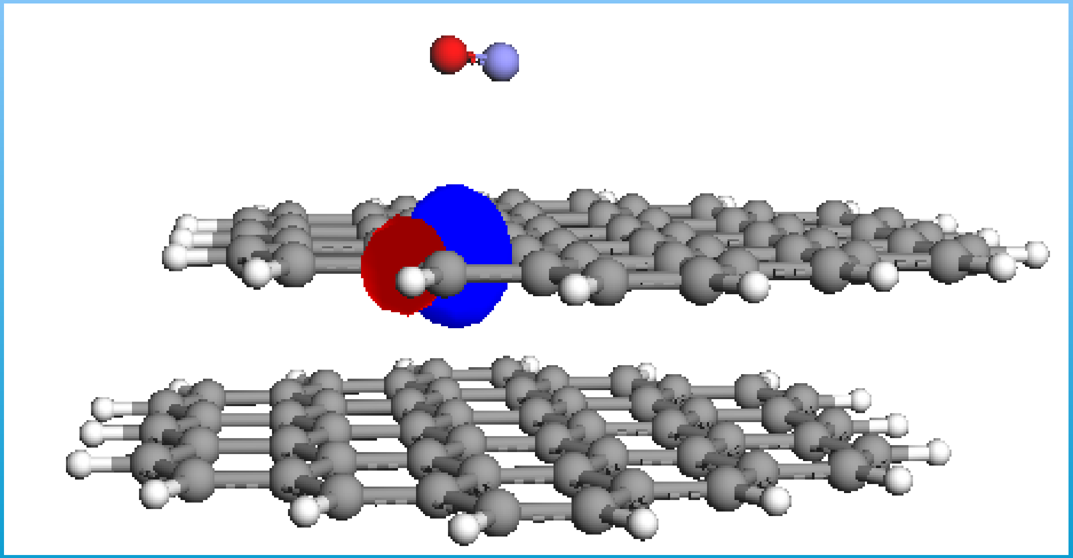
-0.15103 eV

LUMO

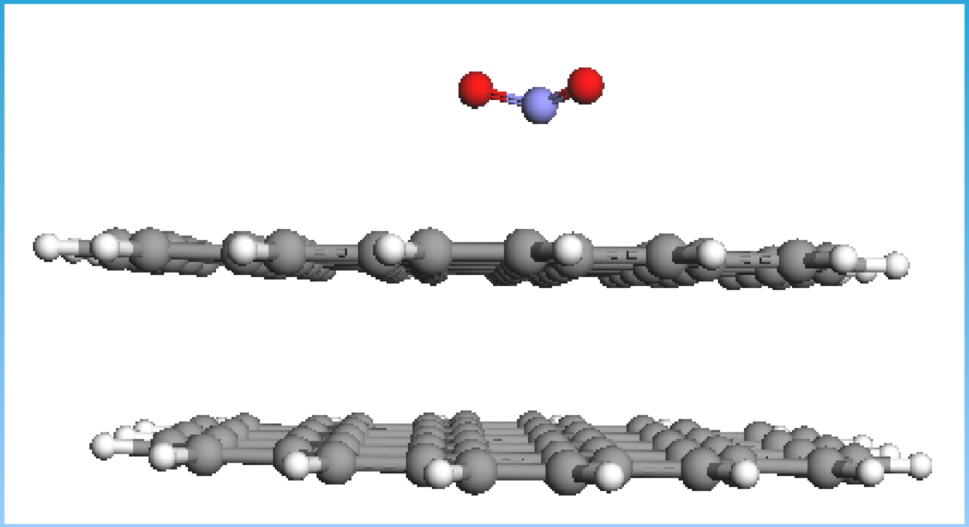
-0.09822 eV

Ea = 0.180 eV

HOMO-LUMO Gap = 1.437 eV



NO2-Graphene Optimized Structure



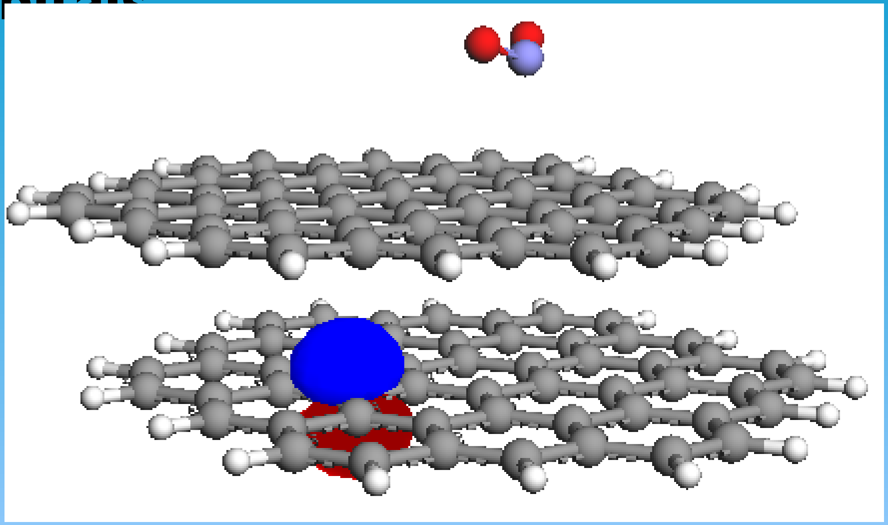
Distance Data

Distance Between layers: 3.23 Å

Bond Length C-C: 1.42 Å

NO2_Graphene Distance: 3.13 Å

NO2-Graphene Frontier Orbitals



HOMO

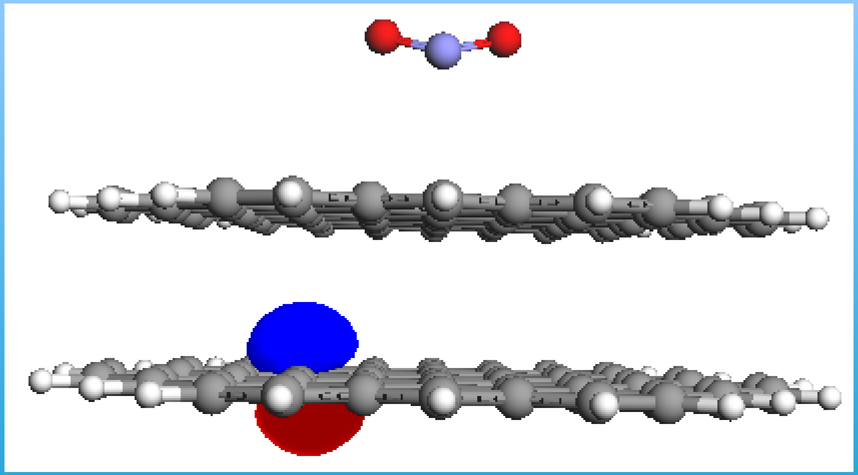
-0.15135 eV

LUMO

-0.14688 eV

Ea = 0.238 eV

HOMO-LUMO Gap = 0.122 eV



Conclusions

In contrast with the inifinite graphene sheet, which gap is zero when irradiated, our molecular model observes a gap of 1.618 eV .

In all cases, there was a reduction in the HOMO-LUMO Gap when the molecule is adsorved in the policircumcoronene.

Fisisorption was found for all the molecule-graphene systems (policircumcoronene), observing the bigger adsorption energy in the carbon dioxide and smaller for the carbon monoxide following the next order:

$CO < NO < NO_2 < CO_2$

These trends in the adsorption energy order are in agreement with trends reported by other authors except that such authors report smaller energy values (Leenaerts, et. al; 2008).

Conclusions

Analyzing the values for the frontier orbitals HOMO and LUMO energy results, it is observed that the order for the HOMO would be as follows:

NO_2<NO <CO
CO_2

A similar trend is observed in the LUMO orbital, except with one small difference between CO and CO_2

NO_2 <NO <CO< CO_2

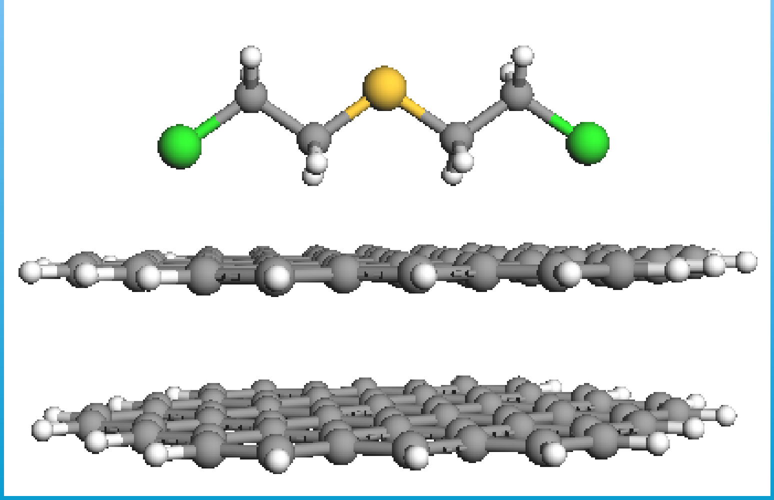
Work under Development

Mustard Gas

•It is important to study the interaction of bilayer graphene with other gases such as mustard gas.

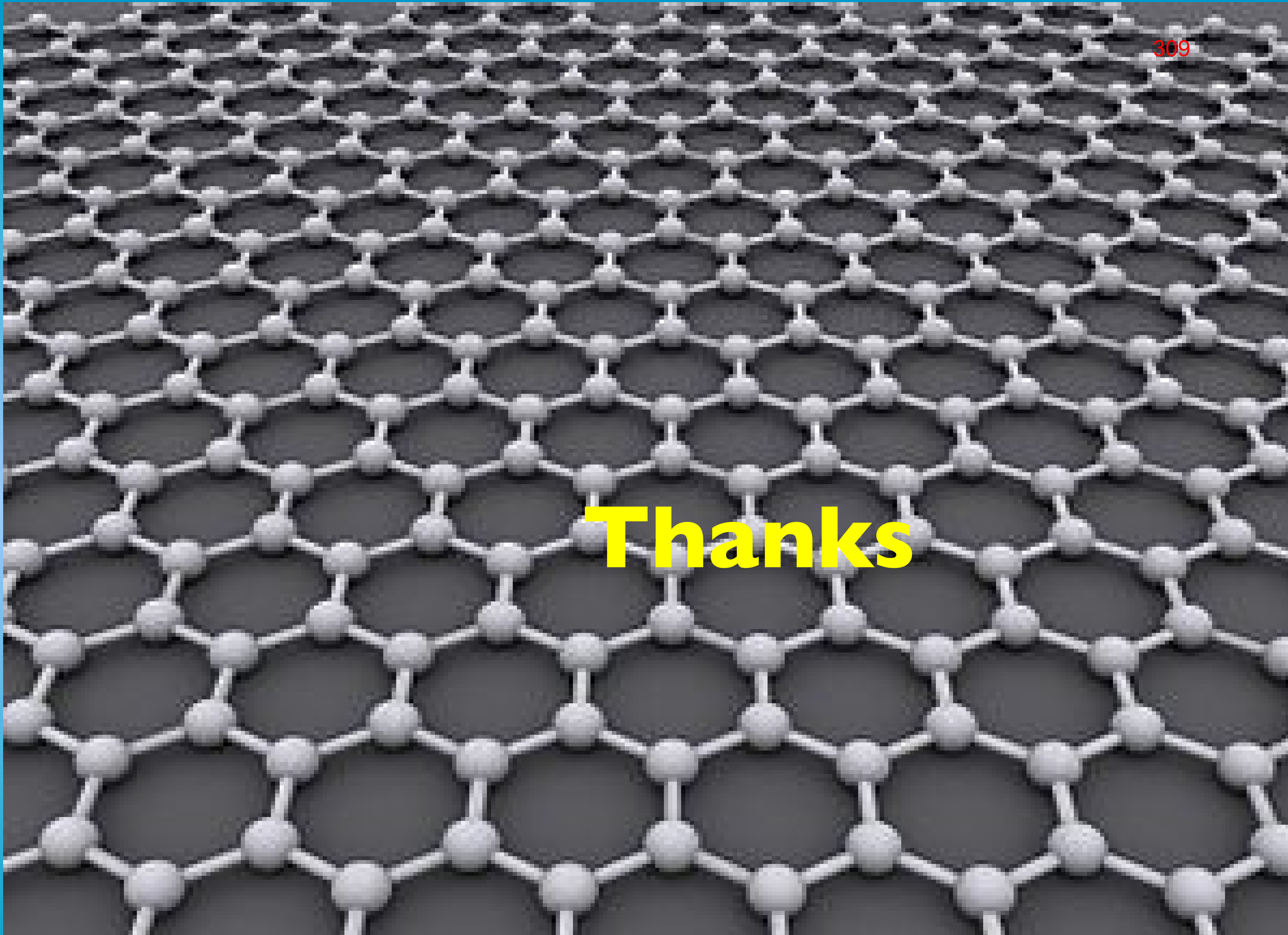
• This gas is used fundamentally in war weapons, produces death by agonic suffocation, as a gas can be odorless.

• It is a bigger molecule which can help in the characterization of bigger gas molecules placed over graphene.



Aknowledgements

Thanks to UNAM supercomputing for the use of Kanbalam supercomputer.
For the computation time for the realization of this work.


Thanks

"Comparative Human Microbiome Analysis"

Remote Video Talk to
CICESE Big Data, Big Network Workshop
Ensenada, Mexico
October 10, 2013

Dr. Larry Smarr
Director, California Institute for Telecommunications and Information Technology
Harry E. Gruber Professor,
Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD
http://lsmarr.calit2.net

J. Craig Venter INSTITUTE

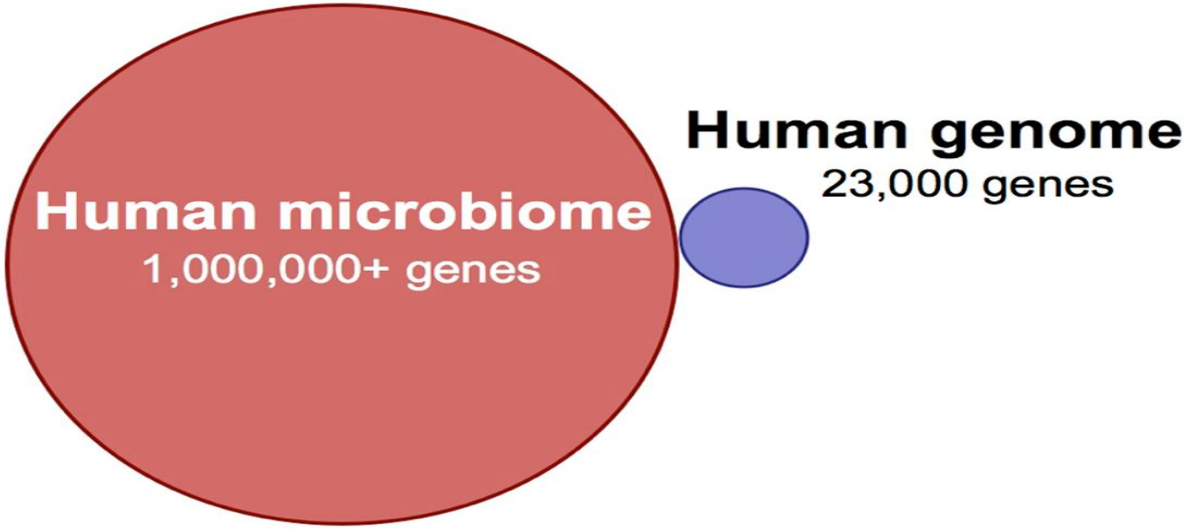
Abstract

We are carrying out very deep metagenomic sequencing of human gut microbiomes from healthy subjects and from people with the autoimmune Inflammatory Bowel Disease. We compare one subject with IBD to metagenomic datasets downloaded from the NIH Human Microbiome Project repository, including 35 healthy subjects and 20 with IBD. We also analyze the changes in this one subject over multiple times, including comparing before and after drug therapy. The dataset of Illumina short reads for one person is ~10GB. The total comparison dataset contains ~0.5 trillion DNA bases. These Big Data had to be moved across the network to the San Diego Supercomputer Center where over 200,000 cpu-hours were consumed in the analysis and then back to Calit2 where a 64 megapixel wall was used for visual analysis. This approach could be extended for cross-border comparisons of human gut microbiomes to examine differences in food intake and various disease states.

Larry Smarr is the Harry E. Gruber Professor in the Department of Computer Science and Engineering of the Jacobs School of Engineering at UC San Diego. He was the founding director of the California Institute for Telecommunications and Information Technology in 2000 and of the National Center for Supercomputing Applications in 1985.

Weizhong Li currently leads a group of researchers funded by NIH and NSF at the Center for Research in Biological System in UC San Diego. He has more than 20 years of experience in bioinformatics, computational biology, and computational chemistry.

Your Body Has 10 Times As Many Microbe Cells As Human Cells



99% of Your DNA Genes Are in Microbe Cells Not Human Cells

Inclusion of the Microbiome Will Radically Change Medicine

J. Craig Venter™ INSTITUTE

Gut Microbiome Metagenomic Datasets Comparing Healthy and Diseased States

Sample Group	Samples	Average PE reads per sample	Total PE reads	Selected samples	Average filtered high quality PE reads per selected sample
LS	3	151,401,139	454,203,418	3	119,694,209
CD	15	91,477,052	1,372,155,785	15	57,624,823
UC	11	14,330,989	157,640,888	6	2,905,681
HE	154	68,609,164	10,565,811,331	34	39,692,865
Total	183	68,578,204	12,549,811,422	58	44,662,870

One "Read" = 100 DNA Bases
Total of 12.5 Billion Reads!

J. Craig Venter INSTITUTE

Source: Weizhong Li, CRBS, UCSD

We Created a Reference Database Of Known Gut Genomes

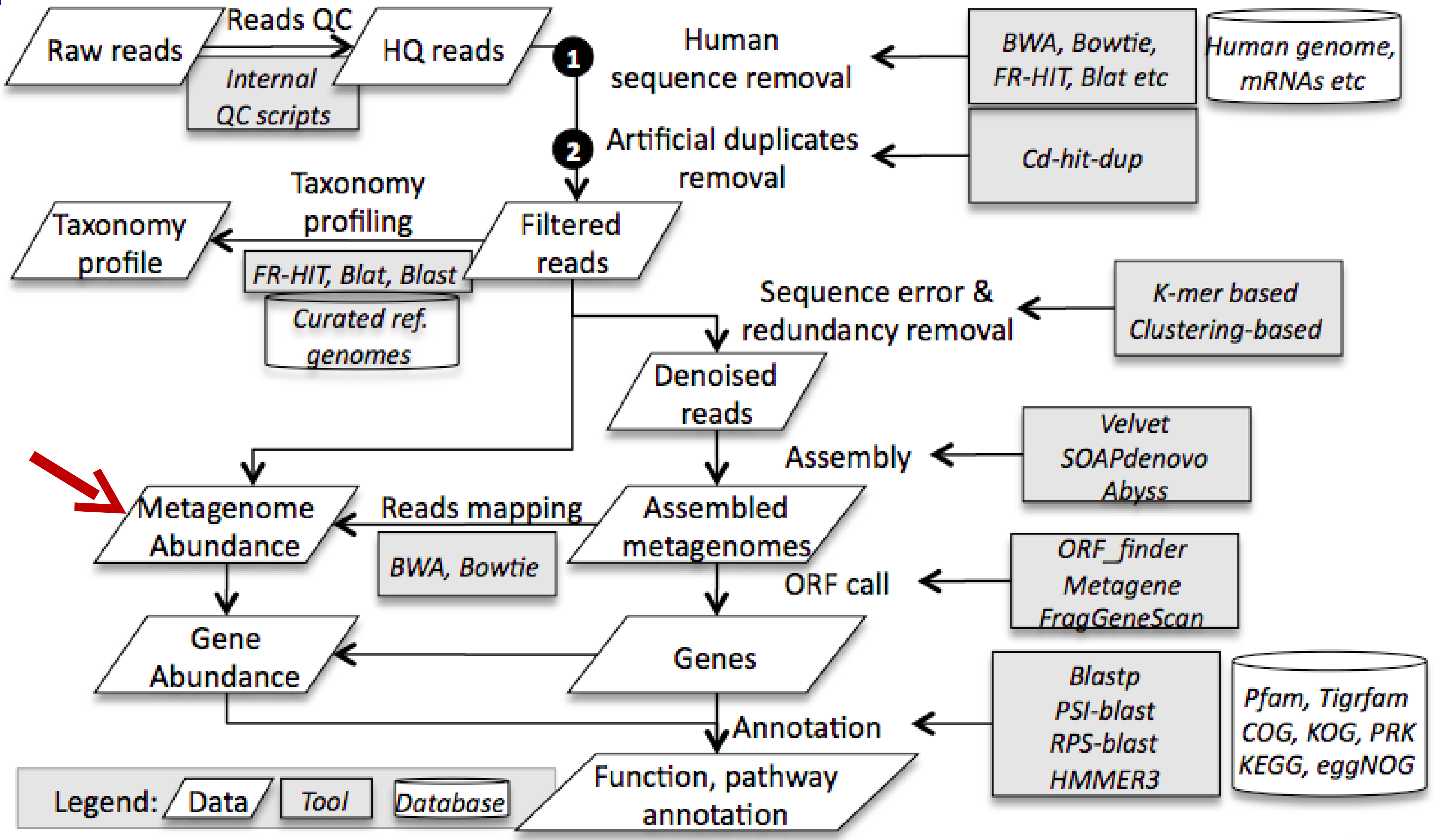
- **NCBI April 2013**
  - **2471 Complete + 5543 Draft Bacteria & Archaea Genomes**
  - **2399 Complete Virus Genomes**
  - **26 Complete Fungi Genomes**
  - **309 HMP Eukaryote Reference Genomes**
- **Total 10,741 genomes, ~30 GB of sequences**

Now to Align Our 12.5 Billion Reads Against the Reference Database

J. Craig Venter INSTITUTE

Source: Weizhong Li, Sitao Wu, CRBS, UCSD

Computational NextGen Sequencing Pipeline: From “Big Equations” to “Big Data” Computing



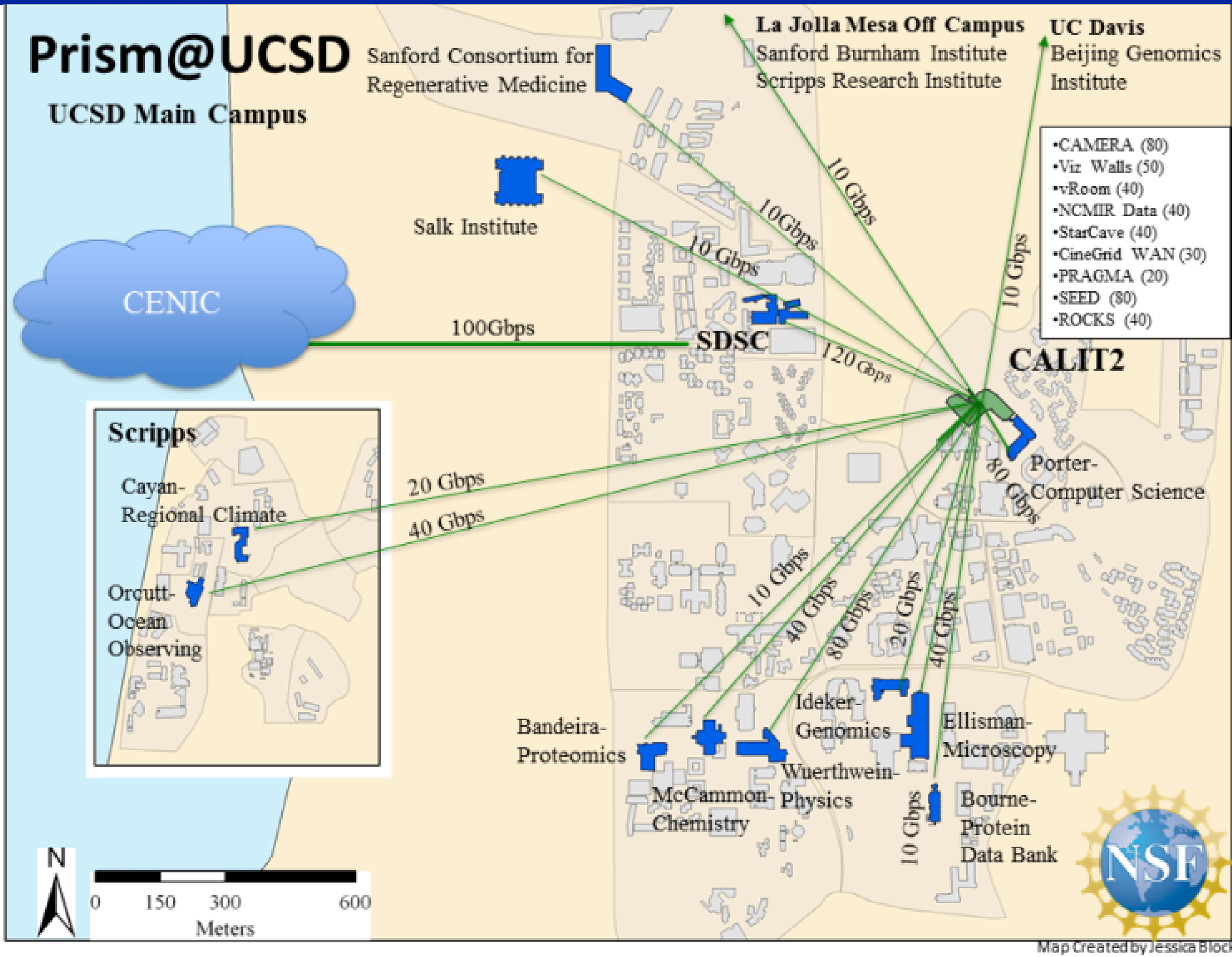
National Human Genome Research Institute

HMP

PI: (Weizhong Li, CRBS, UCSD):
NIH R01HG005978 (2010-2013, $1.1M)

CRBS

Creating a Big Data Freeway System: Coupling 'Omics Data Generators with Supercomputers



SDSC

Using Optical Fiber with 1000x Shared Internet Speeds

We Used SDSC's Gordon Data-Intensive Supercomputer to Analyze a Wide Range of Gut Microbiomes

- ~180,000 Core-Hrs on Gordon
  - KEGG function annotation: 90,000 hrs
  - Mapping: 36,000 hrs
    - Used 16 Cores/Node and up to 50 nodes
  - Duplicates removal: 18,000 hrs
  - Assembly: 18,000 hrs
  - Other: 18,000 hrs
- Gordon RAM Required
  - 64GB RAM for Reference DB
  - 192GB RAM for Assembly
- Gordon Disk Required
  - Ultra-Fast Disk Holds Ref DB for All Nodes
  - 8TB for All Subjects

SDSC

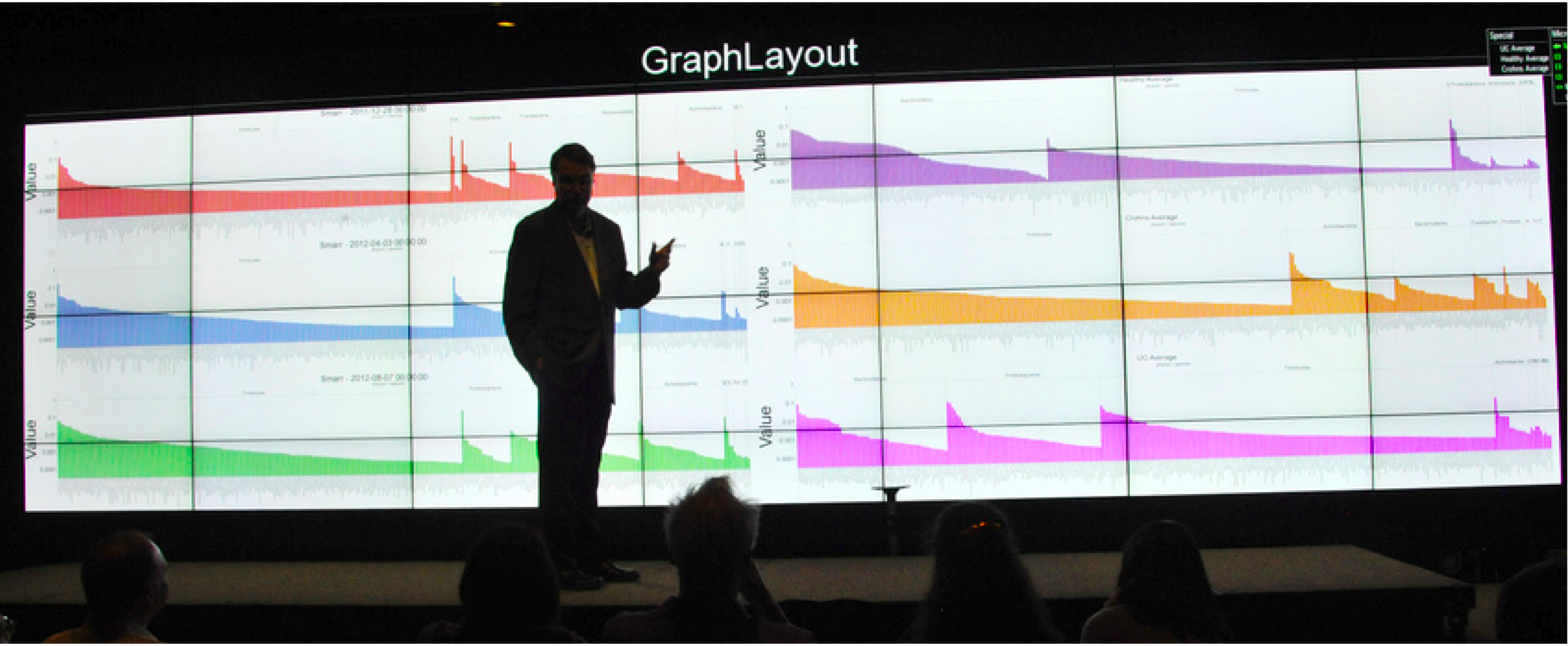
Enabled by a Grant of Time on Gordon from SDSC Director Mike Norman

NSF NIH HMP

J. Craig Venter INSTITUTE

CRBS it2

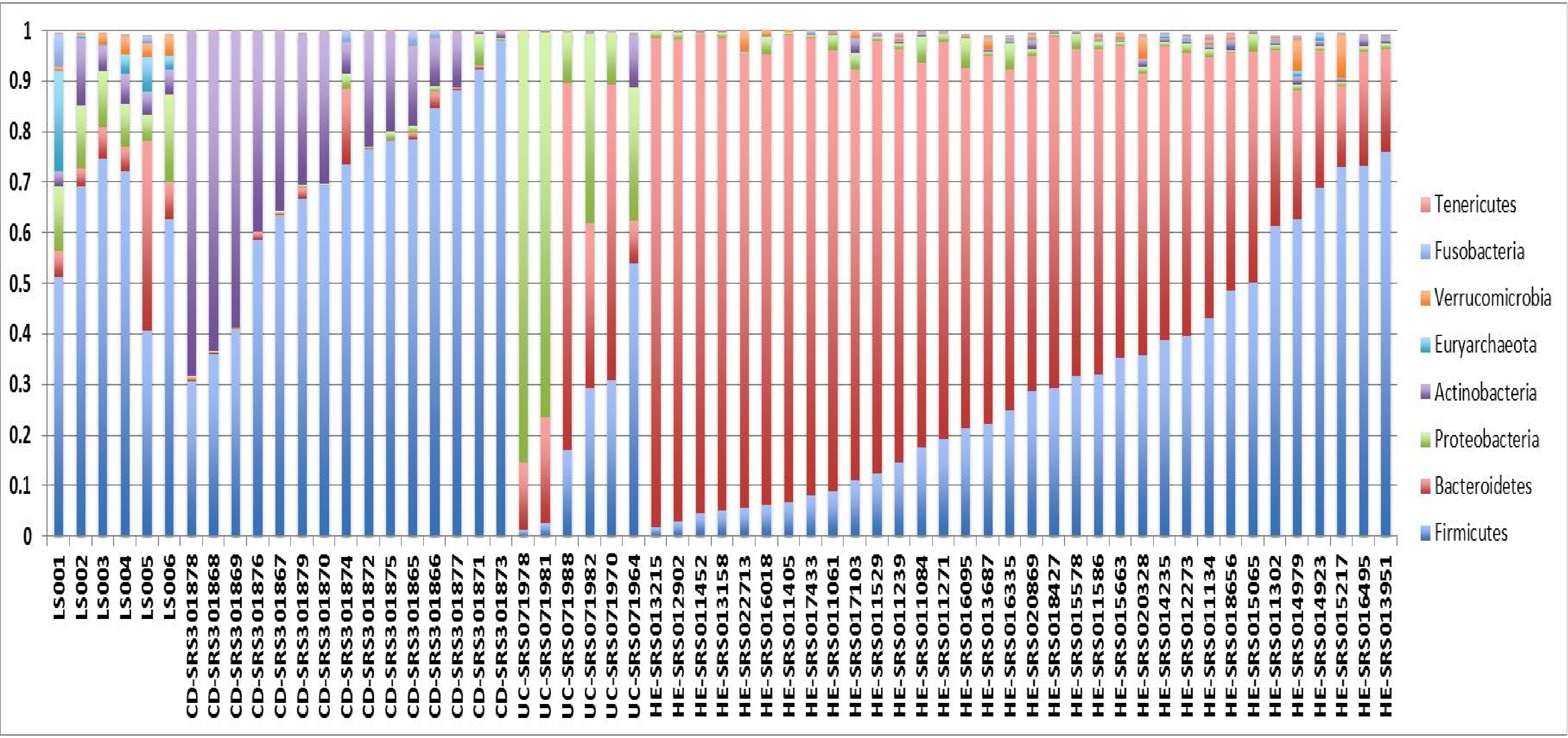
Comparing 3 LS Time Snapshots (Left) with Healthy, Crohn's, UC (Right Top to Bottom)



Calit2 VROOM-FuturePatient Expedition

Phyla Gut Microbial Abundance Without Viruses: LS, Crohn's, UC, and Healthy Subjects

Source: Weizhong Li, Sitao Wu, CRBS, UCSD



LS **Crohn's** **Ulcerative Colitis** **Healthy**

Toward Noninvasive Microbial Ecology Diagnostics

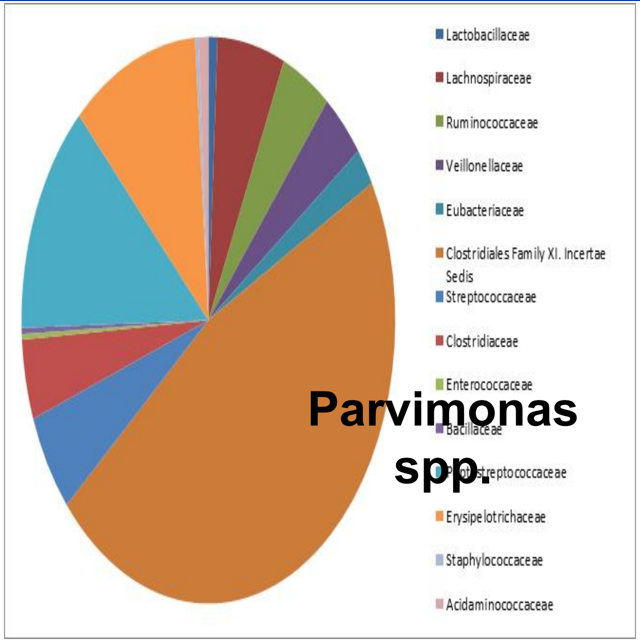
Lessons From Ecological Science:

Invasive Species Dominate After Major Species Destroyed

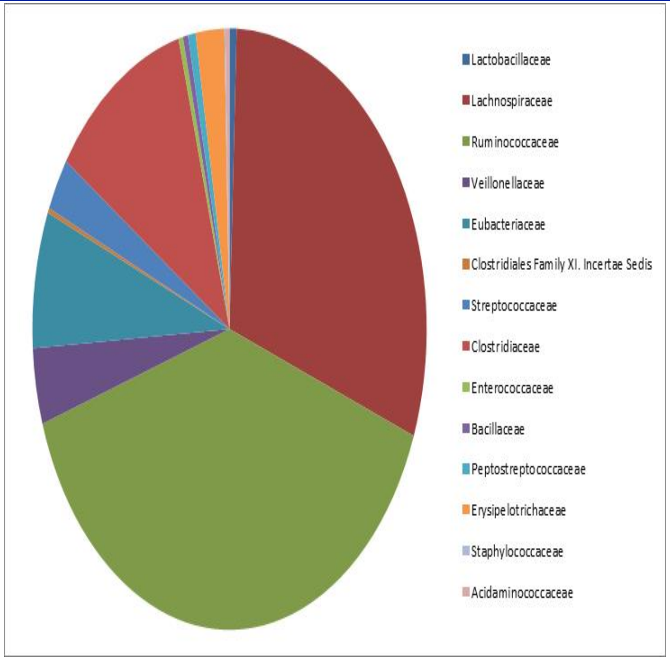
”In many areas following these burns invasive species are able to establish themselves, crowding out native species.”

Source: Ponderosa Pine Fire Ecology
http://cpluhna.nau.edu/Biota/ponderosafire.htm

Rare Firmicutes Bloom in Colon Disappearing After Antibiotic/Immunosuppressant Therapy

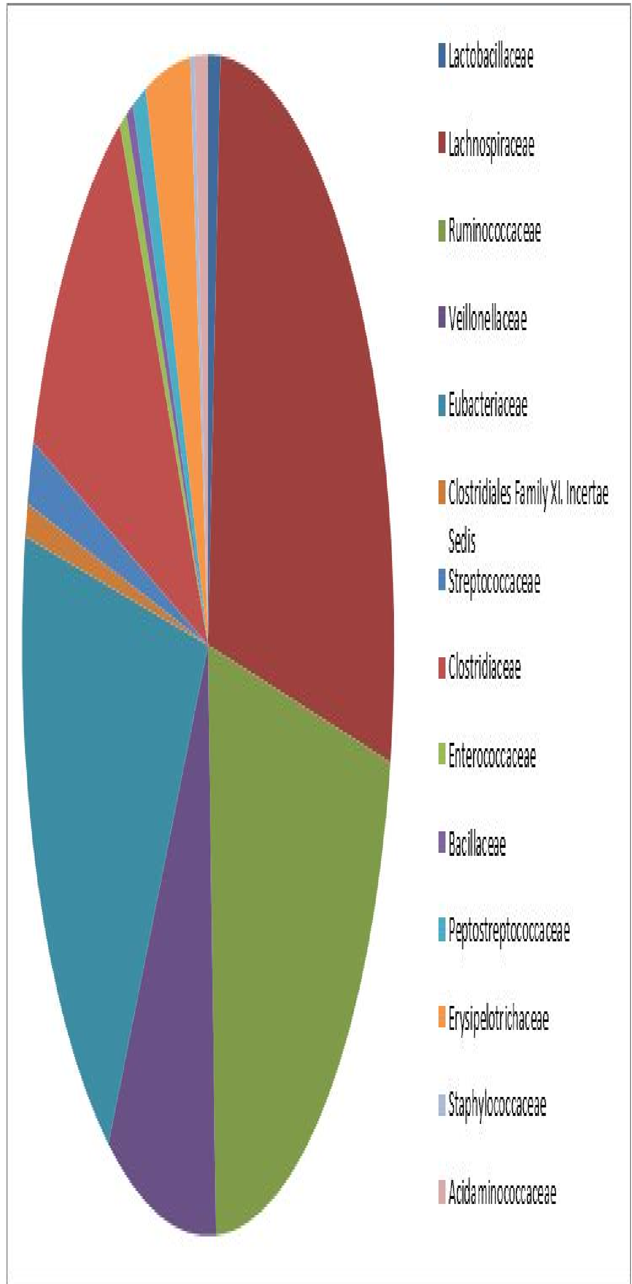


LS Time 1

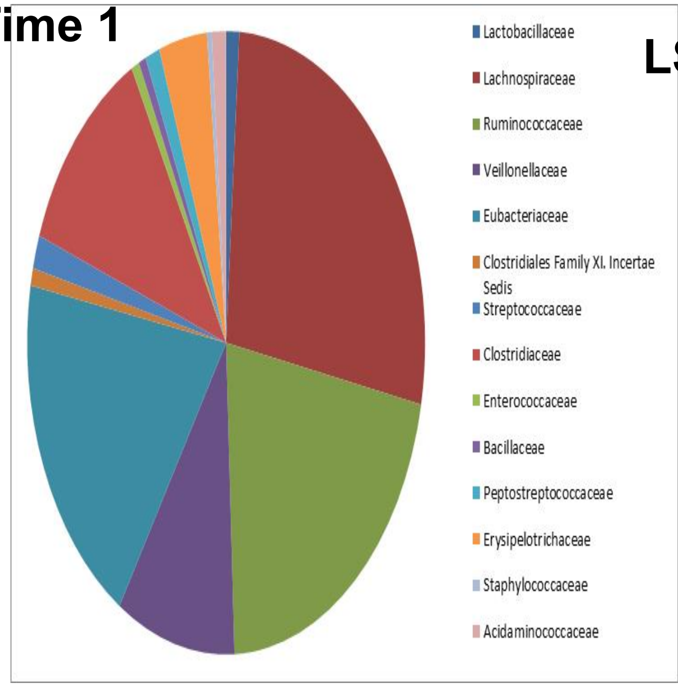


LS Time 2

Firmicutes Families



Healthy Average



Thanks to Our Great Team!

UCSD Metagenomics Team

Weizhong Li
Sitao Wu

JCVI Team

Karen Nelson
Shibu Yooseph
Manolito Torralba

Calit2@UCSD Future Patient Team

Jerry Sheehan
Tom DeFanti
Kevin Patrick
Jurgen Schulze
Andrew Prudhomme
Philip Weber
Fred Raab
Joe Keefe
Ernesto Ramirez

SDSC Team

Michael Norman
Mahidhar Tatineni
Robert Sinkovits

BIG DATA
BIG NETWORK
Big Opportunities for Collaboration
between Mexico and California
10-11 Oct 2013 • Ensenada, Mexico

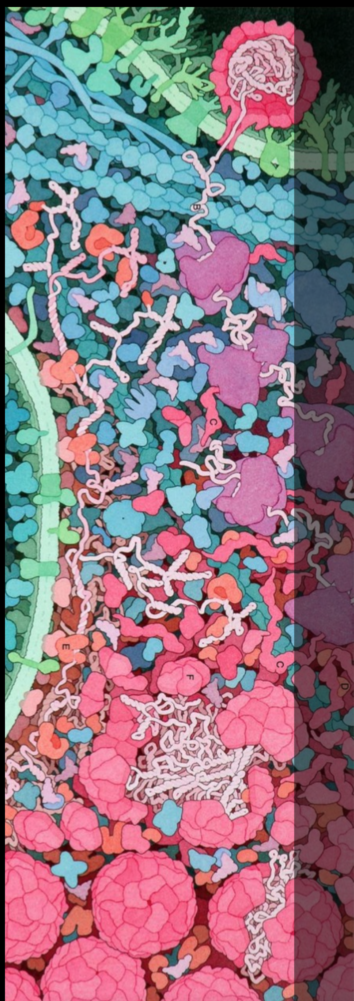
Computational Biology: Computation-Based Discovery in Molecular/Structural Biology

Nobels 2013: Chemistry prize goes to computer modellers

- This year's prize is about taking the chemical experiments to cyberspace
- Chemists used to create models of molecules using plastic balls and sticks. Today, the modelling is carried out in computers
- Computers were made for biology. Biology would never have advanced as it did without the dramatic increase in computer power and availability

Mauricio
Carrillo-Tripp

Biomolecular Diversity Laboratory



The cell is a complex set of interacting biomolecules which show a wide *diversity* of size, shape, properties and functions. Such functions are mainly governed by two related factors: their molecular composition / structure and their thermodynamic behavior.

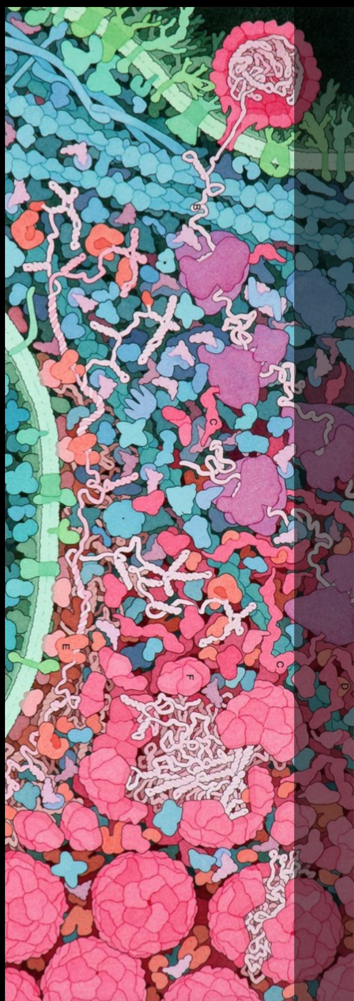
At the *bmd* Lab we study the molecular mechanisms of cellular processes at atomic level through a hybrid approach, trying to understand the relationships between structure, dynamics and function of the *machinery of life*.

D. S. Goodsell "The machinery of life"

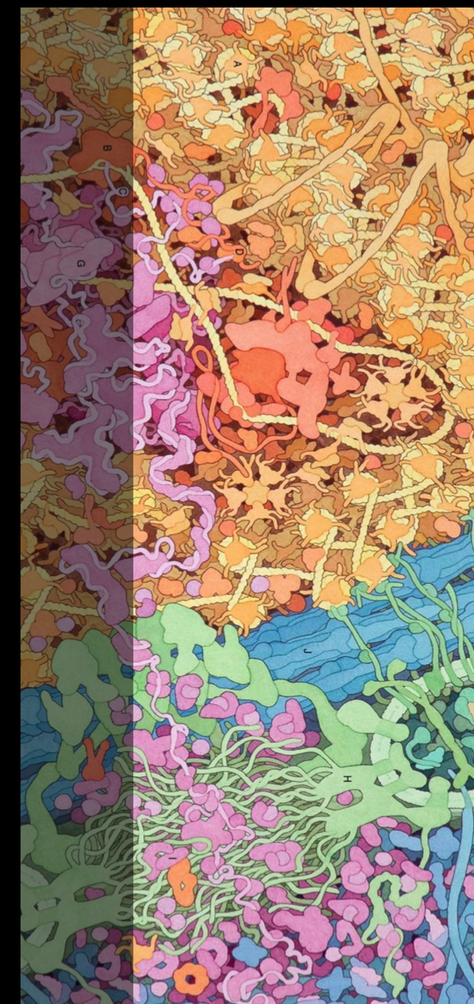
Laboratorio Nacional de Genómica para la Biodiversidad

Mauricio Carrillo-Tripp

Biomolecular Diversity Laboratory



Understanding Life
is a Big Challenge
having **Big Data**
that needs
Big Networks



D. S. Goodsell *"The machinery of life"*

Laboratorio Nacional de Genómica para la Biodiversidad

National Genomics Laboratory for Biodiversity

LANGEBIO

Cinvestav

National Genomics Laboratory for Biodiversity

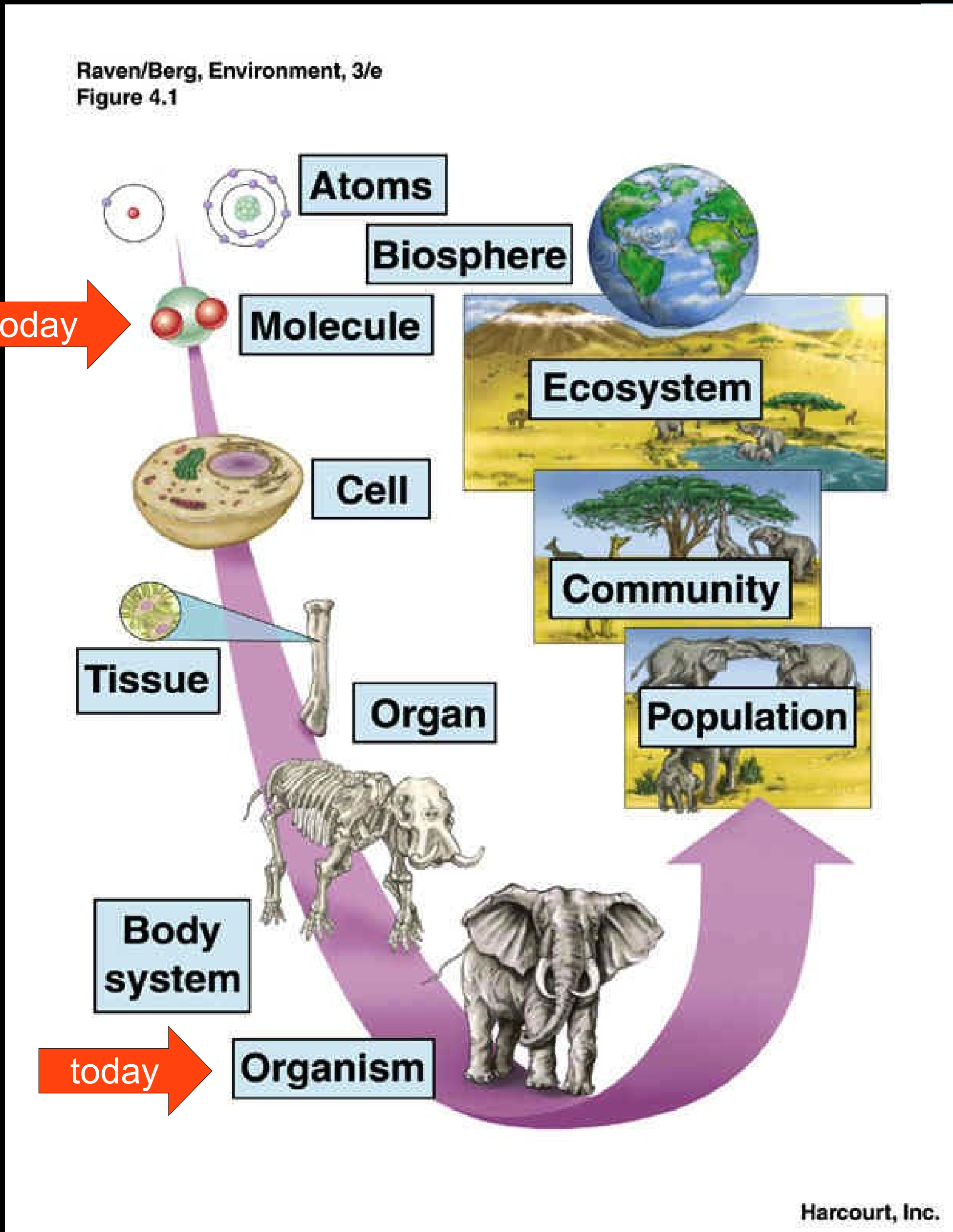
http://langebio.cinvestav.mx

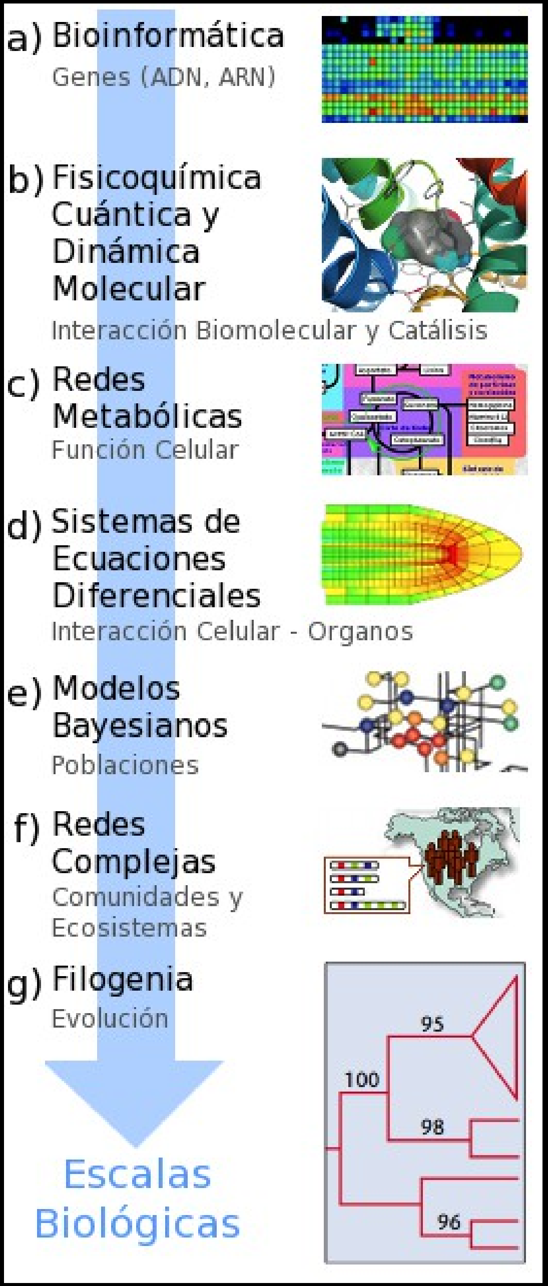
LANGEBIO Cinvestav

VISION LANGEBIO

- Make a difference in the study and sustainable use of National Biodiversity

Currently… ALL Life Scales included at Langebio:





Short-Read Archive (NCBI)

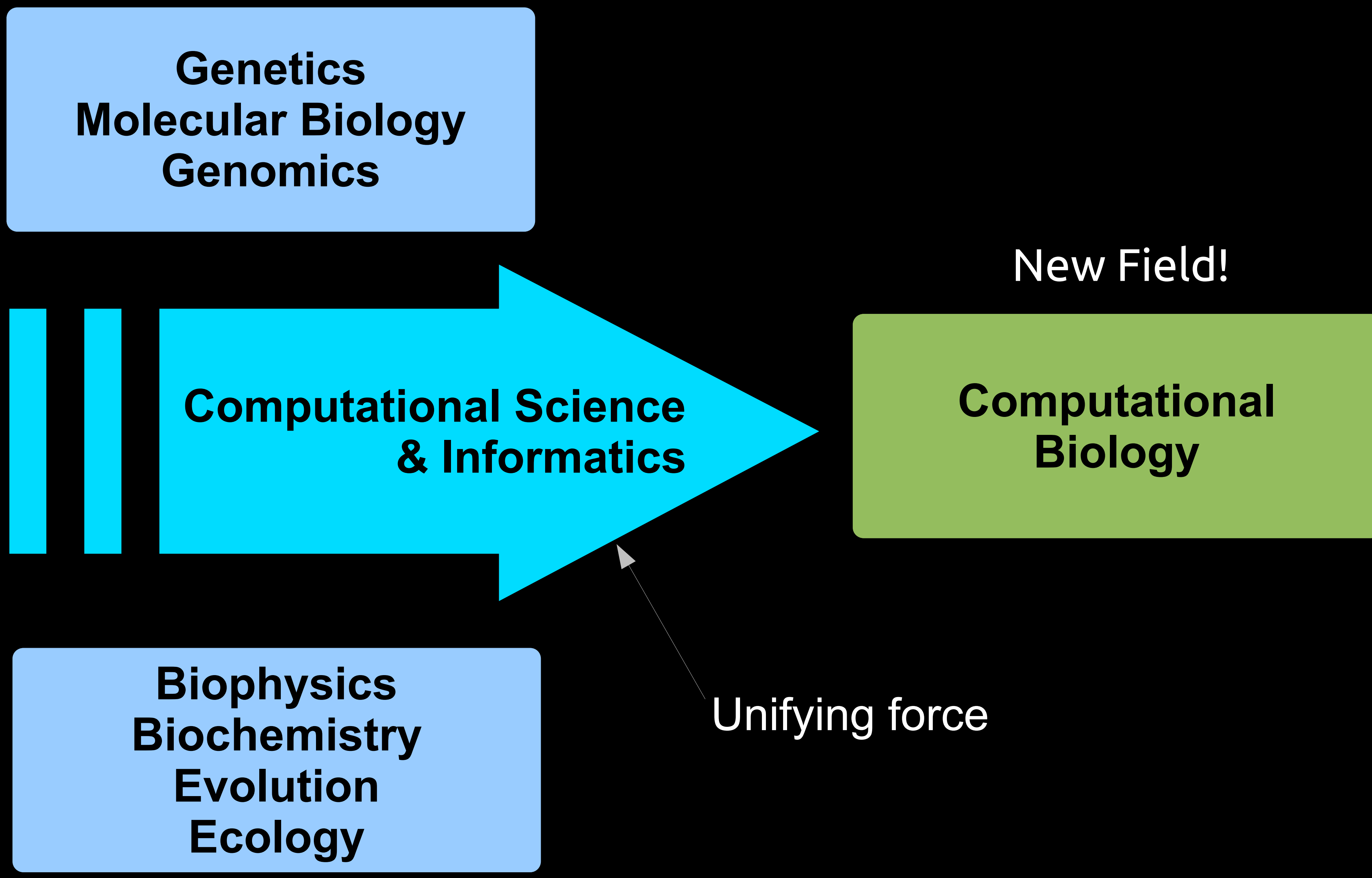
- Main international repository for NextGen sequencing results.
- Current available downloads are:
  - Metadata: 22 Gb
  - Sequences (compressed raw data): 650 Tb
- A large meta-analysis type project requires:
  - Searching metadata for particular experiments.
  - Downloading several Tb of raw data.

DATA / INFORMATION

Computational Biology
IS the BRIDGE

KNOWLEDGE

Biology IS a computational science!



BIG CHALLENGES IN BIOLOGY

NEW PARADIGMS

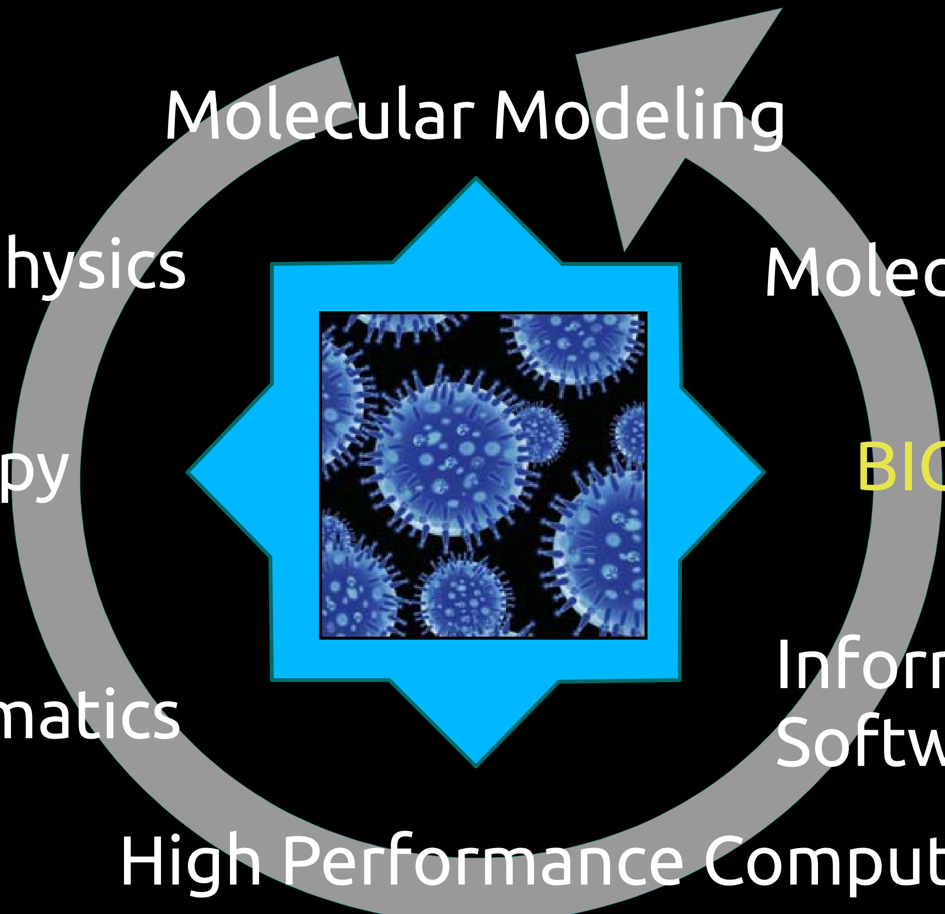
MULTIDISCIPLINARY APPROACH

COLLABORATIONS

334

How can we study the complex relationships between
Molecular Structure, Dynamics and Function
of the diverse cellular machinery?

Hybrid approaches: *in silico* + *in vitro* + *in vivo*



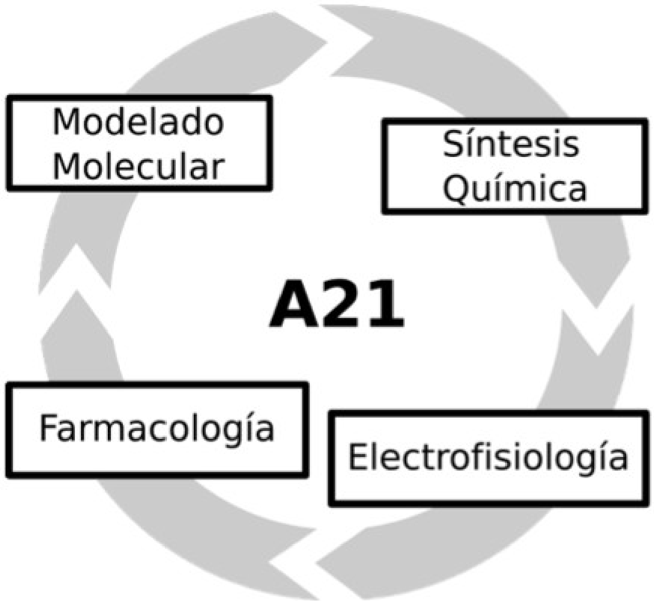
BASIC SCIENCE / KNOWLEDGE

TECHNOLOGY TRANSFER

NEW DRUG DEVELOPMENT

GOAL: GLOBAL HEALTH

Sistema de Innovación en Investigación: Estrategia Multidisciplinaria



1. Modelado Molecular

Se realizaron estudios comparativos a nivel atómico entre la AmB y los derivados, utilizando la metodología de modelado molecular *in silico*. Esto permitió generar hipótesis e indicios del mecanismo de acción de los antibióticos.

2. Síntesis Química

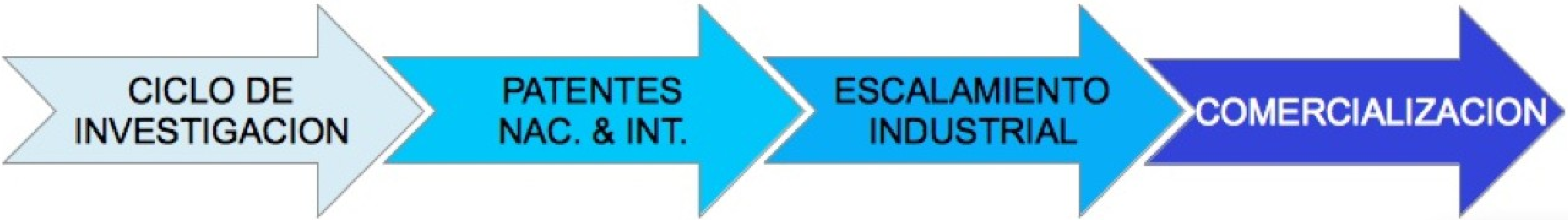
Se sustituyeron diversos grupos amida sintetizando una serie de análogos a la AmB, los cuales se caracterizaron a través de señales infrarojas representativas y espectrometría de masas de alta resolución.

3. Electrofisiología

Los derivados fueron probados para detectar transporte transmembranal *in vitro* usando bicapas lipídicas con colesterol y ergosterol, determinando así paso iónico.

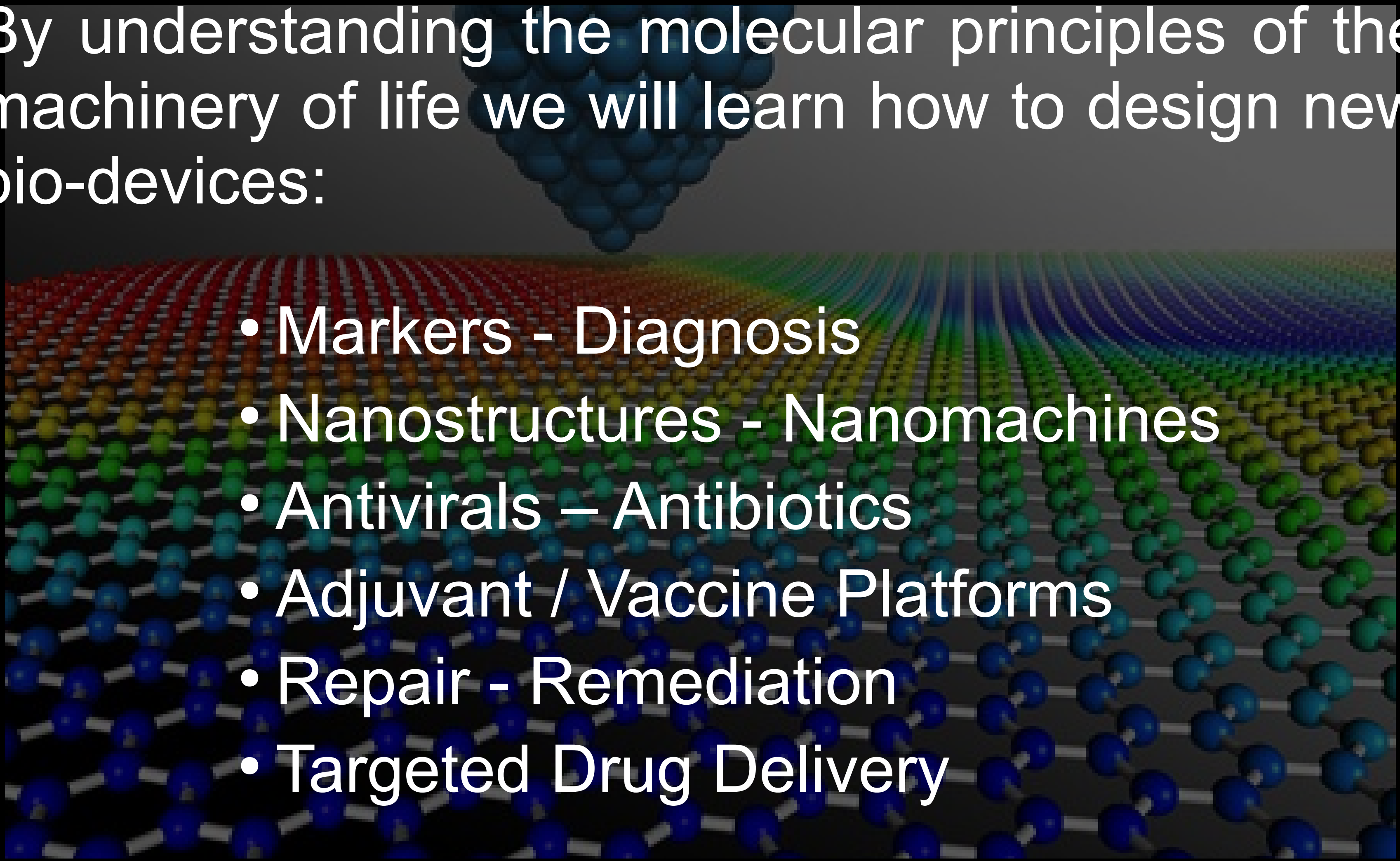
4. Farmacología

Aquellos derivados que presentaron conductividad fueron caracterizados a través de estudios farmacológicos en hongos, eritrocitos y tejido celular renal. El paso final fueron las pruebas pre-clínicas en mamíferos, usando ratones Balb-C.



...at the frontiers of Bioengineering

By understanding the molecular principles of the machinery of life we will learn how to design new bio-devices:



- Markers - Diagnosis
- Nanostructures - Nanomachines
- Antivirals – Antibiotics
- Adjuvant / Vaccine Platforms
- Repair - Remediation
- Targeted Drug Delivery

...at the frontiers of Life Sciences

Integrating the different components of the machinery of life:

Can we MODEL the nucleus, the mitochondria?

What about the whole cell, organ, organism... the World?

What would you do
if you were not afraid?

The Matrix (1999)

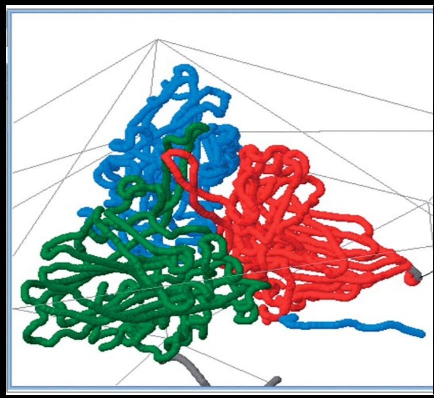
We are DATA isolated... We need Big Networks!

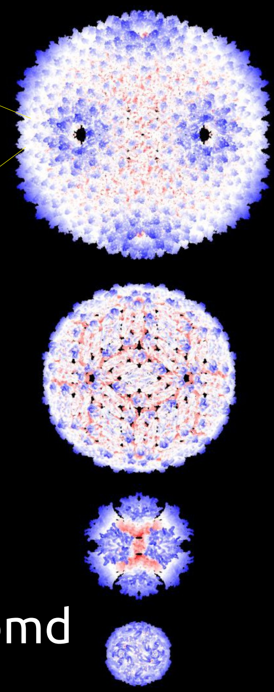
PARTNERS / HPC

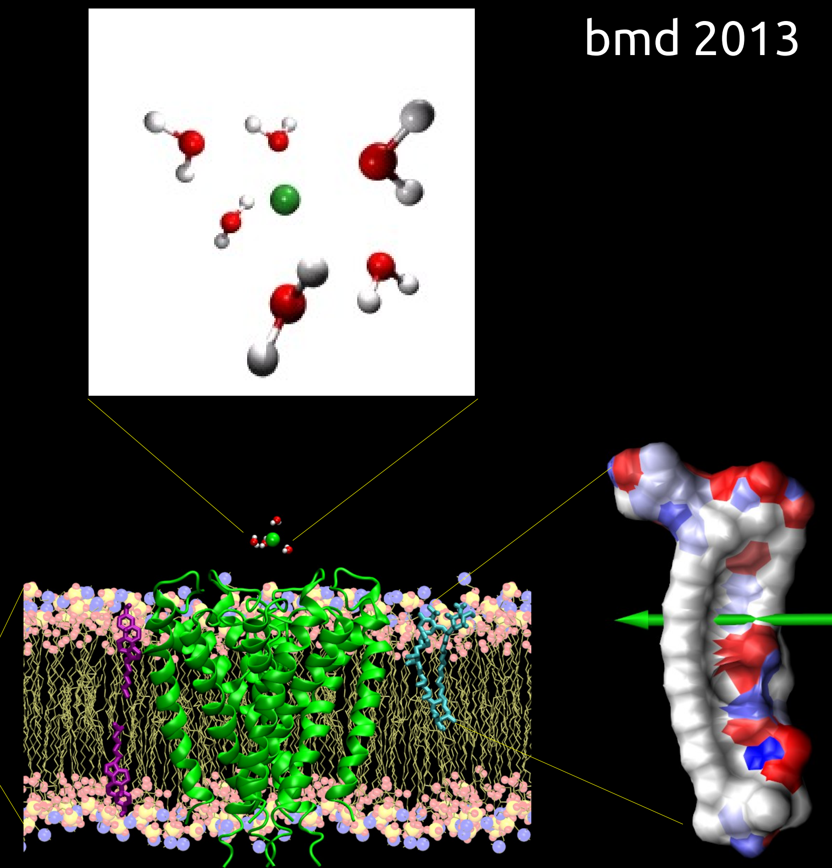
Argonne National Laboratory
The Scripps Research Institute
Groningen U
Vanderbilt U
many more...

CINVESTAV
UNAM
CIMAT
CNS

bmd 2013

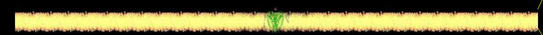


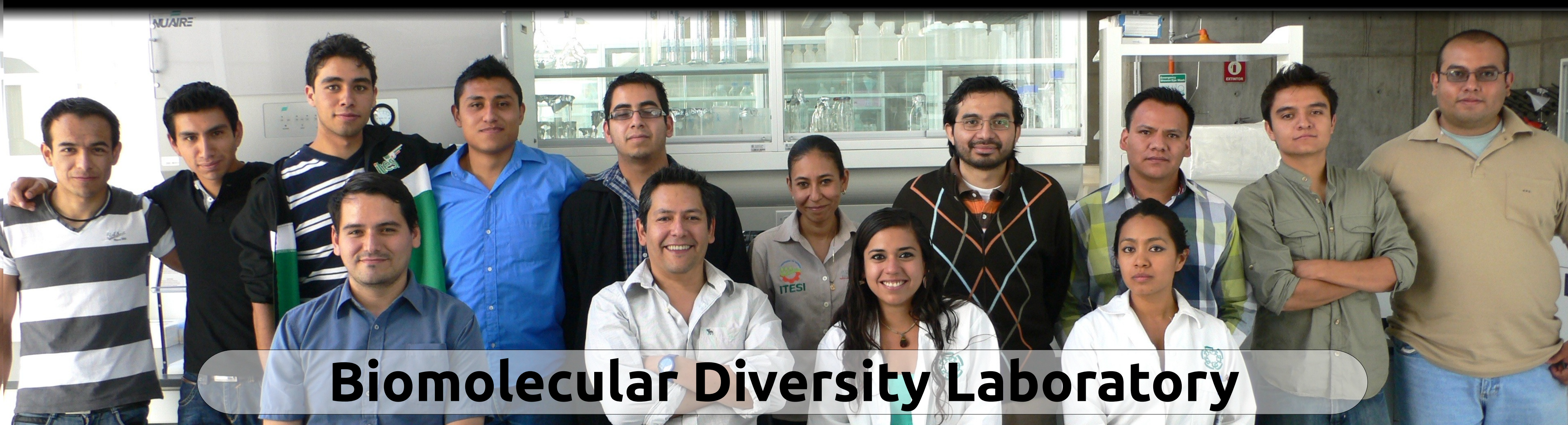




http://langebio.cinvestav.mx/bmd

trippm@langebio.cinvestav.mx





Biomolecular Diversity Laboratory

VIRTUAL MICROSCOPE

Langebio's sequencing capacity

- Illumina HiSeq 2000 (run-time 2 weeks)
  - Complete output: ~3 Tb
  - Minimum set of compressed sequences: ~550 Gb

- Ion Proton (run-time 4 hours)
  - Output: ~10 Gb

- Bandwidth limitations
  - Sending results to users (on or off-site)
  - Transfering results to compute clusters capable of processing (on or off-site).

- Compute limitations
  - Genome/Transcriptome assembly can require compute nodes with >256 Gb RAM

Workshop Big Data, Big Network: Big Opportunities for Collaboration Between Mexico and The United States

Current state of permanent GPS network in northern Baja California, Mexico. Challenges and opportunities for next decade

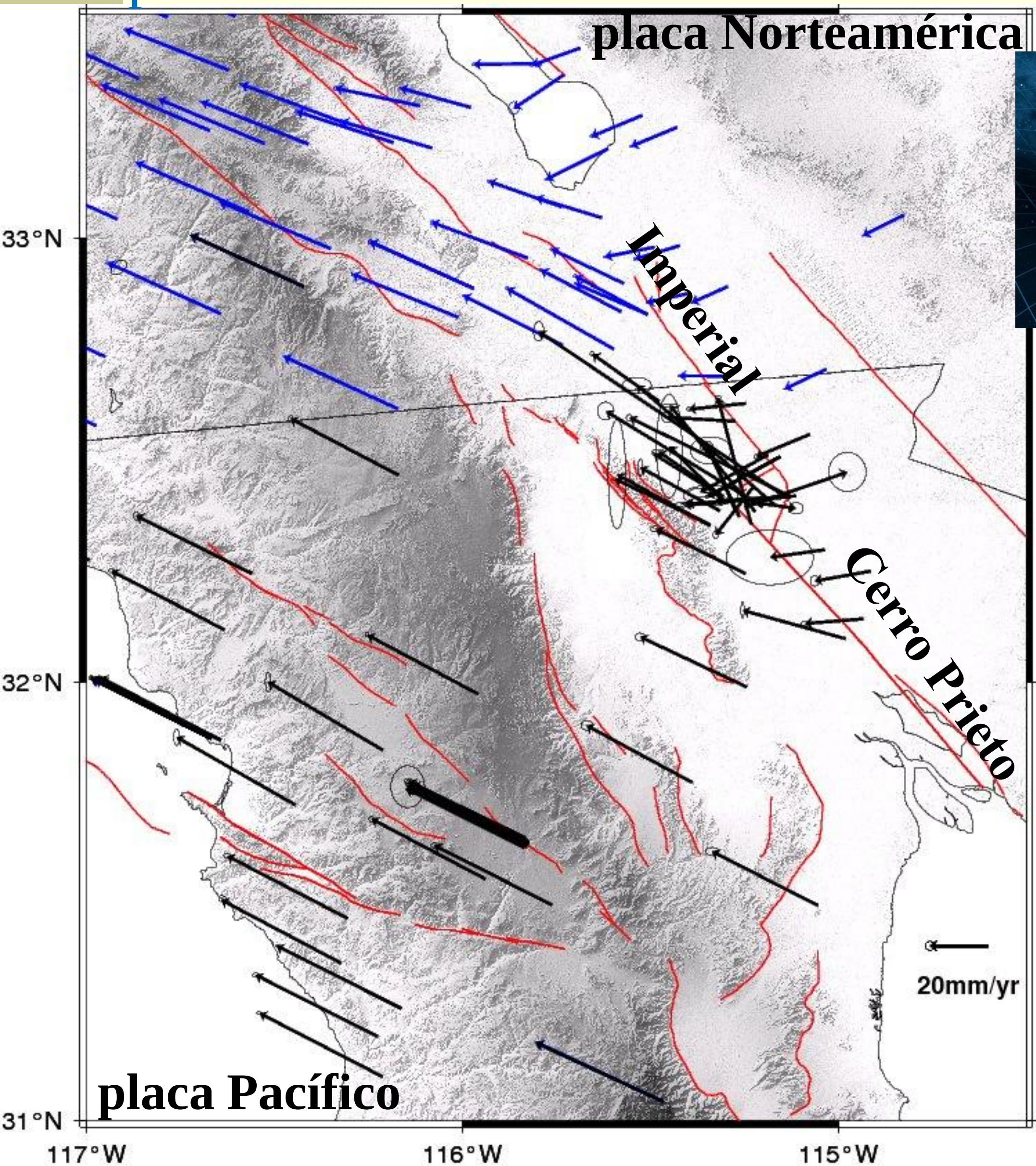
M.C. Javier Gonzalez-Garcia
Departamento de Sismología

M.C. Alejandro González-Ortega
Posgrado en Ciencias de la Tierra. Departamento de Sismología

Ensenada, Baja California, Mexico. 10 de Octubre de 2013

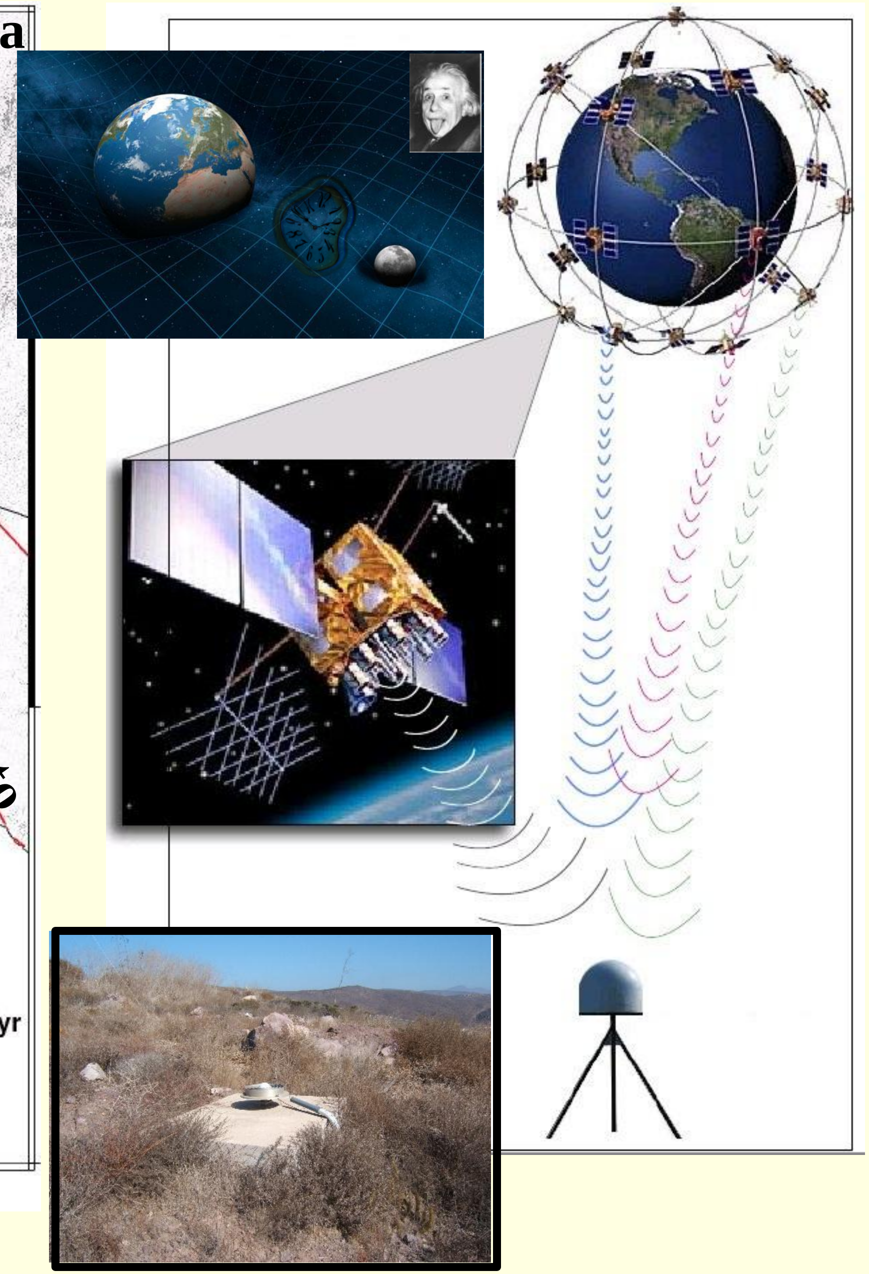
Que es el GPS y para que sirve en geodesia

Mapa de velocidades GPS. Norte B.C.

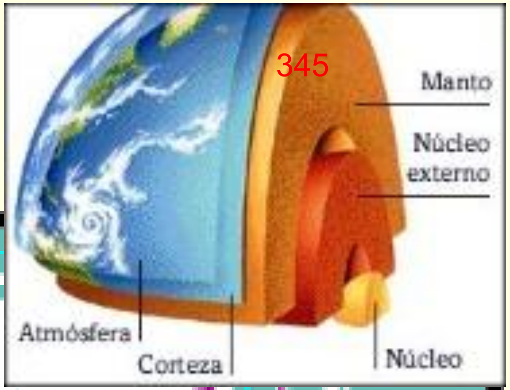


Datos de 1993 a 2008

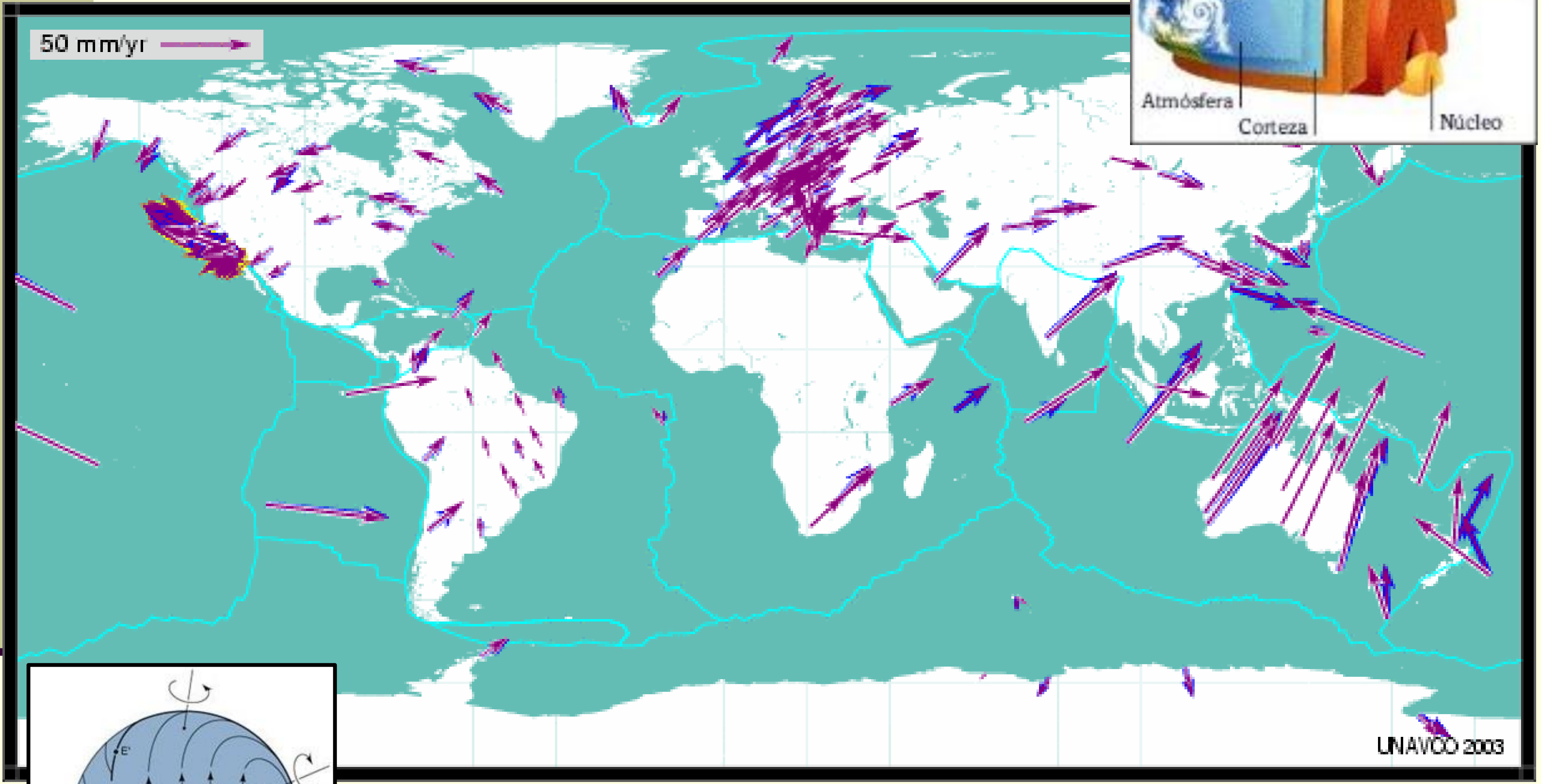
Sistema GPS

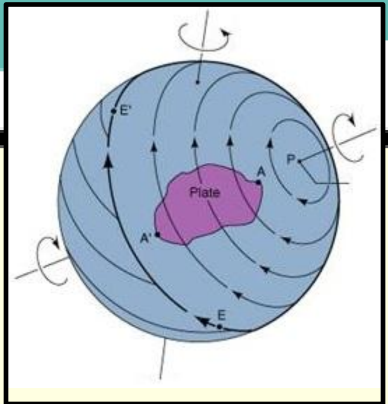


Introducción. Tectónica de placas



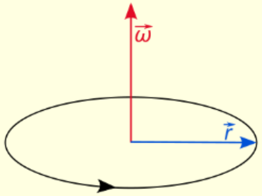
Mapa de velocidades GPS. www.unavco.org





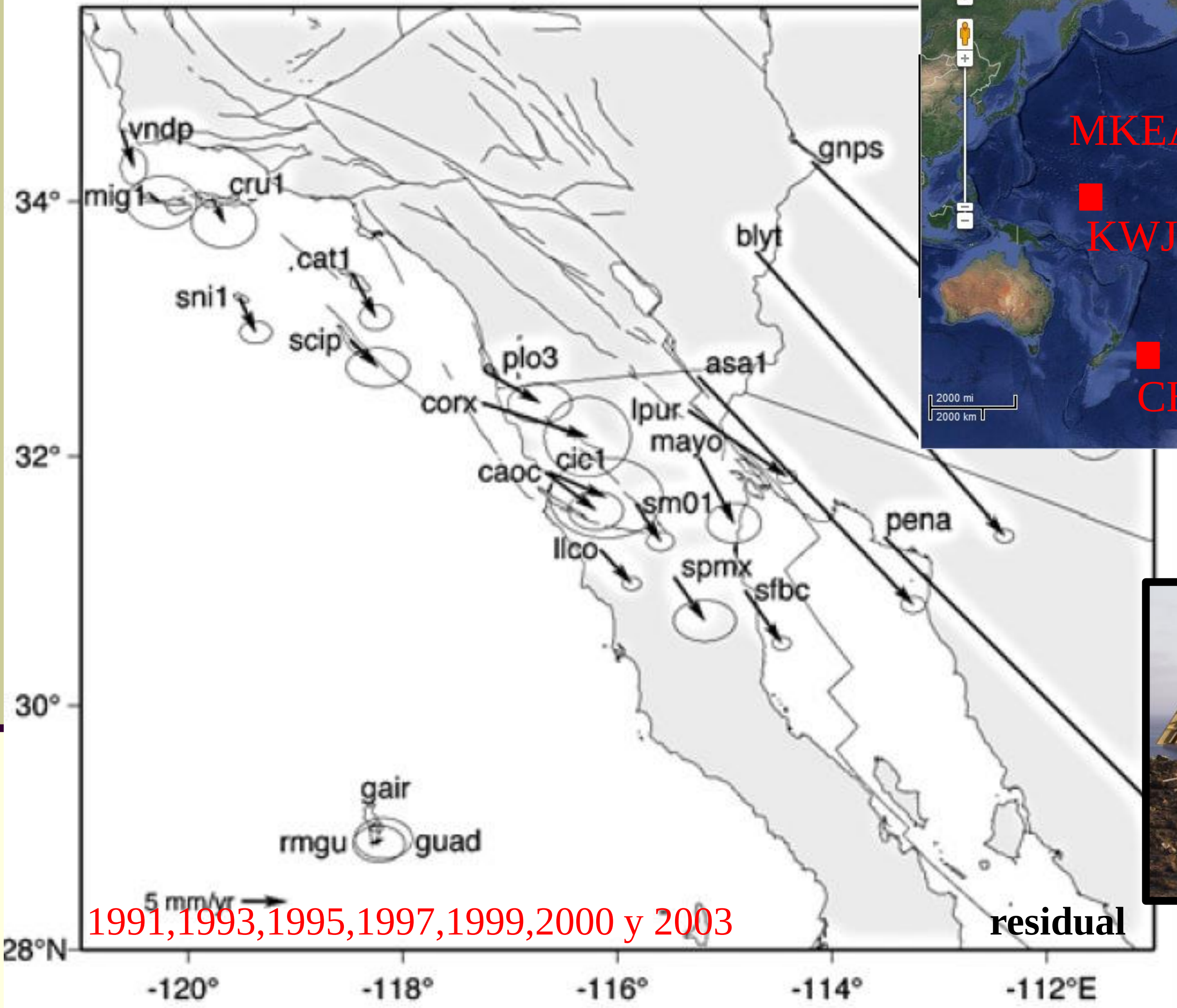
Placas tectónicas en la corteza terrestre con desplazamientos e interacciones.

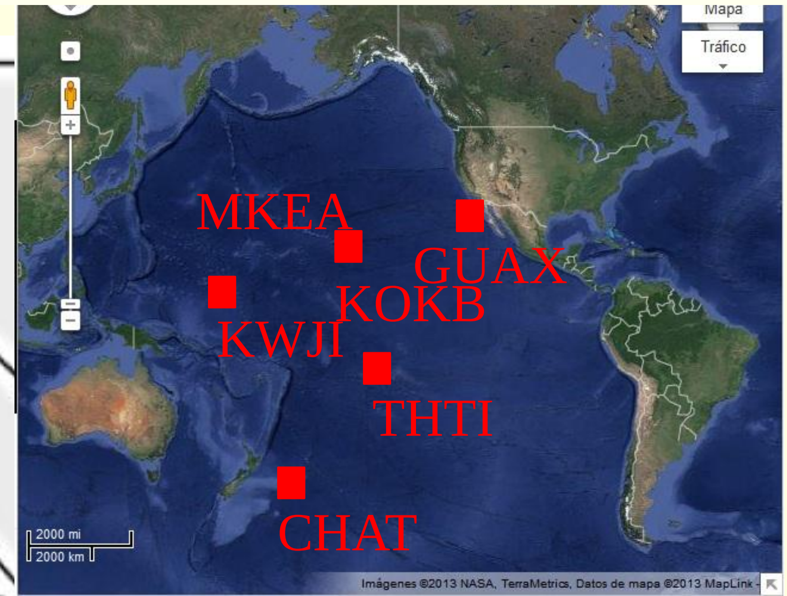
El movimiento de la placa se puede representar: $v = \omega \, x \, r$



Isla de Guadalupe, como restricción de placa Pacifico rígida.

Modelo de NAPA: lat 50.007; long -76.98;w=0.763 /Myr



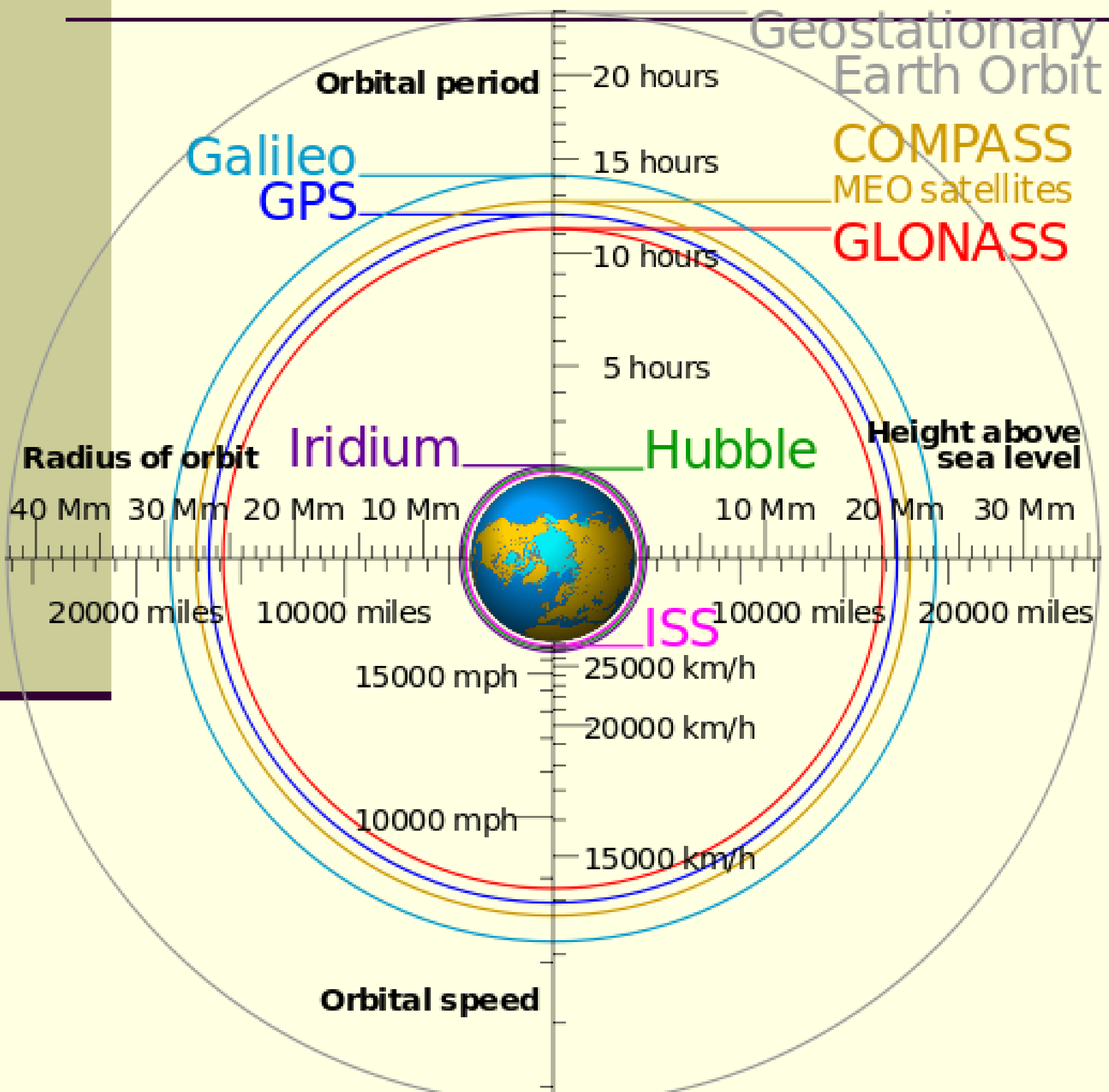


1991,1993,1995,1997,1999,2000 y 2003

residual

Guadalupe Island, Mexico as a new constraint for Pacific plate motion. Gonzalez-Garcia et. al (2003).

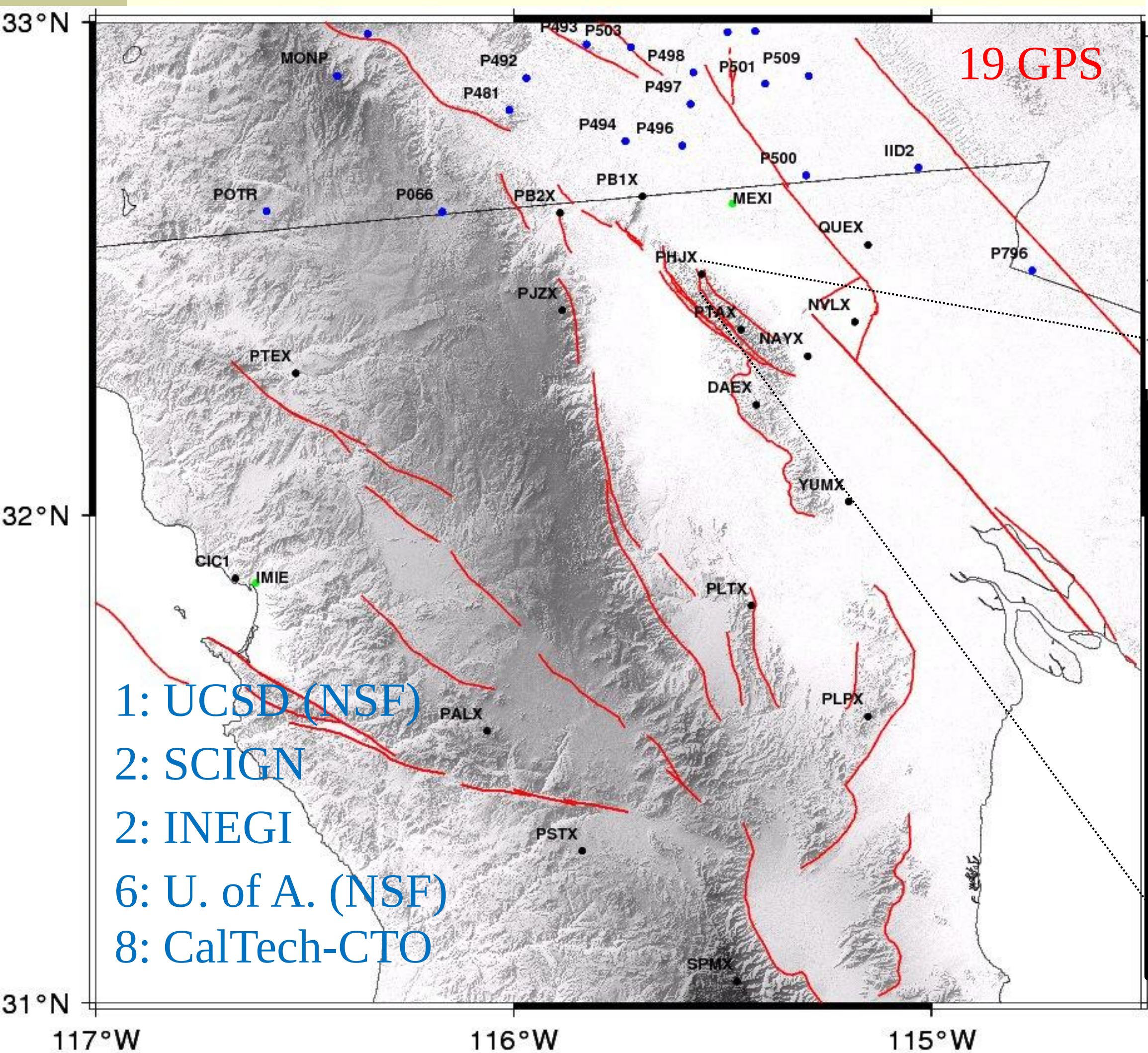
Sistemas de navegación GNSS: GPS, GLONASS, GALILEO, COMPASS



Comparación de los sistemas GPS, GLONASS, Galileo y Compass (órbita terrestre media) órbitas satelitales del sistema de navegación con la Estación Espacial Internacional, el Telescopio Espacial Hubble y las órbitas de la constelación Iridium, órbita terrestre geoestacionaria, y el tamaño nominal de la Tierra. La órbita de la Luna es 9,1 veces más grande (en la radio y la longitud) que la órbita geoestacionaria

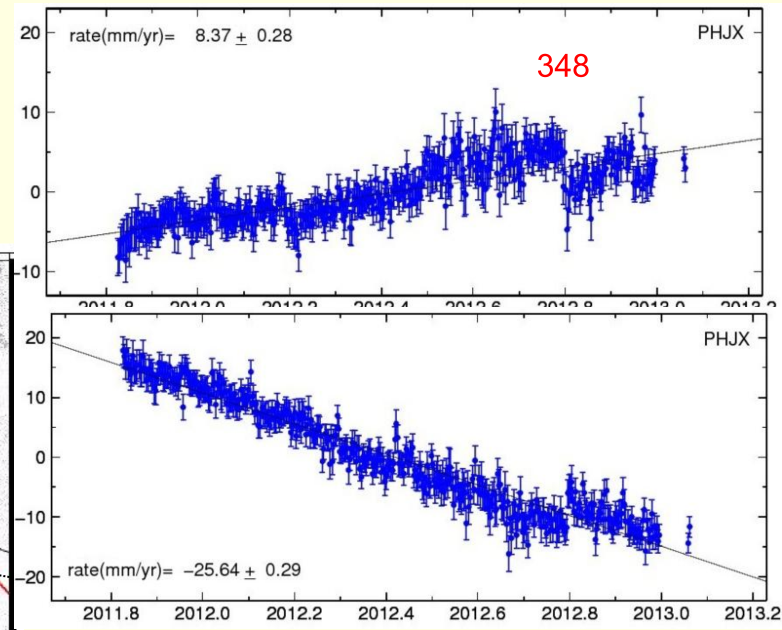
Estaciones GPS permanentes en norte de B.C.

Geodesia



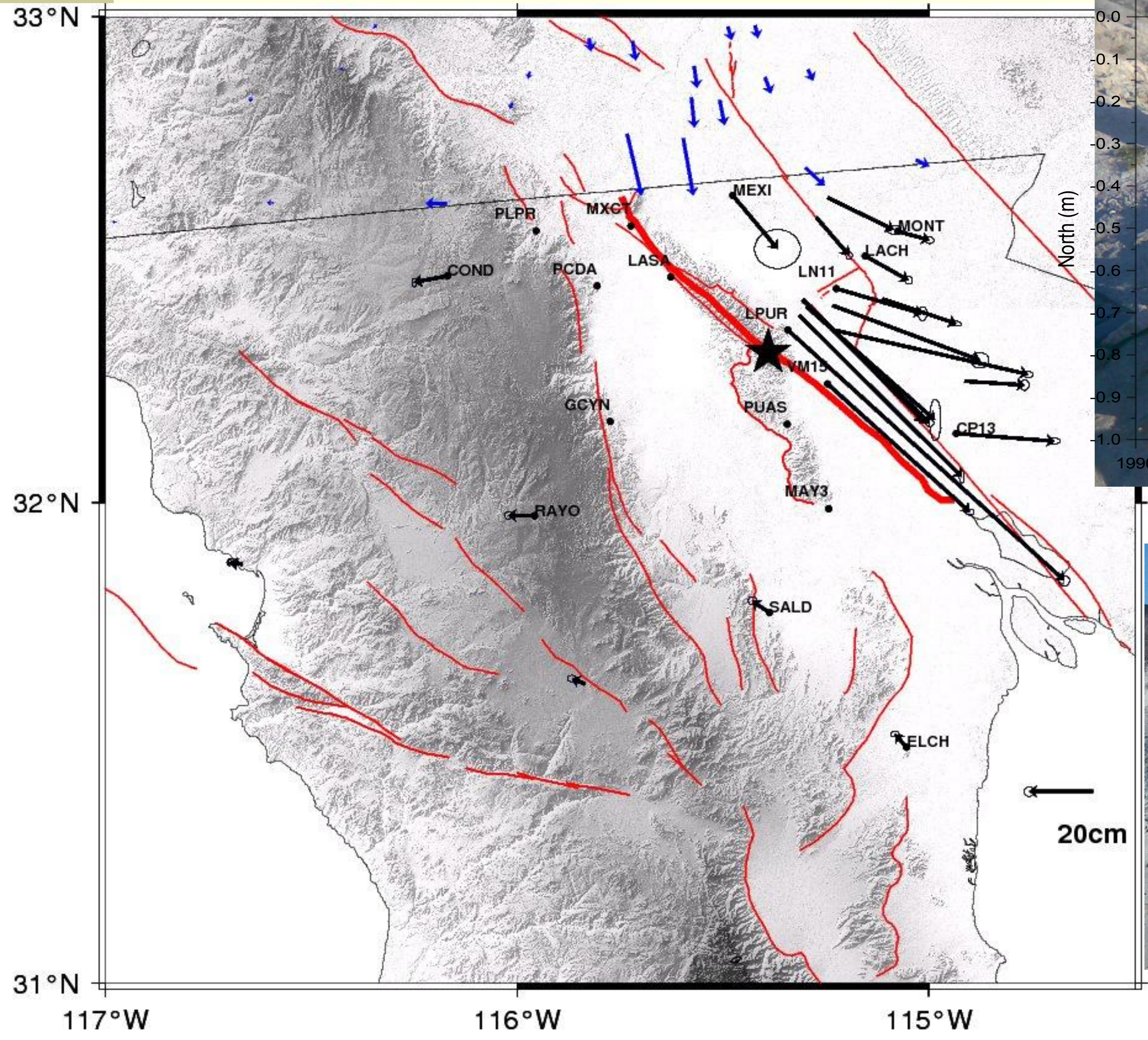
19 GPS

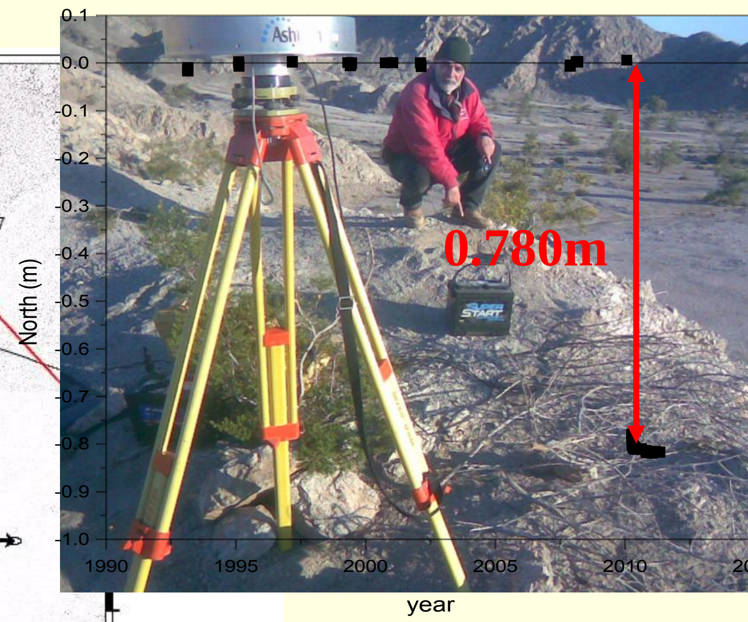
1: UCSD (NSF)
2: SCIGN
2: INEGI
6: U. of A. (NSF)
8: CalTech-CTO

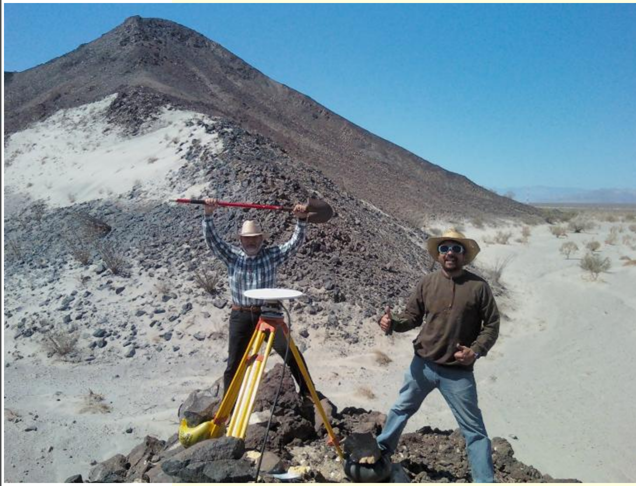


Señal/Ruido ⇨ Meteorología
Física Espacial
Sismología

Desplazamiento cosísmico El Mayor-Cucapah (Mw 7.2)[349]

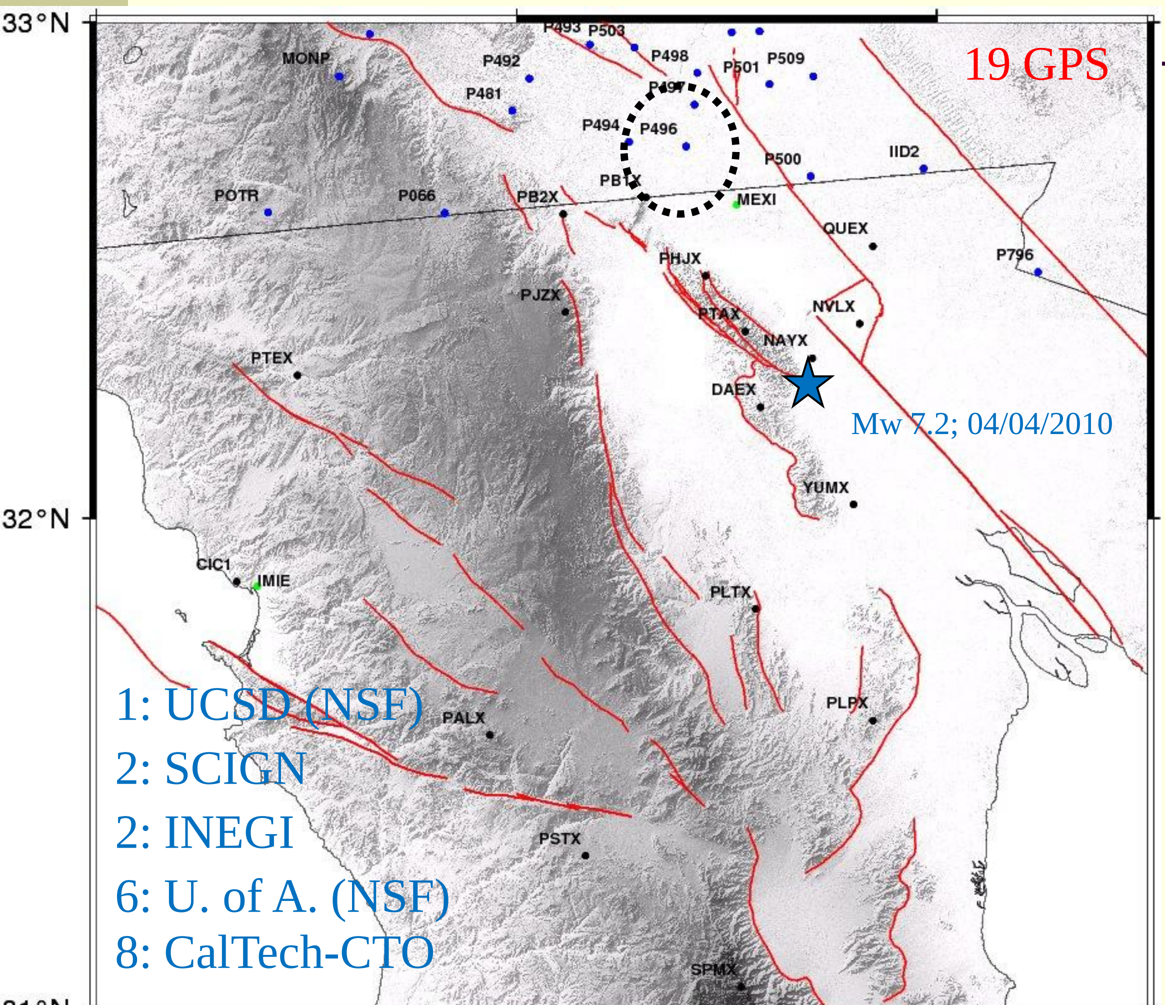


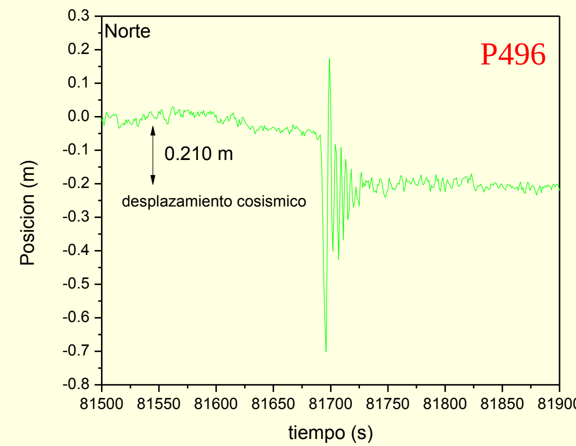


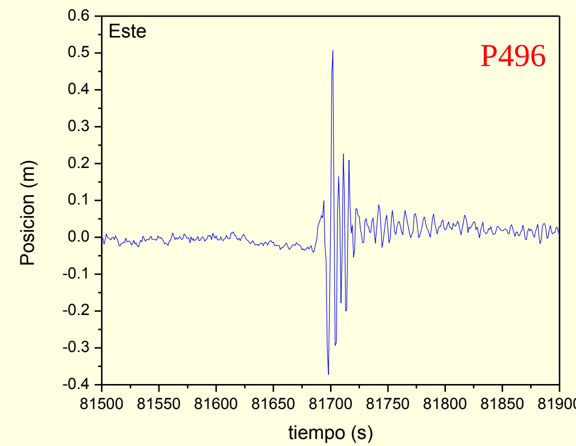


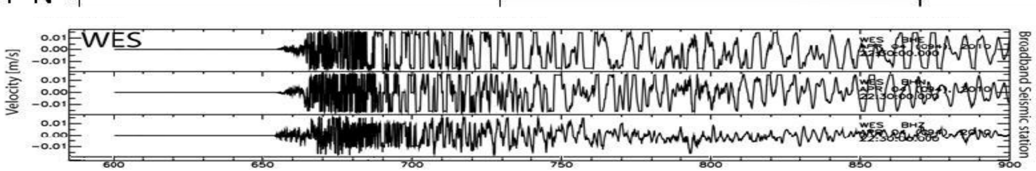
Estaciones GPS permanentes en norte de B.C.

Sismología/Geodesia WES/P496





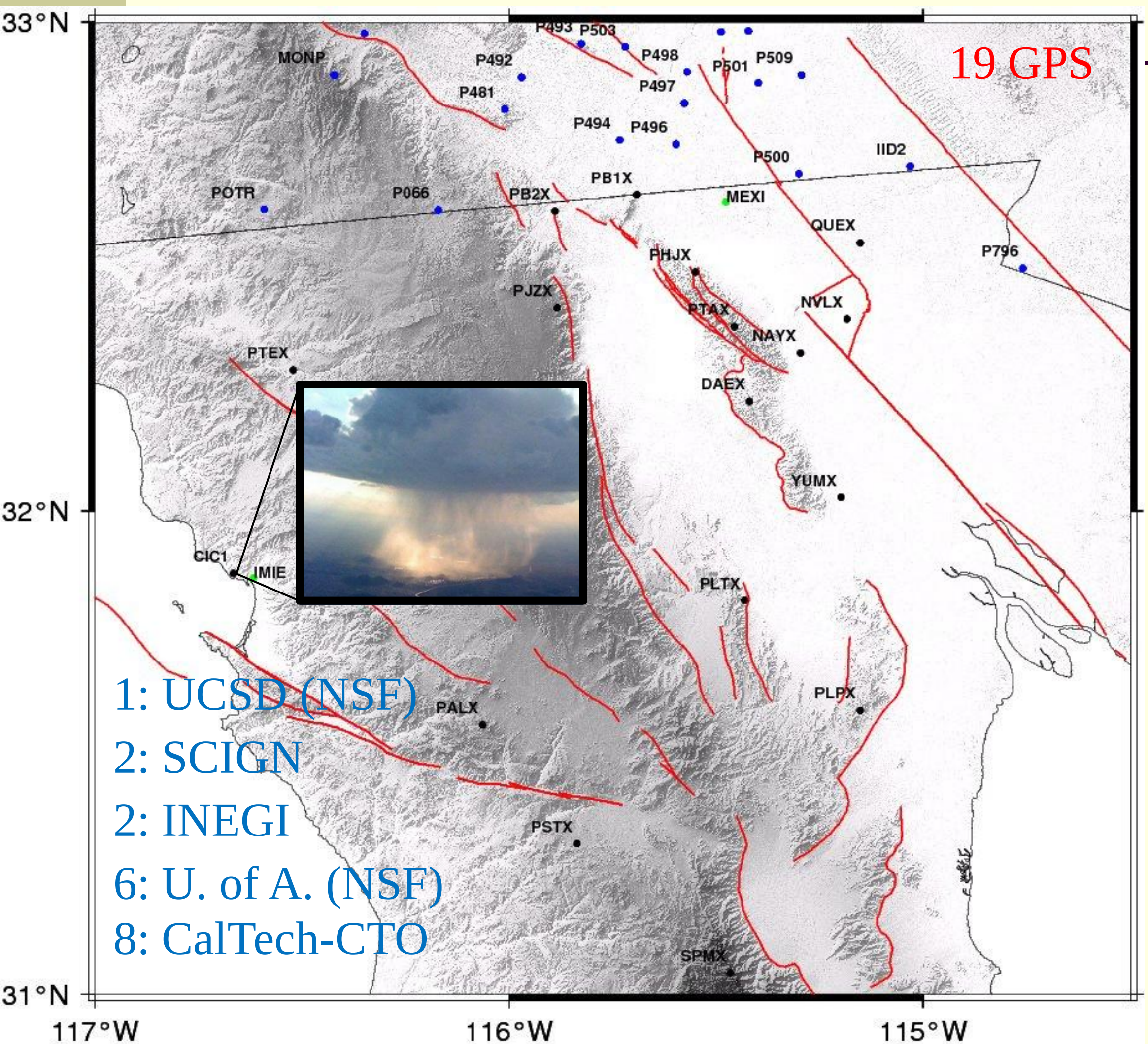




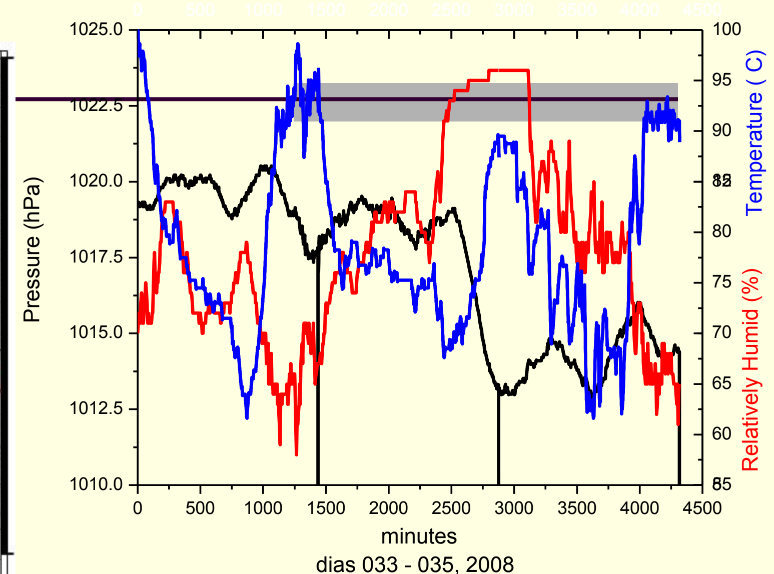
v.g. Real-time inversion of GPS data for finite fault modelling and rapid hazard assessment. Brendan et. al. (2012)
Rapid modelling of the 2011 Mw 9.0 Tohoku-oki earthquake with seismogeodesy. Melgar et. al. (2013)

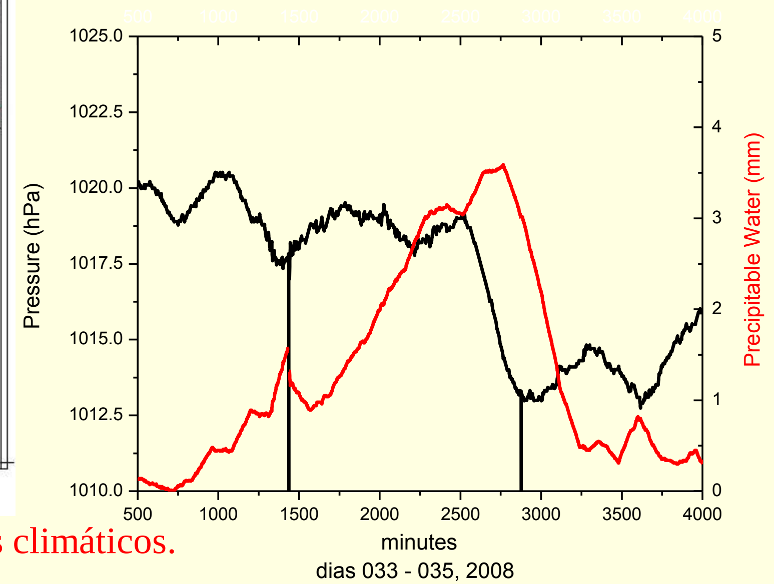
Estaciones GPS permanentes en norte de B.C.

Meteorología



1: UCSD (NSF)
2: SCIGN
2: INEGI
6: U. of A. (NSF)
8: CalTech-CTO



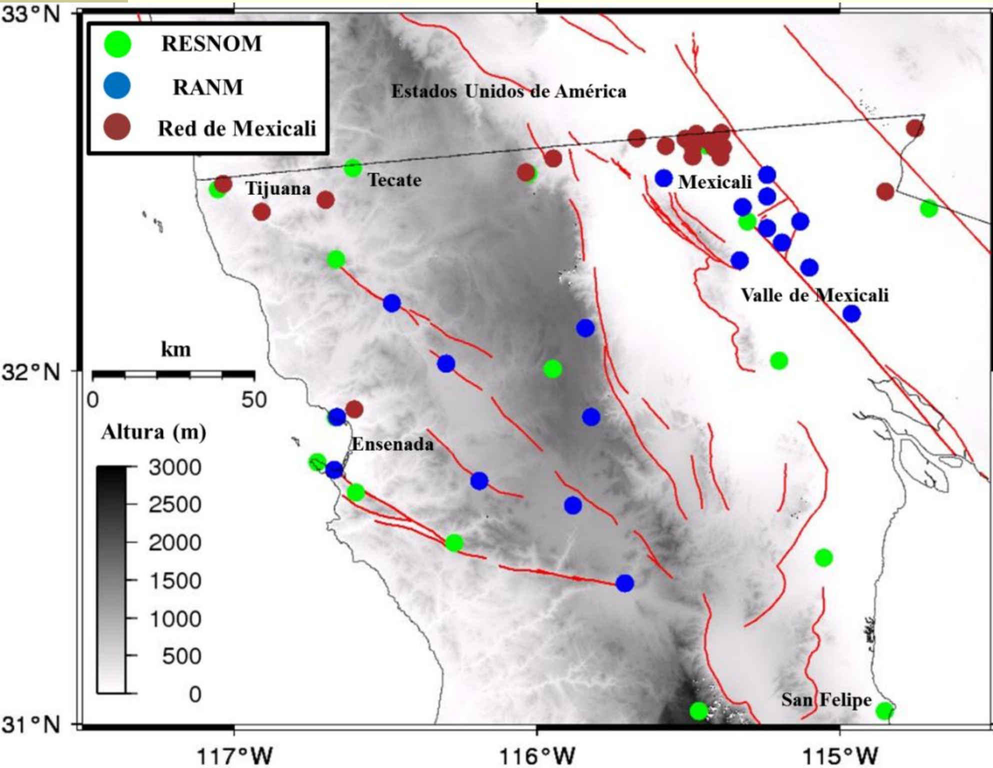


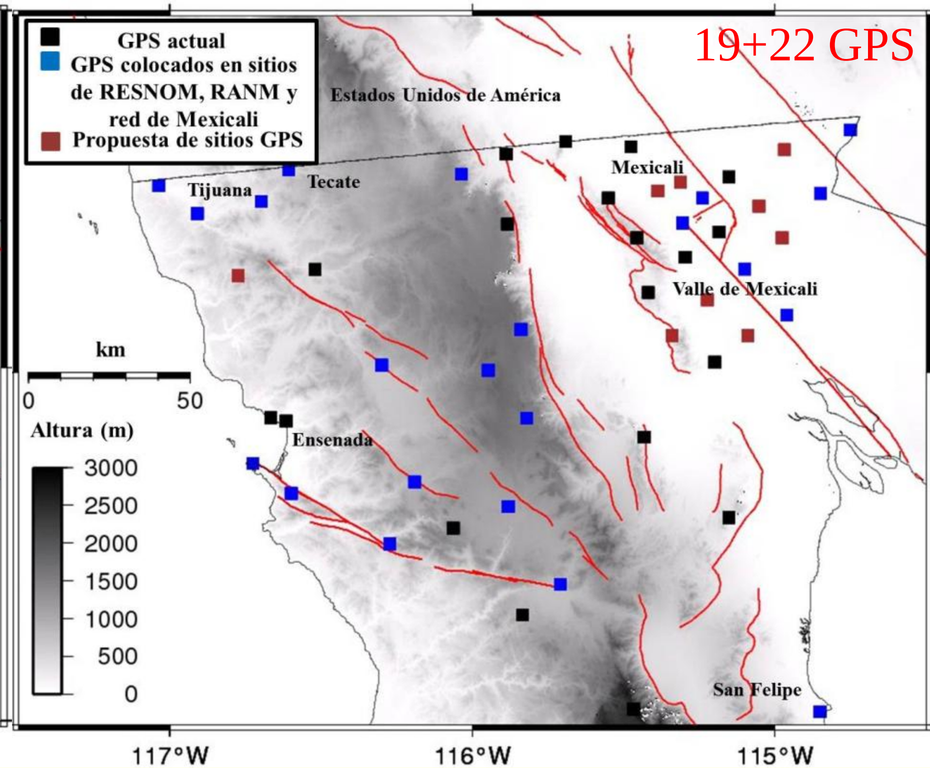
El vapor de agua es un gas de invernadero. Pronóstico de tiempo y modelos climáticos.

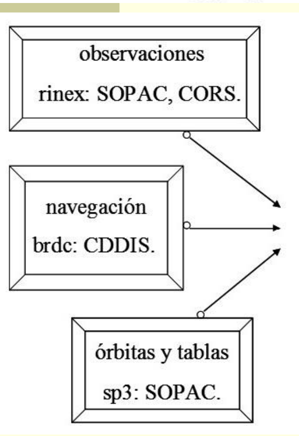
Water in the atmosphere. Bjorn S. et al., Physics Today 2013.

Red de estaciones sismológicas/geodésicas. Integración y ampliación.

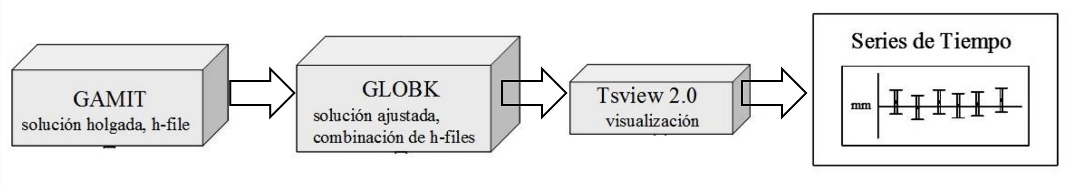
Estado actual de estaciones sismológicas co-localización de red sismológica/geodésica







50 GPS donados por USC a través de USGS.



12 GPS+GLONASS y strong motion+banda ancha donadas por NORTHCOM a través de USGS.

Workshop Big Data, Big Network: Big Opportunities for Collaboration Between Mexico and The United States

GRACIAS POR SU ATENCION

M.C. Javier Gonzalez-Garcia. javier@cicese.mx
Departamento de Sismología

M.C. Alejandro González-Ortega
Posgrado en Ciencias de la Tierra. Departamento de Sismología

Ensenada, Baja California, Mexico. 10 de Octubre de 2013

BIG DATA BIG NETWORK
Big Opportunities for Collaboration between Mexico and California
10-11 Oct 2013 • Ensenada, Mexico

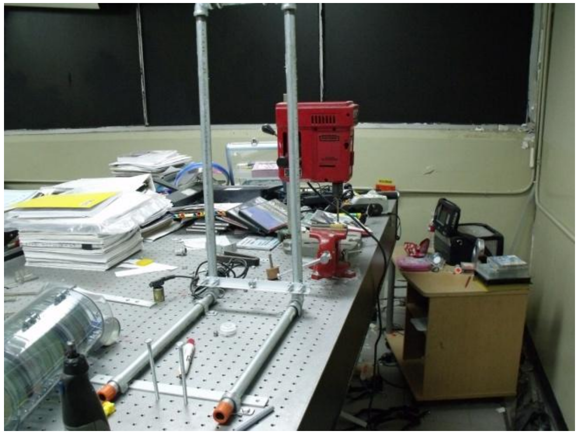
POSSIBLE ENHANCEMENT OF UABC'S CROSS BORDER COLLABORATIONS IN SEISMOLOGY, SEISMIC AND FAULT MAPPING USING GEOPHYSICAL DATA AND LIDAR

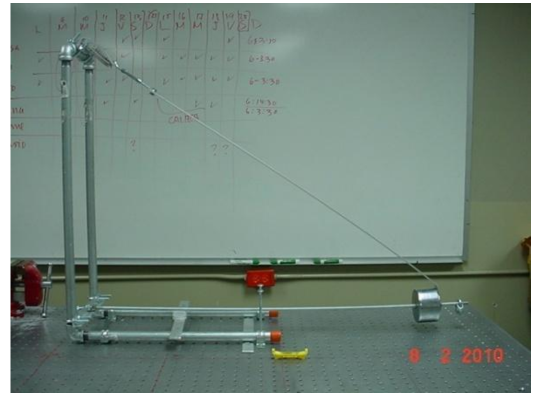
OCTAVIO LAZARO MANCILLA

Laboratorio de Sismología y Geofísica Aplicada, Instituto de Ingeniería, UABC Mexicali

Instituto de Ingenieria, UABC
After earthquake Mw 7.2
April 4, 2010

Seismological and Applied Geophysics Laboratory opened in January 2010 after earthquake





Our Long Period Lehman seismometer under construction was damaged during the earthquake

After the earthquake
3 broadband Caltech's seismometers were deployed
In Mexicali, one of them at Octavio Lázaro's home, the second at the Airport and the third at UABC Engineering Faculty in Ciudad Victoria

Collaboration UABC-CALTECH

We had problems for transmit the seismic records after earthquake
we could not use the Internet, we solved the problem using flash cards that we sent using parcel service to Seismo Lab.

222 Radon studies Collaboration UABC-Clippinger Laboratory, Ohio State University

^{222}RADON CONCENTRATION MEASUREMENTS BIASED TO CERRO PRIETO FAULT FOR VERIFY ITS CONTINUITY TO THE NORTHWEST OF THE MEXICALI VALLEY

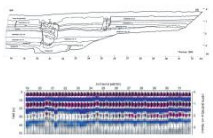
O. Lázaro-Mancilla[1], D. López[2], J. Reyes-López[1], C. Carreón-Diazconti[1], J. Ramírez-Hernández[1]
[1]Departamento de Hidrología, Geofísica e Impacto Ambiental
Instituto de Ingeniería, Universidad Autónoma de Baja California, Campus Mexicali
[2]Department of Geological Sciences, Ohio University
olazaro@iing.mxl.uabc.mx

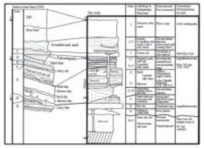
The need to know the exact location in the field of the fault traces in Mexicali has been an important affair due that the topography in this valley is almost flat and fault traces are hidden by plow zone, for this reason, the southern and northern ends of the San Jacinto and Cerro Prieto fault zones, respectively, are not well defined beneath the thick sequence of late Holocene Lake Cahuilla deposits. The purpose of this study was to verify if Cerro Prieto fault is the continuation to the southeast of the San Jacinto Fault proposed by Hogan in 2002 who based his analysis on pre-agriculture geomorphy, relocation and analysis of regional microseismicity, and trench exposures from a paleoseismic site in Laguna Xochimilco, Mexicali. In this study, four radon (^{222}Rn) profiles were carried out in the Mexicali Valley, first, to the SW-NE of Cerro Prieto Volcano, second, to the W-E along the highway Libramiento San Luis Río Colorado-Tecate, third, to the W-E of Laguna Xochimilco and fourth, to the W-E of the Colonia Progreso.

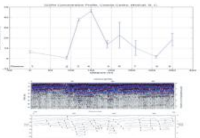
The Radon results allow us to identify in the Cerro Prieto profile three peaks where the values exceed 60 picocuries per liter (pCi/L), these regions can be associated to fault traces, one of them associated to the Cerro Prieto Fault (60 pCi/L) and other related with Michoacán de Ocampo Fault (239 pCi/L). The profile Libramiento San Luis Río Colorado-Tecate, shows four peaks above 200 pCi/L, the highest peak is related to the Michoacán de Ocampo fault. The profile of the Laguna Xochimilco, site used by Hogan et al., (2002), permit us observe three peaks above the 140 pCi/L, we can associate the peak of 184 pCi/L the Michoacán de Ocampo Fault, but none of them to the Cerro Prieto Fault. Finally in spite of the Colonia Progreso is the shortest profile with only five stations, it shows one peak with a value of 270 pCi/L that we can correlate with the Cerro Prieto Fault. The results of this study allow us to think in the possibility that the Michoacán de Ocampo Fault is the Continuation to the South of the San Jacinto Fault, not the Cerro Prieto Fault.

Introduction

The need to know the exact location in the field of the fault traces in the Mexicali Valley has been an important affair for geologist and earth scientists due that the topography there is almost flat and fault traces are hidden by plow zone that cover the thick sequence of late Holocene Lake Cahuilla deposits. For to solve this problem we have followed a two steps calibration procedure, first: we have used 100 MHz Ground Penetrating Radar images comparing them with trench logs (Thomas, 1965), both of the Colonia Castro site where the traces of Imperial fault are known, and second: we have used an outdoor ^{222}Rn Concentration profile comparing it with the GPR image of the same site. The results of this study permitted us relate the radon features profiles with those of GPR images where the fault traces are located (Lázaro-Mancilla, et al, 2008). After the calibration procedure we have used the lateral GPR images and ^{222}Rn Concentration profile for to infer another traces.

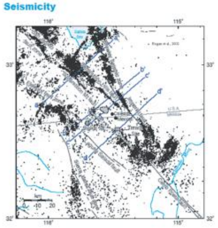


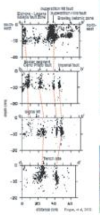


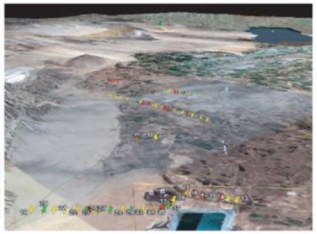


The purpose of this study was to prove how good are our results using only ^{222}Rn Concentration method for to verify if Cerro Prieto Fault is the continuation to the southeast of the San Jacinto Fault proposed by Hogan et al. (2002); they based their proposal on the results of three points analysis, first: the analysis of the New River trench site where is exposed a broad, diffuse zone of vertical strike-slip faults. They interpret these faults as the northwest continuation of the Cerro Prieto fault zone in an extensional jog or step-over, second: Geomorphic analysis suggests that several linear, northwest-trending shallow lakes along the New River represent fault-controlled transtensional grabens along a right stepping en echelon fault zone. This geomorphic pattern suggests an active continuation of the Cerro Prieto fault zone to the northwest from Mexicali toward the Heber geothermal field and third: Relocated earthquakes that support this interpretation. A northwest trending seismicity lineament extends from Cerro Prieto volcano to the west side of the Heber geothermal field. Hypocenter alignment also define the southeast extensions of the Superstition Mountain and Superstition Hills faults of the San Jacinto fault zone for nearly 30 km beyond their mapped fault traces; the Superstition Mountain fault extension reaches the Heber geothermal field. The Heber field apparently occupies a right step in the Cerro Prieto-San Jacinto fault zone. The other two major geothermal fields in the region, near Cerro Prieto and Salton Buttes, are also associated with major fault step-over.

Seismicity







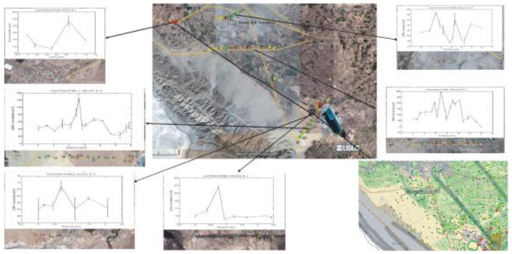
Presented at the American Geophysical Union 2009 Joint Assembly, May 24-27, Toronto, Canada.

Location of the ^{222}Rn Concentration profiles

In this study, four radon (^{222}Rn) profiles were carried out in the Mexicali Valley, first, to the SW-NE of Cerro Prieto Volcano, second, to the W-E along the highway Libramiento San Luis Río Colorado-Tecate, third, to the W-E of Laguna Xochimilco and fourth, to the W-E of the Colonia Progreso.

Method

For Radon concentration measurements we used a vacuum soil probe assembly model 154 connected to a Pylon AB-5 Portable Radiation Monitor. For interpretation of the ^{222}Rn concentration profiles we check the Reconnaissance Geologic Map of the State of Baja California (Gastil et al., 1971) and the Geologic Map D-MN5 of the Instituto Municipal de Investigación y Planeación Urbana de Mexicali (2008).



Results

The Radon results allow us to identify in the Cerro Prieto profile three peaks where the values exceed 60 picocuries per liter (pCi/L), these regions can be associated to fault traces, one of them associated to the Cerro Prieto Fault (60 pCi/L) and other related with Michoacán de Ocampo Fault (239 pCi/L). The profile Libramiento San Luis Río Colorado-Tecate, show four peaks above 200 pCi/L, with the highest peak related to the Michoacan fault. The profile of the Laguna Xochimilco, site used by Hogan (2002), permit us observe three regions above the 140 pCi/L, but we only can associate the peak of 184 pCi/L to the Michoacán de Ocampo Fault, but none peak to the Cerro Prieto Fault. Finally in spite of the Colonia Progreso is the shortest profile with only five stations, it shows one peak with a value of 270 pCi/L that we can correlate with the Cerro Prieto Fault.

Discussion

For interpretation purpose we have overlapped the Cerro Prieto fault from the Gastil Reconnaissance map in the picture of the study area and we observed that the Cerro Prieto fault changes his direction about 7° to the northwest of the Cerro Prieto volcano as we can observe in the dashed line in the picture, if we draw the fault to the northwest in that direction we can observe that the projection of the trace fault to the North West cross the Colonia Progreso, in this picture we have redrawn the site of the trench. On the other hand in the Geologic Map D-MN5 of the Instituto Municipal de Investigación y Planeación Urbana de Mexicali we can observe the extension of the Michoacán de Ocampo fault which trace is better related with description reported by Hogan et al., (2002). Finally to the left to this discussion we present a panoramic view that visually permit us align the San Jacinto fault with the Michoacán de Ocampo fault but not with Cerro Prieto fault.

Conclusion

In agreement with the results of this study we think that the Cerro Fault is no the continuation to the Southwest of the San Jacinto Fault but the Michoacán de Ocampo Fault. This conclusion pose more questions now: really the Michoacán de Ocampo continue to the north as Hogan propose? , beside, the Cerro Prieto Fault continue to the northwest?, if yes, that is no contrary to the dispersion center model? If not, its contribution is partial to the tectonics and seismic behavior?

Acknowledgements

This investigation was supported by funds of the Programa de Apoyo para Movilidad Académica 2008 and 2009 of the Universidad Autónoma de Baja California. Beside, the first author wishes to express his gratitude to the Geological Sciences Department of the Ohio State University by the loan of the Pylon AB-5 Portable Radiation Monitor and a vacuum soil probe assembly.

References

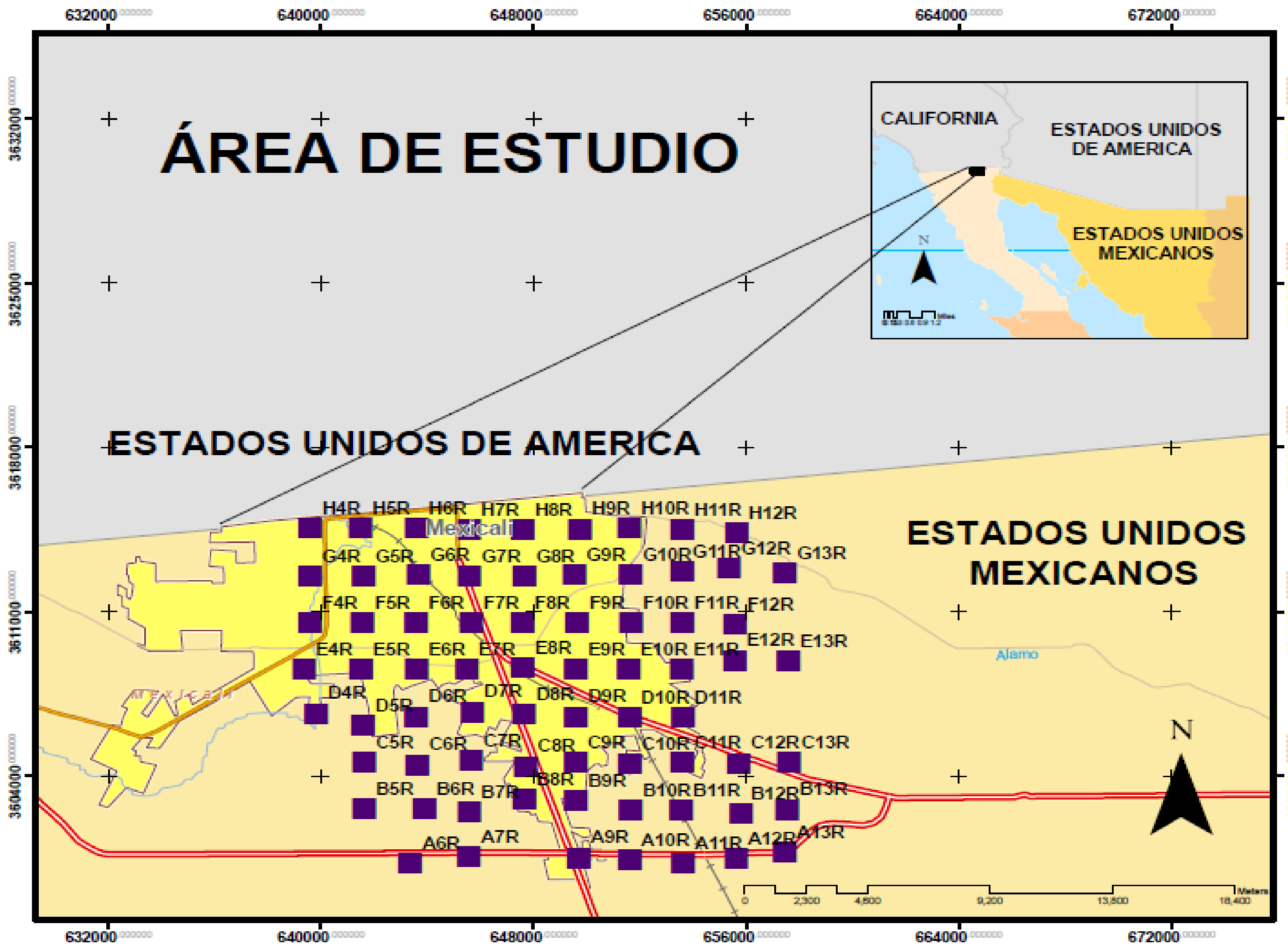
Hogan, P. J., Lindvall, S. C., Magistrale, H. and Rockwell, T. K. Continuity of the San Jacinto and Cerro Prieto fault zones: Missing slip found in Mexicali?. 2002 Southern California Earthquake Center Annual Meeting , September 8-11.

Lázaro-Mancilla , O., López, D., Reyes-López, J. , Carreón-Diazconti, C. and Ramírez-Hernández , J. Ground Penetrating Radar Images compared with a trench log and a radon profile for field location of the Imperial fault traces in Colonia Castro, Mexicali, Baja California, México. 2008 Southern California Earthquake Center Annual Meeting, September 2008. Palm Springs, CA, USA. Proceedings and Abstracts., Volume XVIII, pp.194-195.

Thomas, A. P. 1995. Seismotectonics of the Imperial-Mexicali Valley, Northern Baja California, México. Master Science Thesis. San Diego State University.

222 Radon Mexicali research Area

Grid of
222 Radon
Research
Area

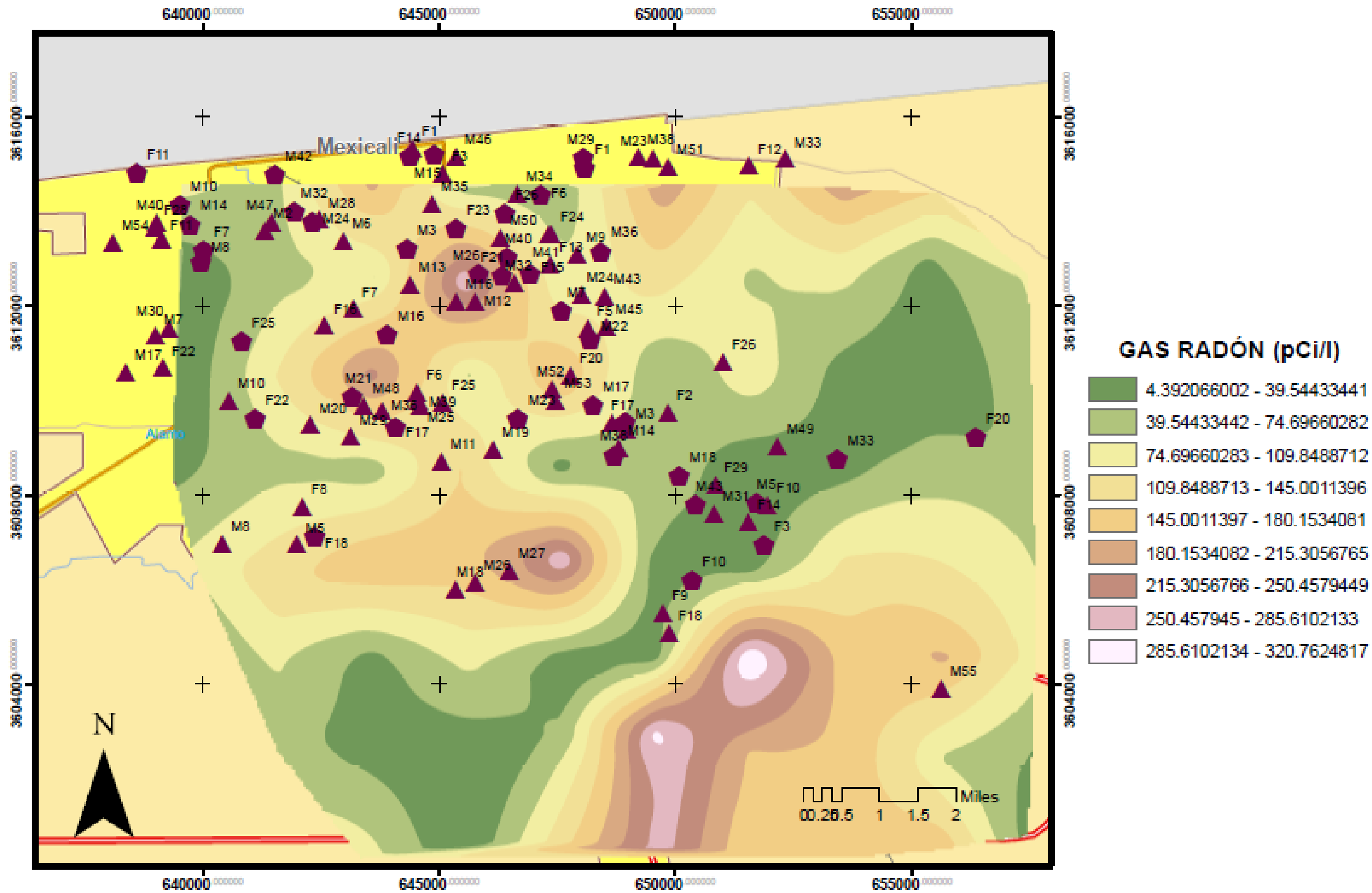


Results of
Radon
Research.

The zones
with high
Radon
concentration
are related
with clinical cases of lung
cancer.

In the future we
will need
monitoring
continuosly
222 Radon
emissions,
this task imply the use
of big amount of data
Because previously high
radon emission in some
cases are related to
ocurrence of earthquakes.

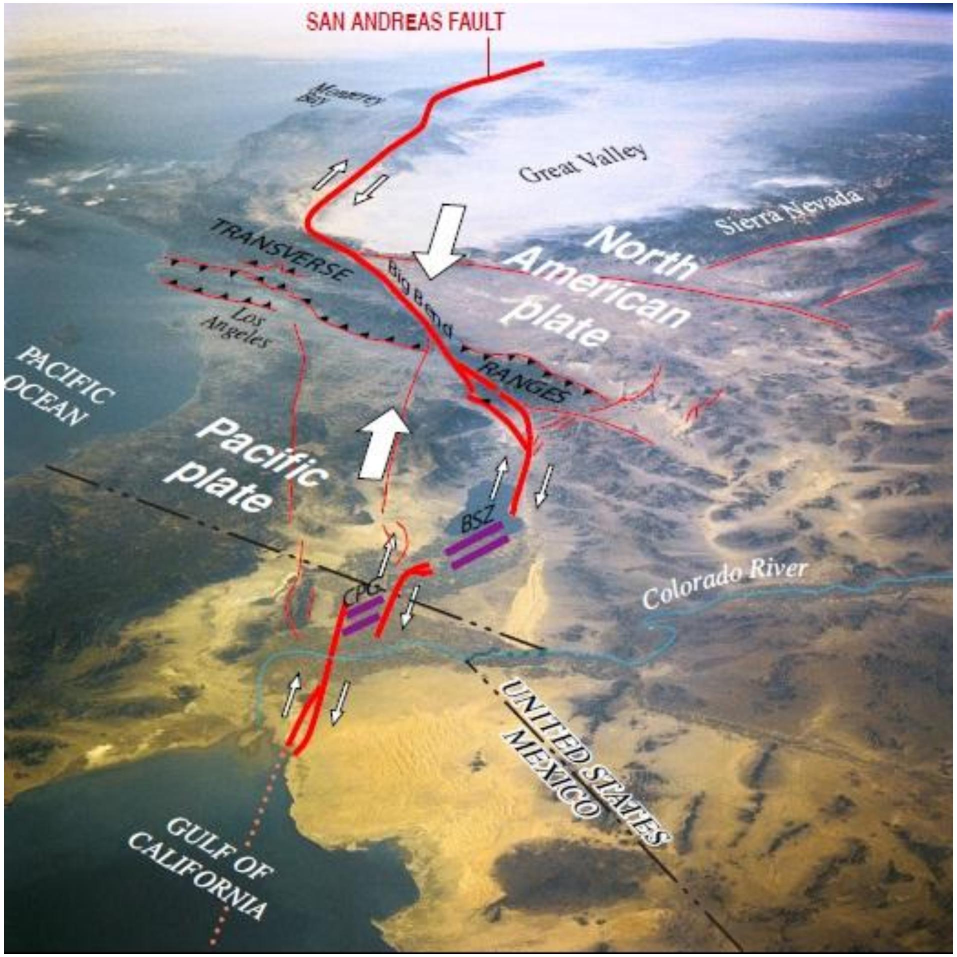
EMANACIÓN GAS RADÓN - MUERTES CÁNCER PULMONAR 2004 Y 2011

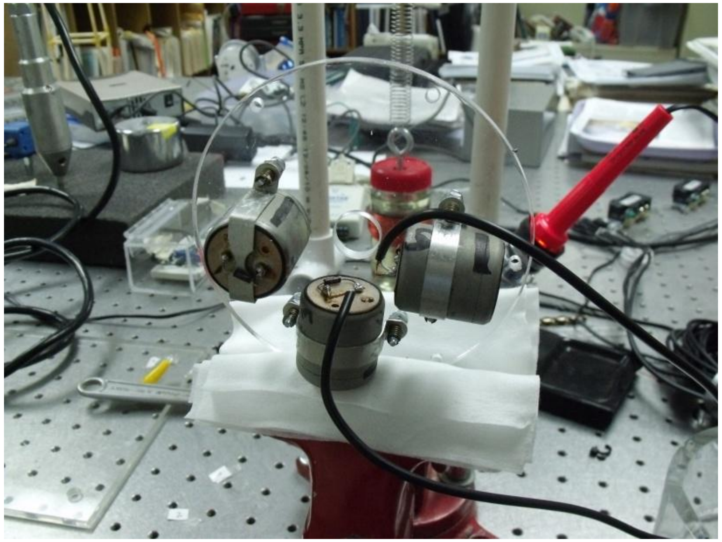


MUERTES CANCER PULMONAR 2004 MUERTES CÁNCER PULMONAR 2011

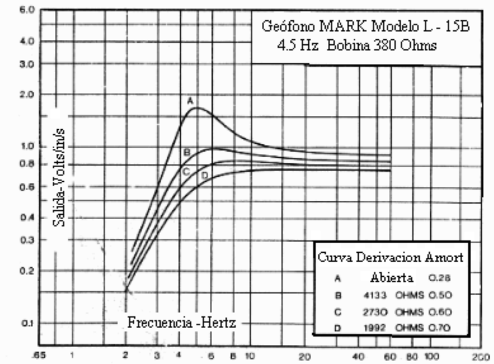
Salton Seismic-Imaging Project (SSIP): A Survey to Evaluate Earthquake Hazards and Structure of the Earth's Crust in Imperial and Coachella Valleys

CALTECH-VIRGINIA TECH-CICESE-UABC


SAN ANDREAS FAULT
Great Valley
Sierra Nevada
North American plate
TRANSVERSE
Big Bend
RANGES
Los Angeles
PACIFIC OCEAN
Pacific plate
BSZ
CPG
Colorado River
UNITED STATES
MEXICO
GULF OF CALIFORNIA



Intervalo de Frecuencia Standard, Hz	4.5 - 10
Tolerancia en Frecuencia	± 0.5 Hz
Resistencia de la Bobina Standard, Ohms	150/240/380
Tolerancia Resistencia, %	5 5 5
Distorsión Máxima @ 0.7 in/s @ 12 Hz ó Resonancia	0.2 %
Constante de Transducción, V/in/s ± 10%	$0.047\sqrt{Rc}$
Amortiguamiento Circuito Abierto, ± 10%	$\frac{1.233}{f}$
Amortiguamiento Corriente Bobina	$\frac{11.84\,Rc}{f\,(Rc+Rs)}$
Masa Suspendida, Gramos	23
Sensibilidad de potencia, mW,in/s	2.21
Movimiento Carcasa a Bobina, en p-p	0.080
Diámetro Unidad Básica, in	1.25
Altura Unidad Básica, in	1.4
Peso Unidad Básica, oz	5.0



Construction of 3 Components
Short Period
Seismometer using 4.5 Hz
Mark Geophones following the Philosophy
of Larry Crochane

USGS
science for a changing world

USGS Home
Contact USGS
Search USGS

Earthquake Hazards Program — Home | About Us | Contact Us | Search

EARTHQUAKES | HAZARDS | LEARN | PREPARE | MONITORING | RESEARCH

- Past
  - Past 8–30 days
  - Significant Earthquakes
  - Earthquake Lists & Maps
  - Search for an Earthquake
- Present
  - Real-time - CA/NV
  - Real-time - USA
  - Real-time - Worldwide
  - About Earthquake Maps
  - KML / RSS Feeds & Data
  - Earthquake Notifications
  - Seismogram Displays
  - Earthquake Animations
  - Did You Feel It?
  - ShakeMaps
  - PAGER
  - EQ Summary Posters
- Future
  - Earthquake Scenarios
  - Prediction
  - Probabilities
- Location
  - Info by State
  - Info by Country/Region

Magnitude 4.4 - BAJA CALIFORNIA, MEXICO

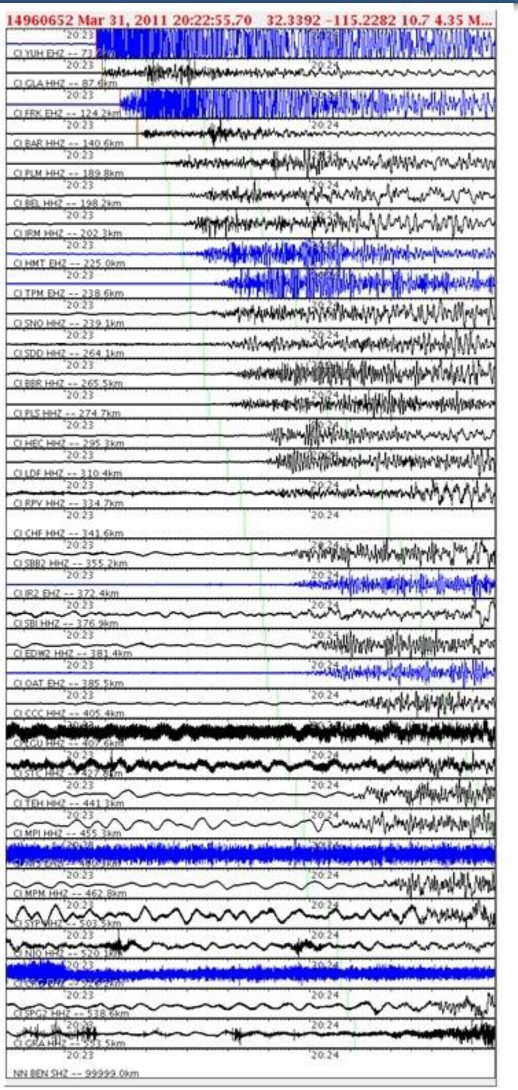
2011 March 31 20:22:55 UTC

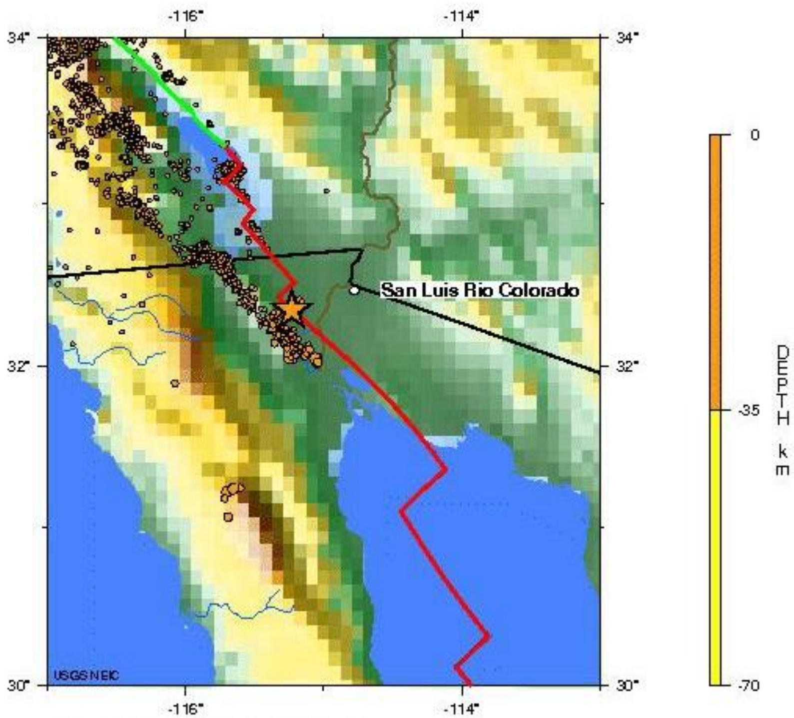
Details | Maps | Scientific & Technical | Additional Info

Earthquake Details

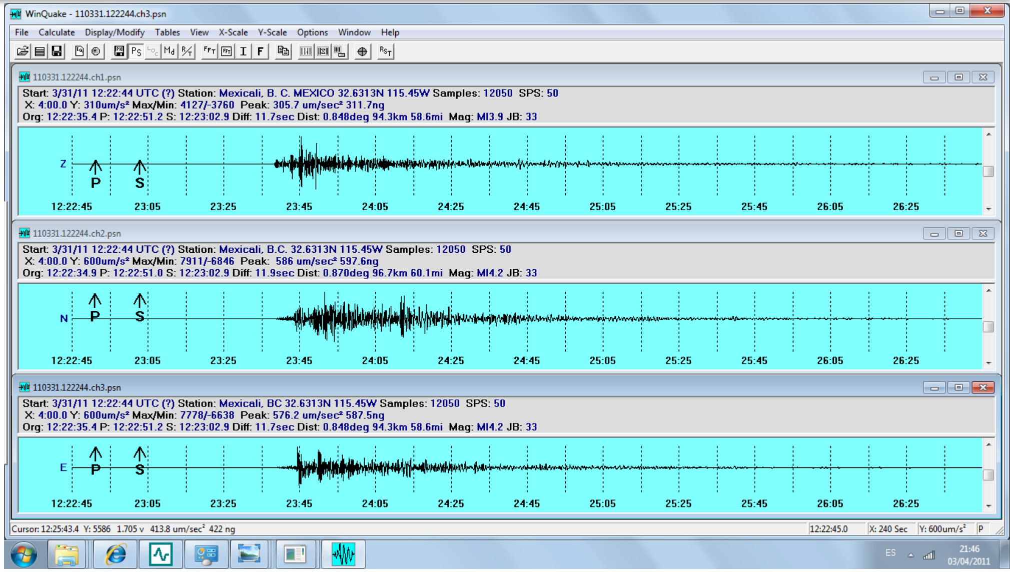
This event has been reviewed by a seismologist.

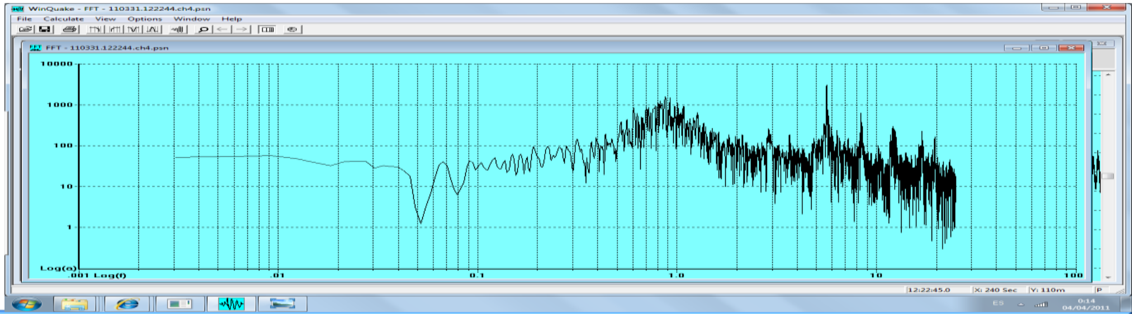
Magnitude	4.4
Date-Time	**Thursday, March 31, 2011 at 20:22:55 UTC** Thursday, March 31, 2011 at 12:22:55 PM at epicenter Time of Earthquake in other Time Zones
Location	32.339°N, 115.228°W
Depth	10.7 km (6.6 miles)
Region	BAJA CALIFORNIA, MEXICO
Distances	12 km (8 miles) WNW (296°) from **Guadalupe Victoria, Baja California, Mexico** 41 km (26 miles) SSE (147°) from **Mexicali, Baja California, Mexico** 45 km (28 miles) SE (146°) from **Calexico, CA** 169 km (105 miles) E (97°) from **Tijuana, Baja California, Mexico**
Location Uncertainty	horizontal +/- 2 km (1.2 miles); depth +/- 3.6 km (2.2 miles)
Parameters	Nph= 15, Dmin=11 km, Rmss=0.29 sec, Gp=230°, M-type=local magnitude (ML), Version=5
Source	California Integrated Seismic Net: USGS Caltech CGS UCB UCSD UNR
Event ID	ci14960652





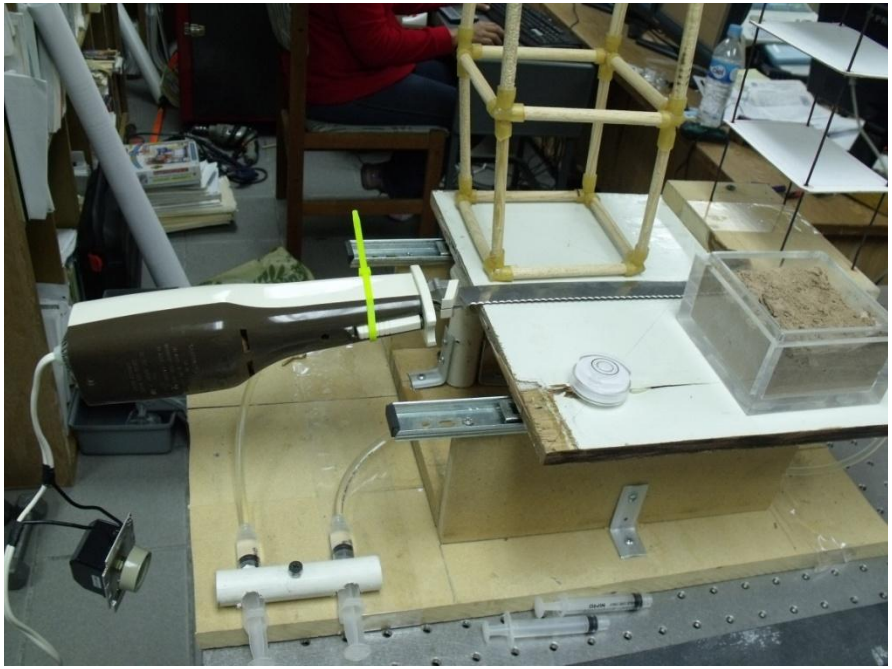
BAJA CALIFORNIA, MEXICO
2011 03 31 20:22:55 UTC 32.34N 115.23W Depth: 10.7 km, Magnitude: 4.4
Seismicity in 2011

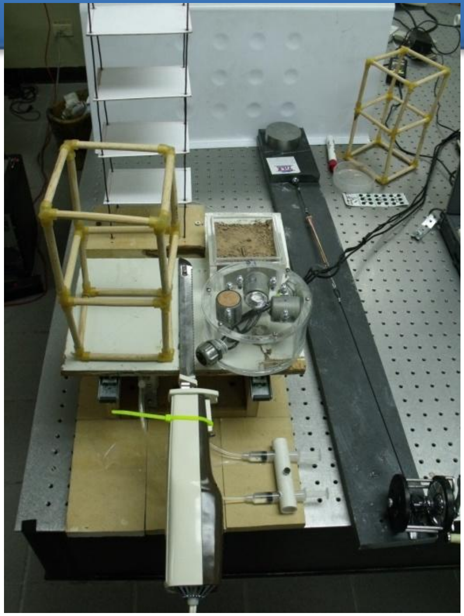


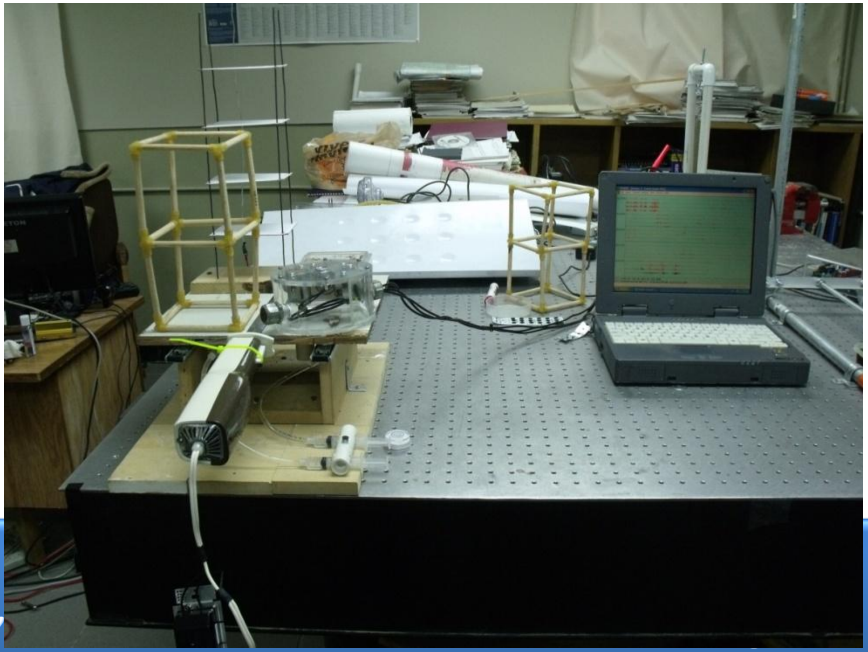


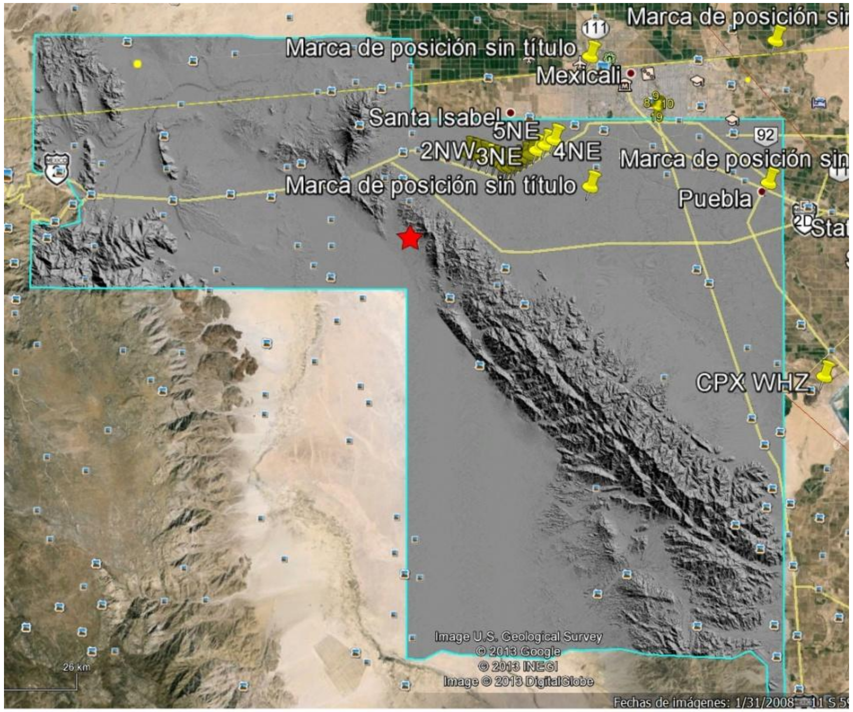
UNAM' S
Shaking Table

UABC's Shaking Table









Shortly we will be developing a project with CICESE and CALTECH on Geological Faults in the urban area of Mexicali and perhaps we will need LIDAR data, that purpose imply handle large amount of data, over all if we will required information from different dates.

For future collaborations

We can help

We need your help

To discuss what really matters

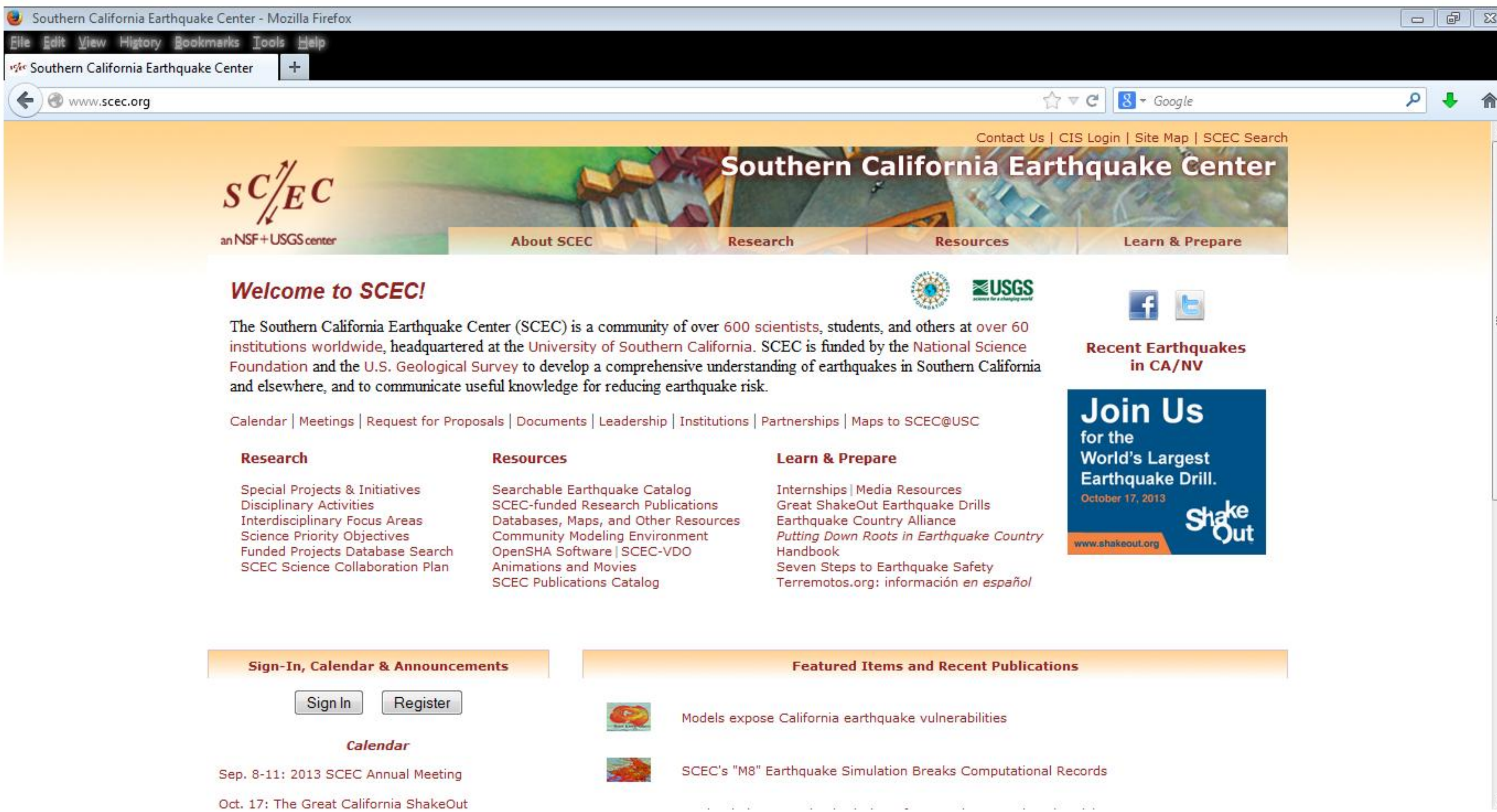
To contribute to the development of our Lab and for prepare us to use BDBN

Thank you very much

Enhancement of Caltech's cross-border collaborations in Geology and Geophysics

Joann Stock, Robert Clayton
California Institute of Technology

SCEC – gateway, includes individual wiki pages for rapid earthquake response



Did You Feel It? - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Did You Feel It?

earthquake.usgs.gov/earthquakes/dyfi/

USGS science for a changing world

USGS Home
Contact USGS
Search USGS

Earthquake Hazards Program Home About Us Contact Us Search

EARTHQUAKES HAZARDS LEARN PREPARE MONITORING RESEARCH

Due to a lapse in Federal funding, the USGS Earthquake Hazards Program has suspended most of its operations. While the USGS will continue to monitor and report on earthquake activity, the accuracy or timeliness of some earthquake information products, as well as the availability or functionality of some web pages, could be affected by our reduced level of operation.

Did You Feel It?
- Home
- Archives
- Background
- Comment
- FAQ
- Disclaimer

Did You Feel It?

Report Unknown Event

View Archives

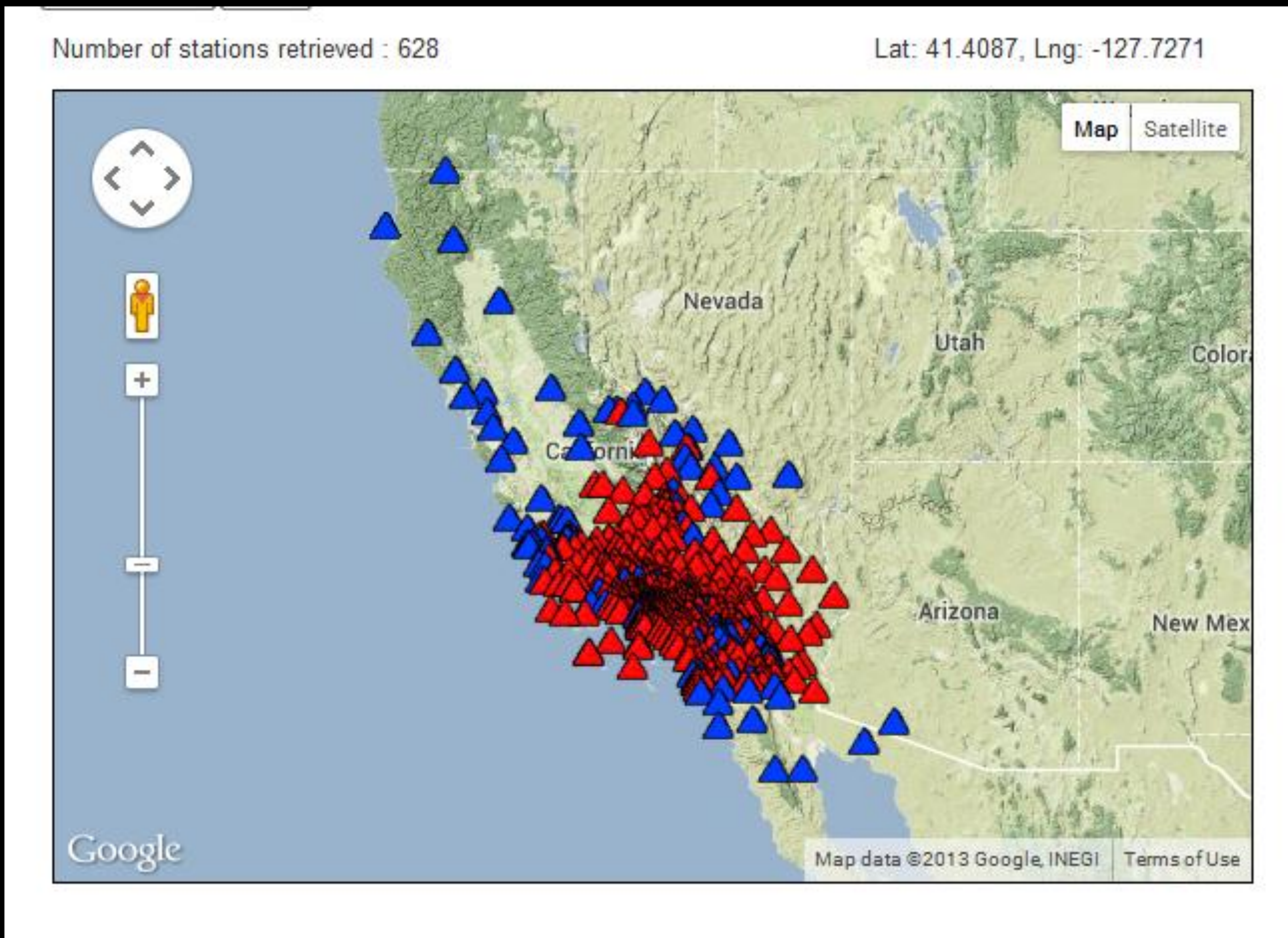
Found 4 matching results Events - Last 24 Hours

MMI	Mag	Location	Event Time	Event ID	Response

Our current collaborators (Mexican institutions)

- CICESE Ciencias de la Tierra
- UNAM Instituto de Geofísica, México D.F.
- UABC Ensenada
- UABC Mexicali
- Universidad de Sonora, Hermosillo, Sonora

California Integrated Seismic Network – 628 stations



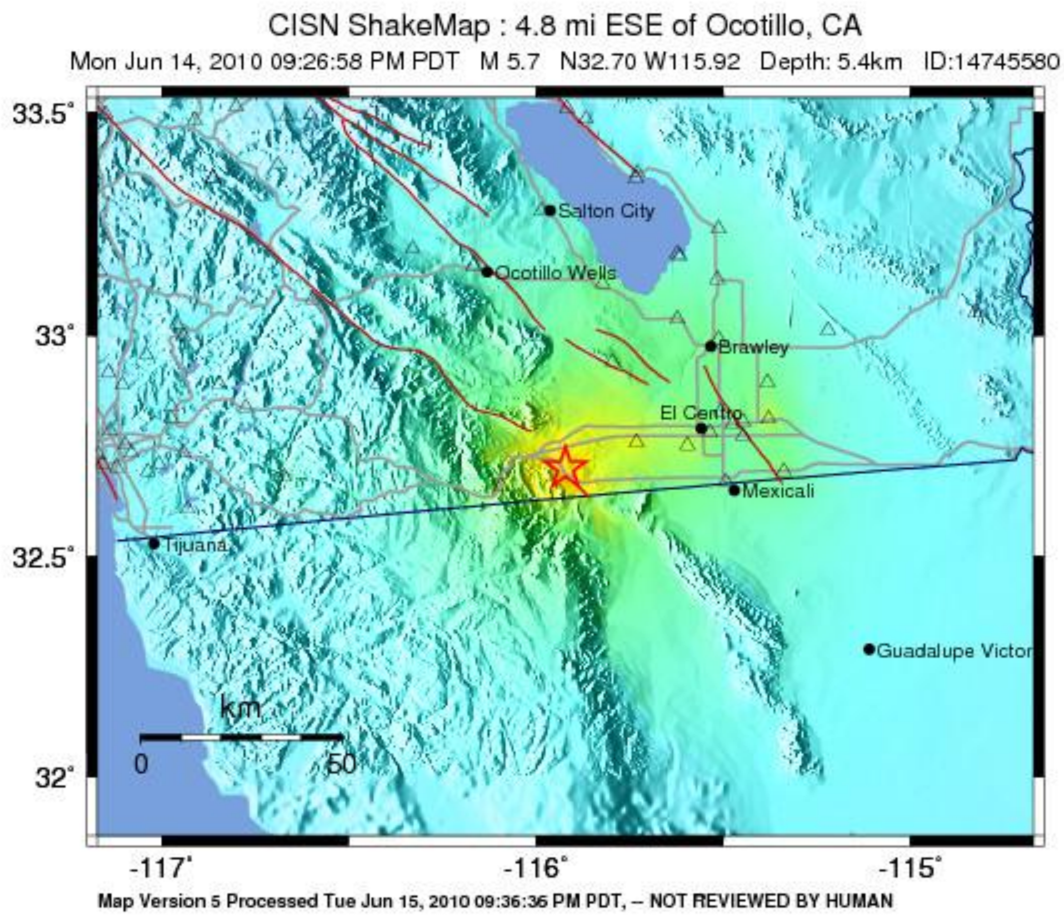
Red: our network
Blue: external but contributed to our network

CGS USGS CalOES Caltech UC Berkeley

CISN main data products

- Rapid locations of earthquakes
- ShakeMaps (observed & predicted ground shaking due to an earthquake)
- Software for rapid distribution of earthquake information to emergency management centers (ShakeCast)
- Strong motion records of engineering interest
- Seismological data for all recorded earthquakes for Southern California, archived in the Southern California Earthquake Data Center (SCEDC)

- Currently beta testing Earthquake Early Warning system

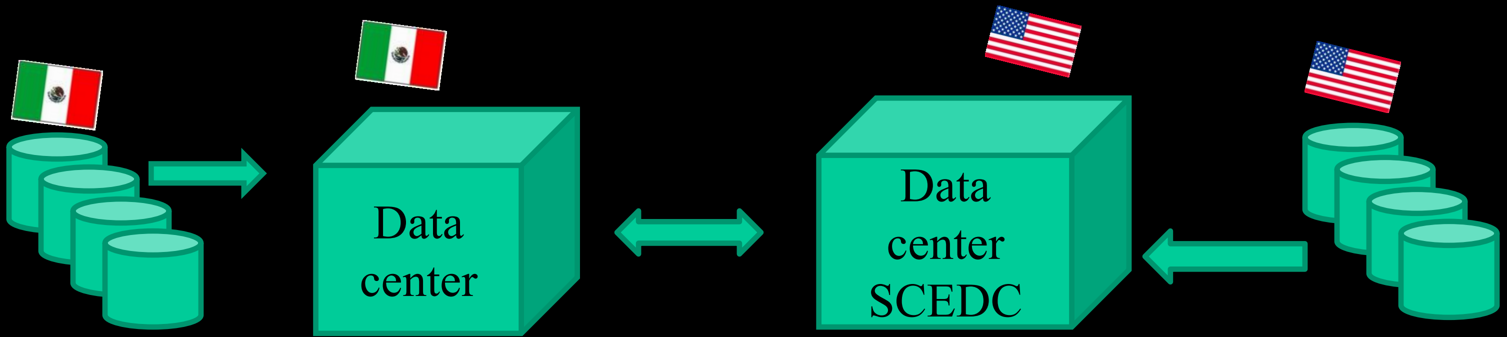


PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	<.17	.17-1.4	1.4-3.9	3.9-9.2	9.2-18	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	8.1-16	16-31	31-60	60-116	>116
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

Example of ShakeMap product

- Transmission of seismic data in both directions across the border:

from seismometers in Mexico to data center in Mexico City and from there to the Southern California Seismic Network Data Center; and vice versa



Research and societal goals: better locations of earthquakes near the US-Mexico border; duplicate archiving of data for regional earthquakes; understanding geometry of sources of earthquakes in Mexico and USA

Seismology collaborations, 1

Data volume: **2.6 Gigabytes** of data transferred in both directions each day

Takes anywhere from **1 to 3+ hours** to transfer these data from Mexico City to Pasadena

Data volume expected to increase to **5 Gbyte/day** as network is upgraded

Current network speeds

Seismology collaborations, 2

- Deployment of temporary web-enabled seismometers in the Mexicali Valley, after the April 4, 2010 El Mayor-Cucapah earthquake

Seismometer
& data logger

Goal: better locations of earthquakes in the aftershock sequence

Data volume: 4 Gbyte CF card stores data (2 – 4 weeks for typical recording parameters)

Instrument can be directly connected to the web via an internet drop (no need for local computer)

Challenge encountered: inability to connect to internet

Seismology collaborations, 3 (CICESE and UABC-Mexicali)

Remember the question during Minster's talk re: why did the earthquake simulation stop at the border?

Active-source seismic imaging experiment in the Salton Trough (Mexicali-Imperial Valleys) in 2011

- 126 explosion sources, about 4000 seismometer locations, recording rate 250-1000 Hz x 7 hours /day/deployed sensor
- **1 Terabyte** total shot data from main experiment

Goals: improve knowledge of fault geometry, basin seismic velocity model for better understanding of earthquake hazards;

Understand geological rifting processes

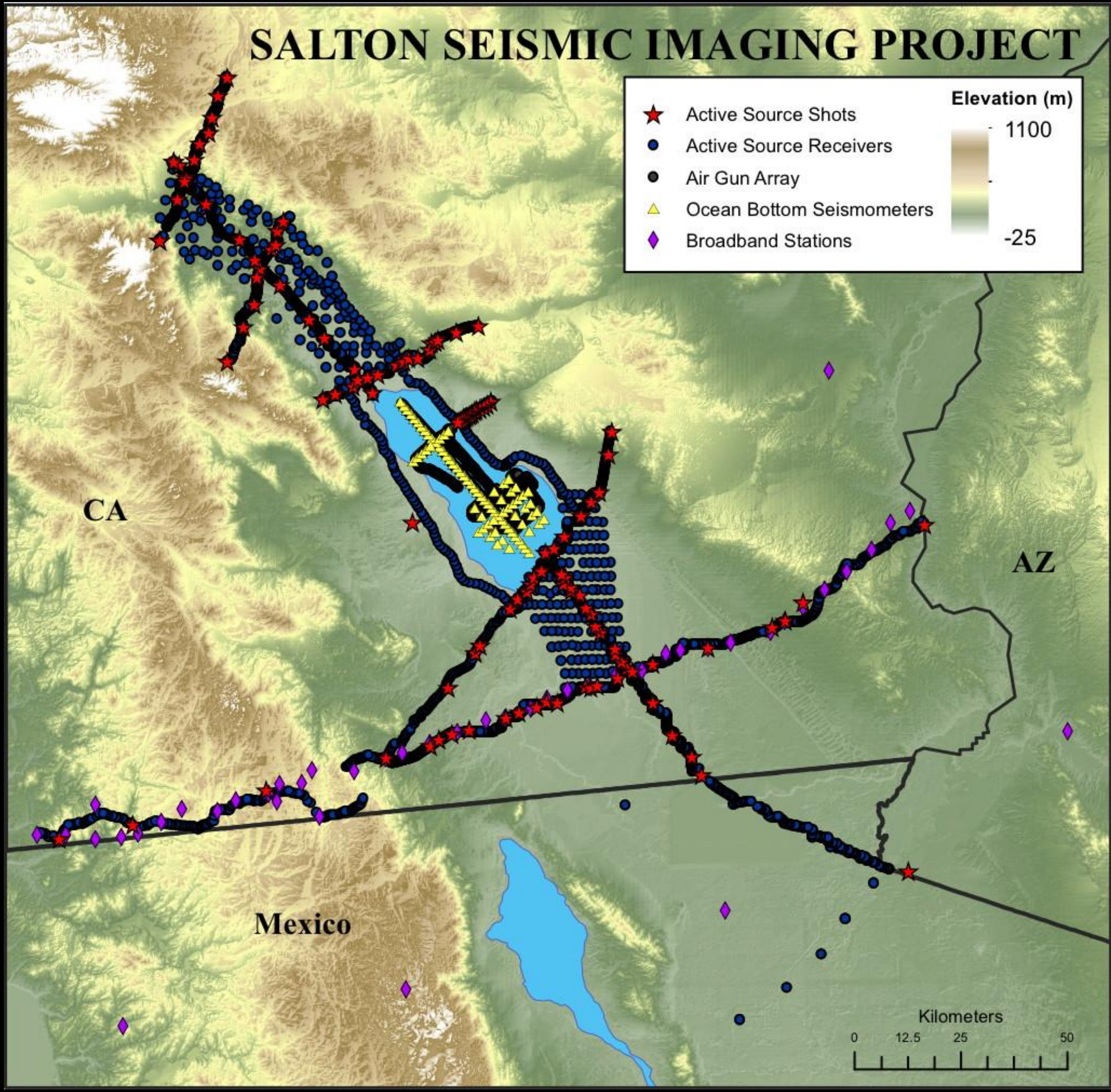
Current collaboration model: project Wiki (based on MindTouch) for planning purposes, evolved to interchange data and keep record of all presentations & products

SSIP

SSIP
onshore seismic refraction & reflection

Wet-SSIP
marine seismic refraction & reflection

Broadband-SSIP
onshore broadband teleseismic

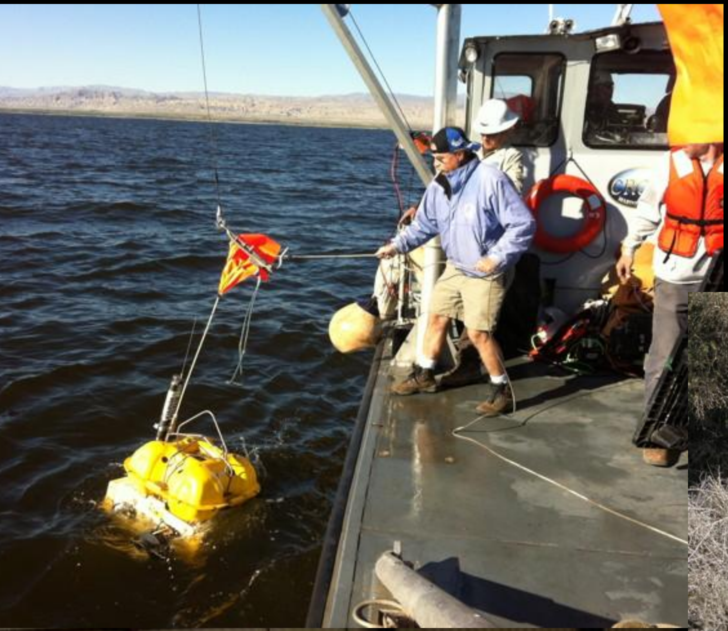


SSIP

Feb. – March 2011

explosive shots
- **126 shots**
- **median 115 kg**
- **up to 1400 kg**

SSIP+Wet-SSIP



Feb. – March 2011

seismometers:

- **3856 1-component Texan sites**
- **272 3-component RT130 sites**
- **78 3-component OBS sites**

Swath bathymetry data

The multibeam system is built into the hull of the ship, or in a separate towed unit. Transducers in a cross-array pattern emit and receive 12 kHz energy to image the seafloor beneath the array

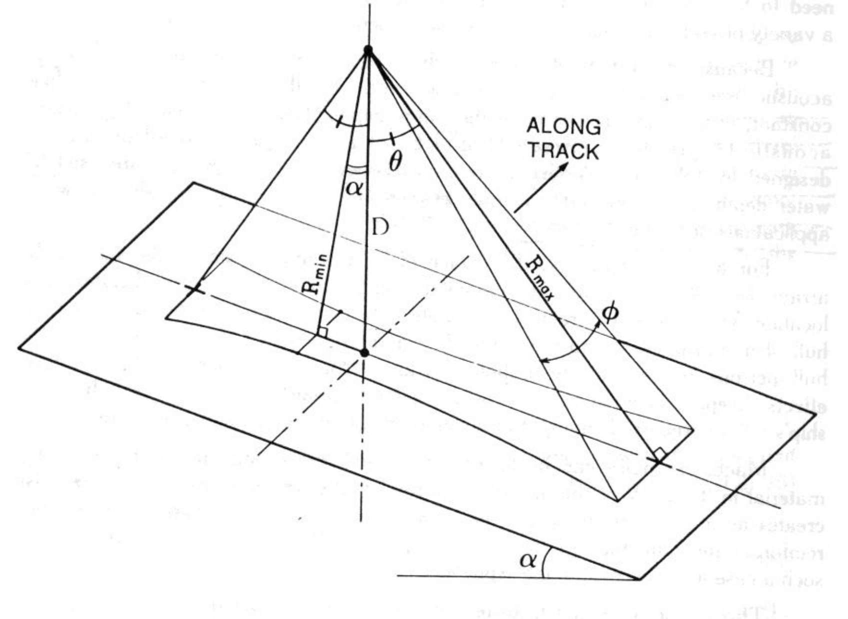


FIG. 3. — Ensonification geometry over a sloping bottom.

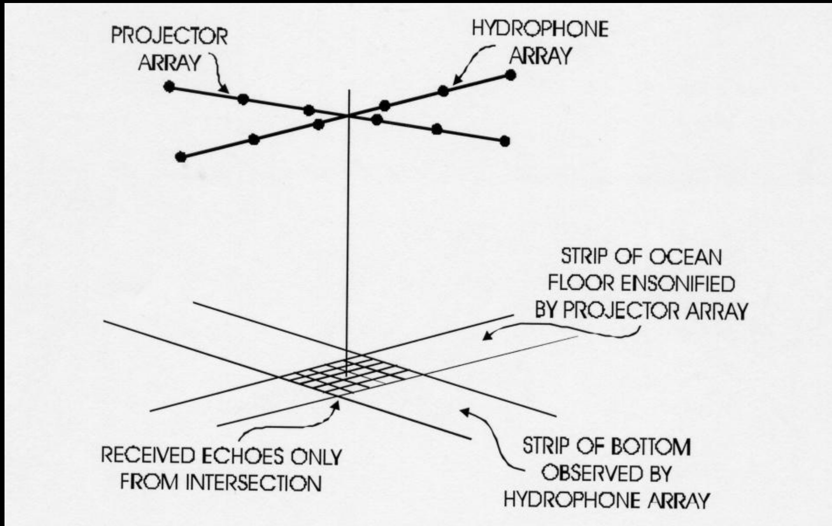
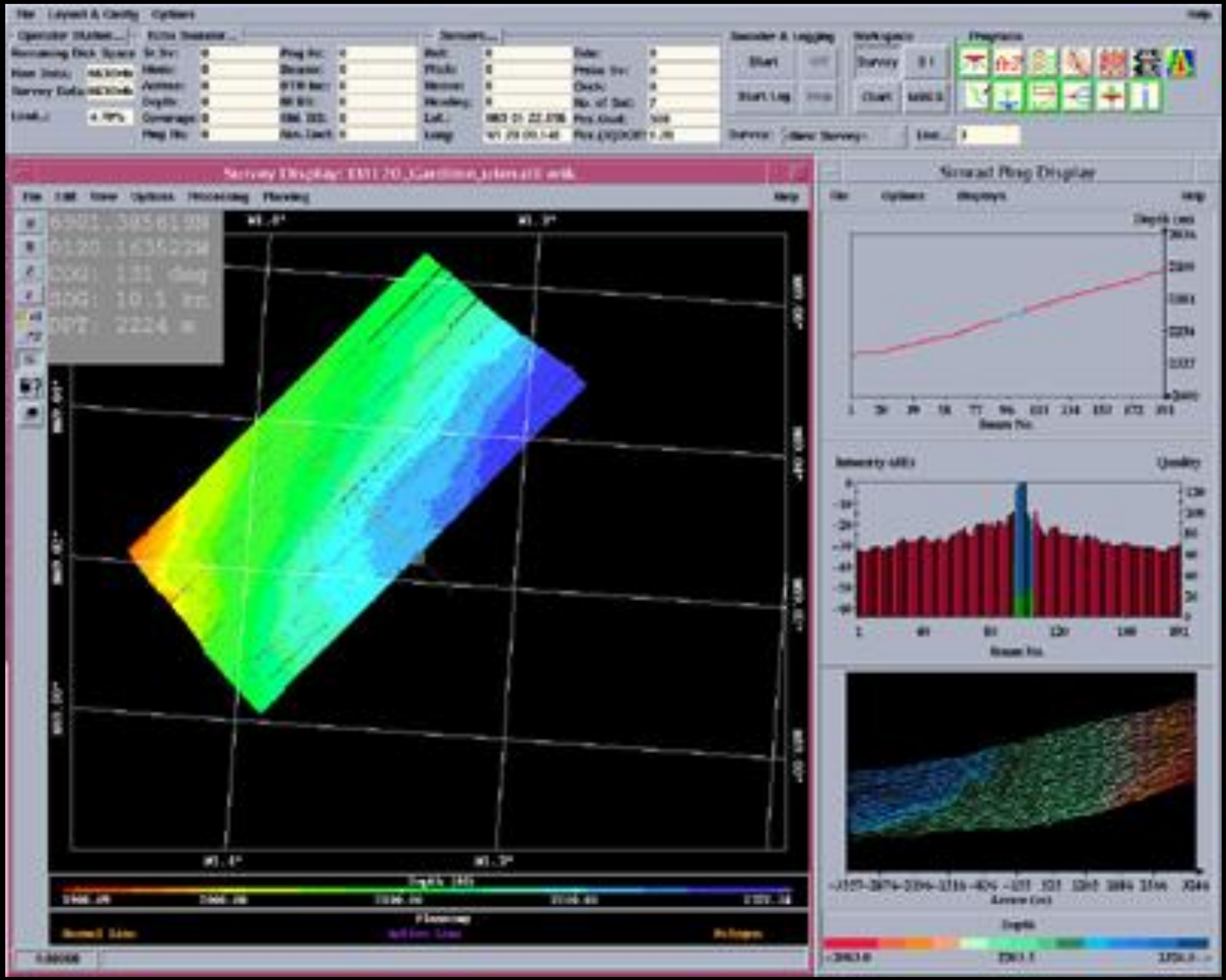
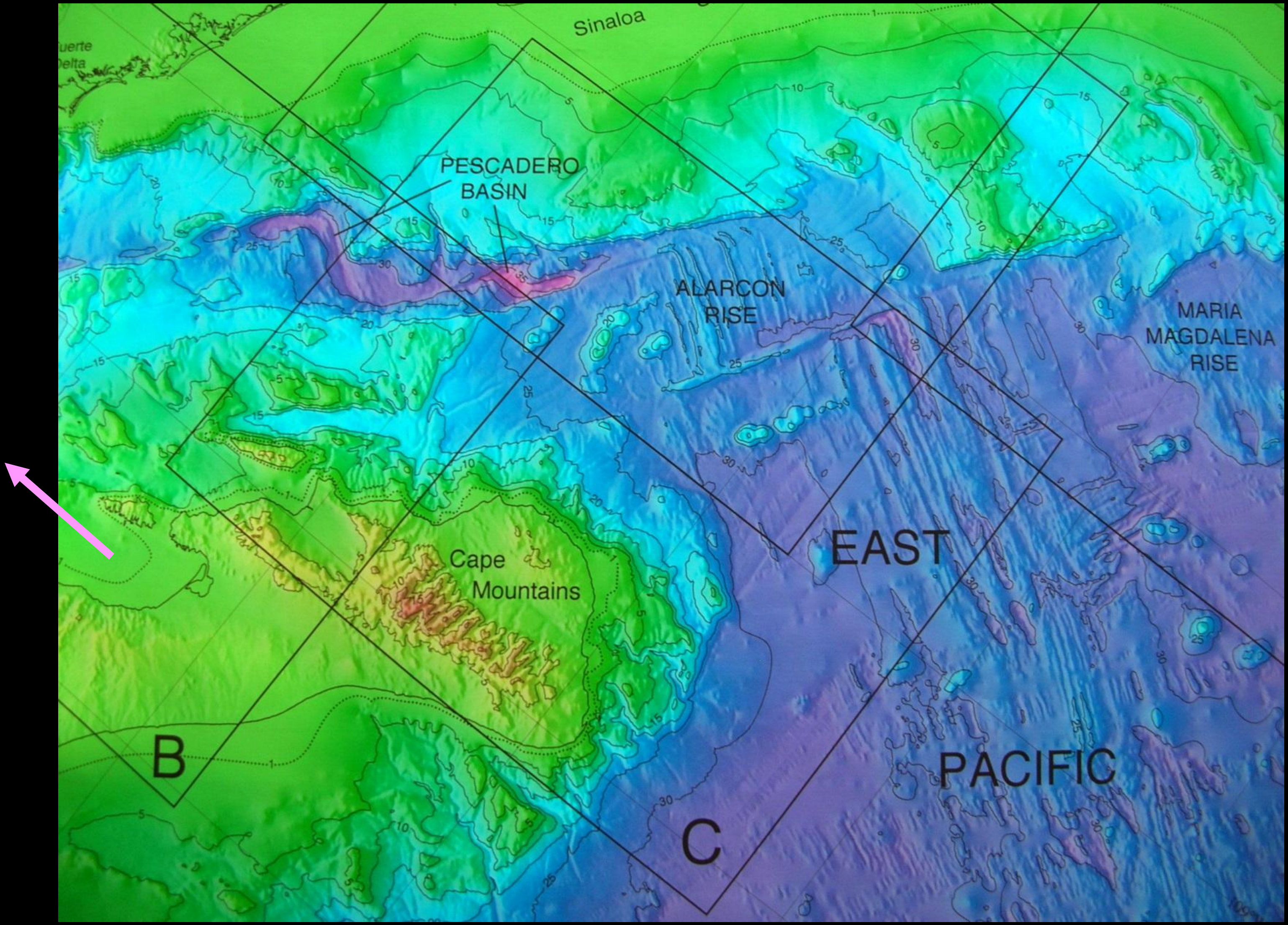


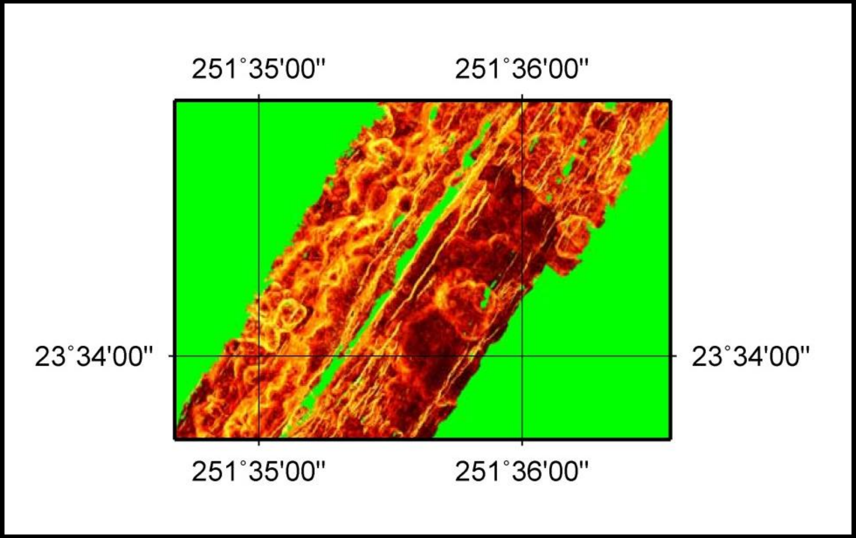
Figure Chapter 3 - -20: Projector and Hydrophone Arrays Arranged in a "Mills Cross"



Multibeam echo sounder shows the shape of the seafloor below



Bathymetry of the mouth of the Gulf of California from Lonsdale et al.
Alarcon Rise is the southernmost Pacific-North America spreading segment



- Example data file from East Pacific Rise in southern Gulf of California
- Colors show slope of seafloor
- Data analyzed using MB-system software

We would like to have joint real time work sessions to analyze these data – but data files are large (compressed size 15-100 Mbytes/ hour of survey, depending on water depth and type of system used)

Marine Geophysical Data

Wish List:

Fast, redundant connections between seismic data centers in both countries

- MOOC – virtual geology field trips
- needs 3D capability;
- can accommodate larger groups of students than possible on regular field trips

Participation by cross-border researchers on committees of graduate students (MS, PhD) at CICESE, UniSON, UNAM (D.F.). We need technology to have video conferencing routinely available with ability to see & hear all participants, share data files in real time

Expand capability & integration of mobile devices

Mexican Geophysical Union webinar series

Once – per – month presentation over the internet on a science topic of interest to the members of the Mexican Geophysical Union.

One hour talk + discussion; viewers register from any web enabled computer, and receive link to connect and listen to and watch the presentation.

Talks can originate at various locations (UNAM, Mexican Petroleum Institute, ERNO, etc). Powerpoint file with presentation can be downloaded later

Difficulties encountered with speed of transmission, loss of audio

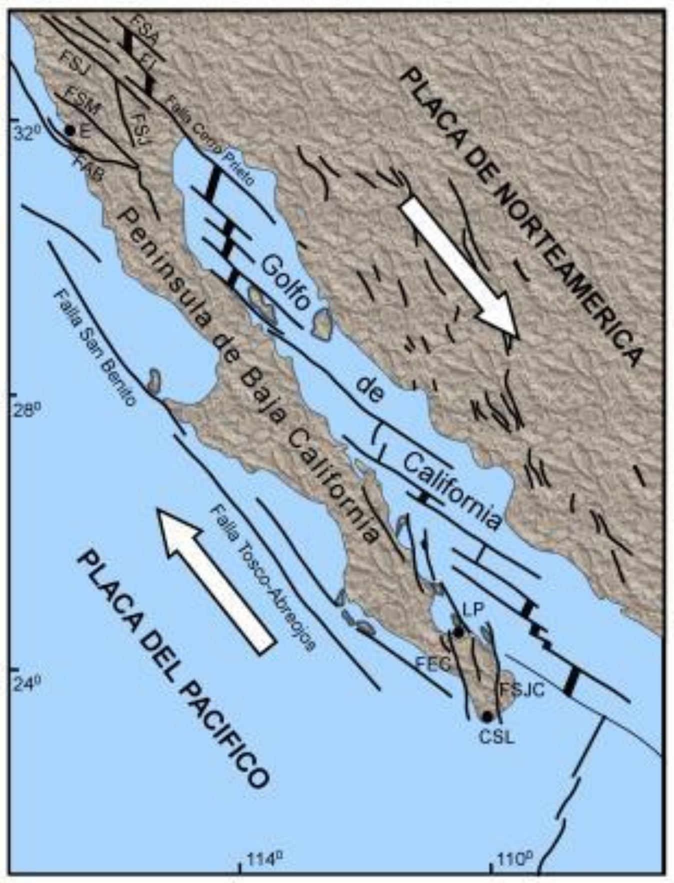
Earthquakes Monitoring and Sharing Data in Real Time in Baja California

Victor Wong Ortega
Seismology Department
Earth Sciences Division - CICESE

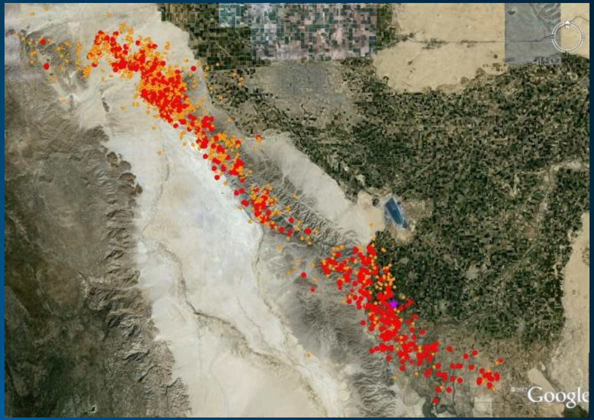
October 10-11, 2013

BAJA CALIFORNIA:

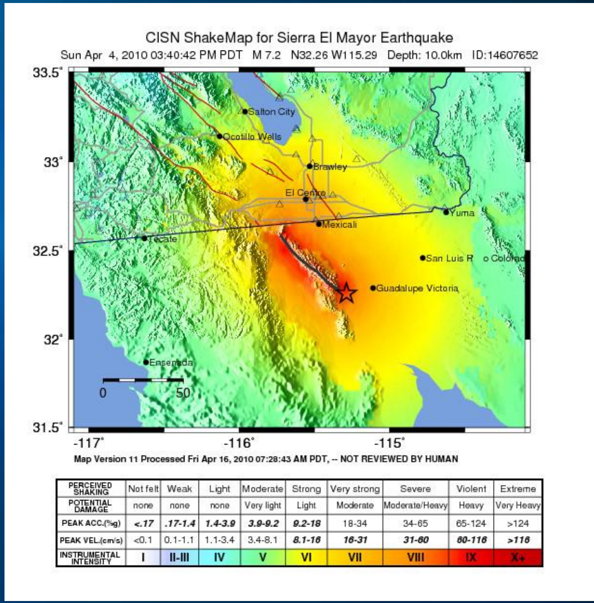
- Baja California is located in the tectonic border of the Pacific and the North America plates.
- This region is tectonically active and generates seismic activity continuously.
- This region is capable of generates moderate to big earthquakes that could be dangerous to populated region



After the bigger earthquake never recorded in Baja



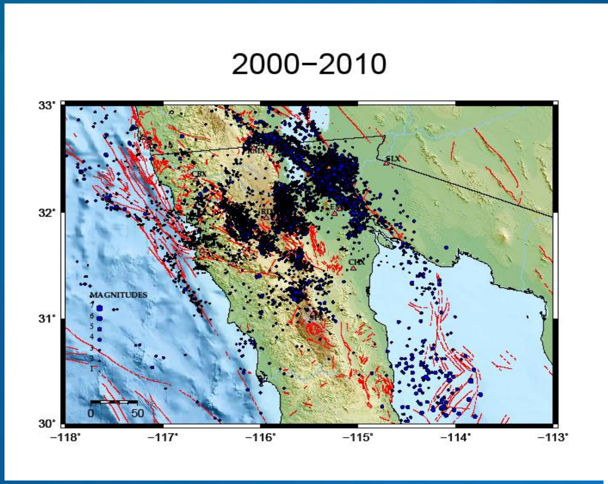
The 7.2 (Mw) magnitude El Mayor Cucapah earthquake of April 4 of 2010, evidence the lack of a well instrumented and dense seismic network in the region that could prevent the local and federal Mexican authorities from rapidly assessing the impact of the quake and its effects on the peoples and infrastructures.

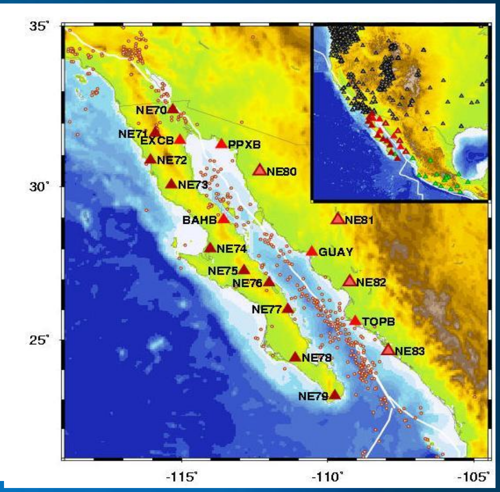


What have we done?

- Instrumentation of 42 new seismic stations in the region of Mexicali Valley, Mexicali and Tijuana cities connected to CICESE in real time by internet.
  - 23 accelerometers stations, 5 in Tijuana and 18 in Mexicali.
  - 26 Broadband and accelerometer seismic stations in the Mexicali Valley and in the Baja California state.
- Agreements of collaboration with Caltech-USGS-CENAPRED-CICESE to monitor and share seismic data in real time.

We know were the earthquakes occurs … but no when.

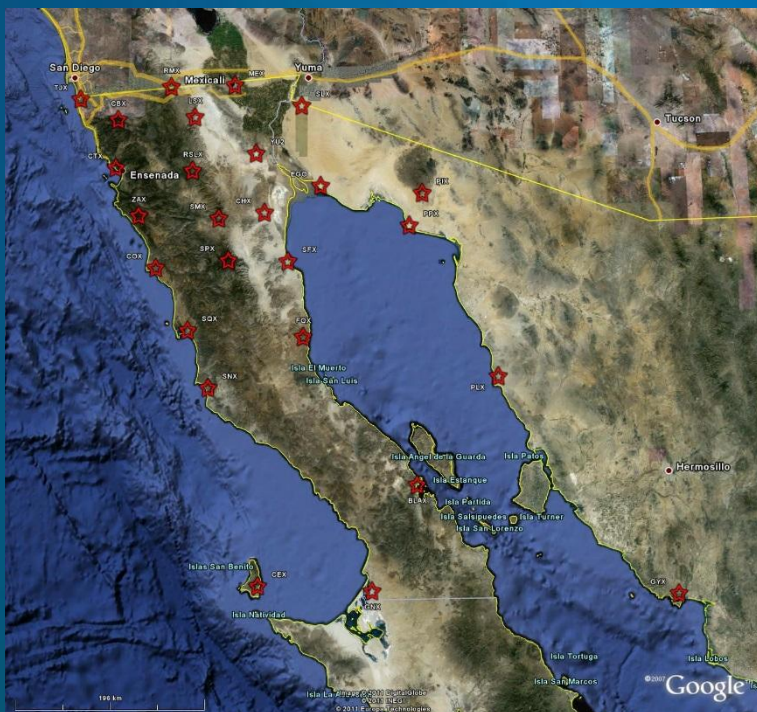




That why we need to be monitoring 24/7 the seismic activity to observe any change in the seismic activity behavior.

¿What next...?

Monitoring the seismic activity in real time to inform the local, state and federal authorities about the earthquakes with relevant magnitude that are occurring in the region.



¿How we can do this:

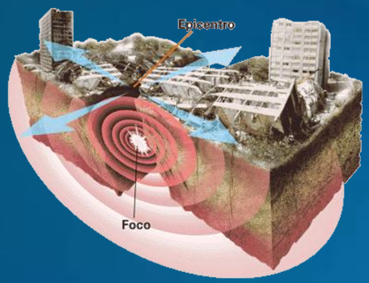
1. radio?
2. Telephone?
3. Commercial Internet?
4. Satellite internet?

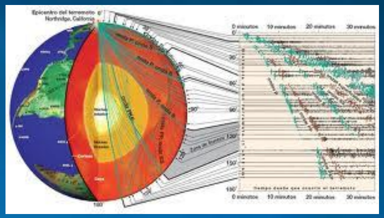
Each red star is a 6 channel (3 acceleration, 3 BB velocity) seismic station, 100 sps, 24 bits resolution.

¿Services... seismic report and seismic alert in real time ?

- Automatic earthquake location and magnitude, 5 minutes after the occurrence.
- Earthquake information by SMS and WEB pages to people, media and authorities.
- Earthquake higher intensities recorded, 30 minutes after the earthquake and the intensities distribution maps (Shake Maps).
- Sharing seismic data with SSN, SCSN-Caltech and others...

Seismic monitoring and its objective…

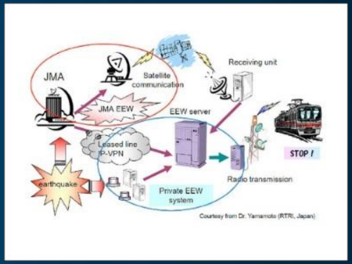




Earthquake
Origen time

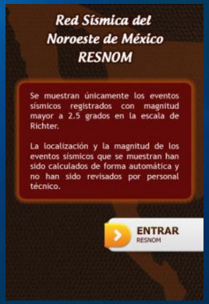
Wave
propagation

Earthquake
analysis



Information
flow

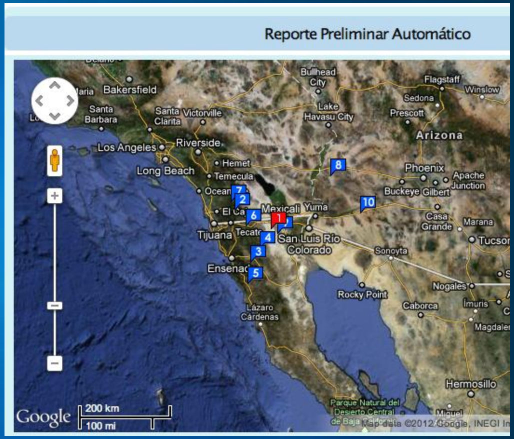
Seek for a safe
place



That mean sharing scientific data…

400

- CICESE and USGS will deploy an information-sharing system to be better prepared when a seismic event occurs.



- This implies transferring very large volumes of data among these institutions – on the order of hundreds of Gigabytes. (Middle of 2013)

Big Data and Network Link

- This will allow the implementation of high performance cyber infrastructure for research and development.
- Establish new ways to collaborate and sharing knowledge in USA and Mexico countries.
- It also represent challenges in order to find new ways of sharing information.

Thanks…

Real - Time Station

Big Data Big Network
Big Opportunities for Collaboration between Mexico and California
10-11 Oct 2013 • Ensenada, Mexico

Space Image Repository

Enrique Pacheco-Cabrera
Deputy Director for Space Science and Technology
Mexican Space Agency

BOARD OF GOVERMENT

FEDERAL GOVERMENT

- PRESIDENT
  Minister of SCT
- 6 Deputy Ministries
  SEGOB
  SRE
  SEP
  SHCP
  SEDENA
  SEMAR

PUBLIC ENTITIES

- CONACYT
- ANUIES
- INEGI

ACADEMIC

- UNAM
- IPN
- Mexican Academy of Sciences
- Engineering Academy
- National Academy of Medicine

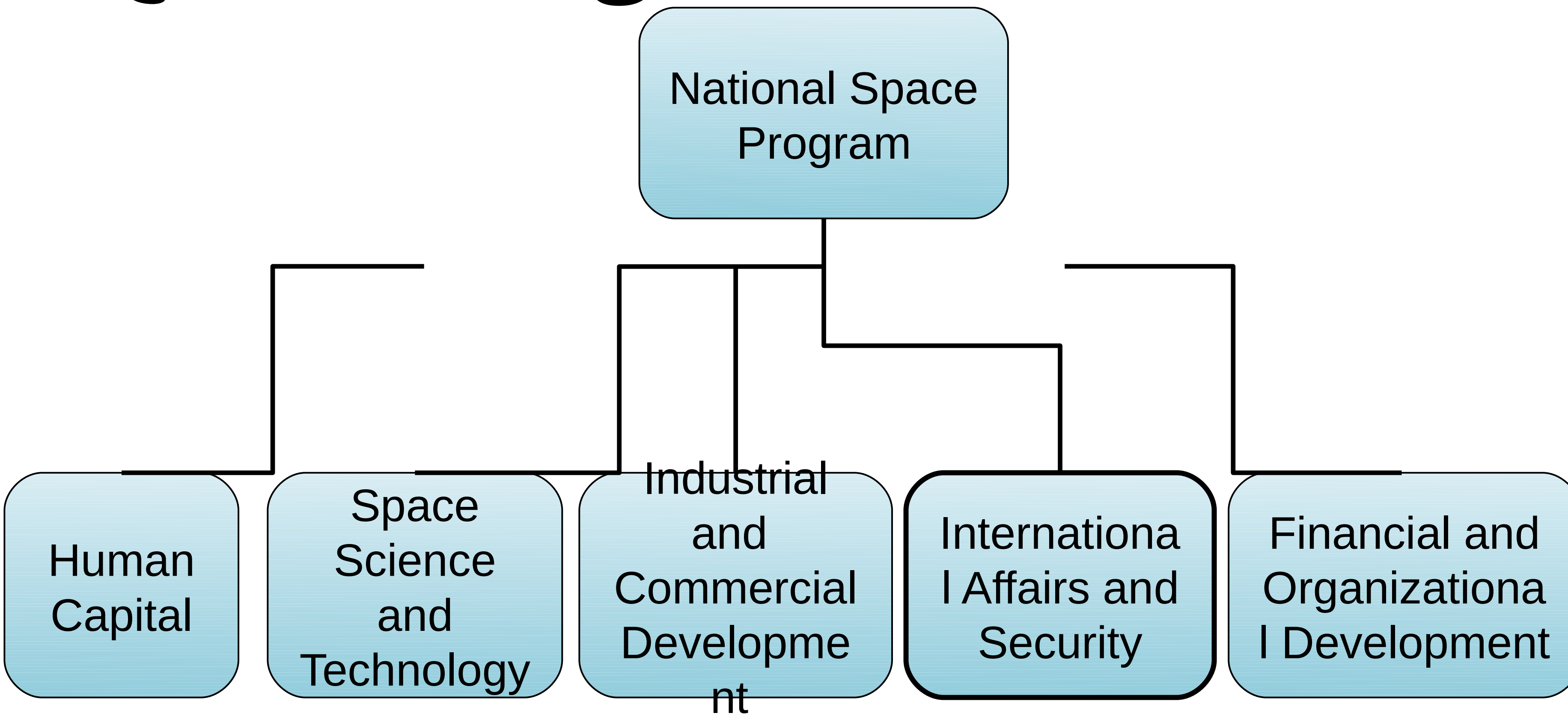
National Space policy

1. The State as authority and responsible	7. Productive sector development
2. Autonomy of the Country	8. Human capital
3. Sovereignty and Security Protection	9. Coordination, regulation and certification
4. Protection of Mexicans	10. International cooperation
5. Environment sustainability	11. Space activities public awareness
6. Scientific research, technology and innovation development	12. Funding
	13. Organization and management

Priorities for the AEM

Collaboration Agreements and MoU's

Structure of the National Space Program



Organization

Space Science and technology Development Coordination

Strategies on the PNAE for Science and Technology

Strategies on the PNAE for Science and Technology

Strategies on the PNAE for Science and Technology

Challenges

Economic

Time and Cost of Projects

Government as sole source of funding

Necessities and short time vision

Management

Technical Knowledge

Long Term policies

Social impact perception

Space Infraestructure

Space infrastructure is the backbone that supports and connects satellite space systems for national security applications, disaster management and early warning, connectivity, social benefits, environmental sustainability and scientific and technological research

Space Infraestructure

The set of tangible and intangible assets needed for the study, access, exploration, use and exploitation of space

- Tangible assets include rockets, launch platforms and systems, suborbital vehicles, satellites and other spacecraft, instrumentation, payloads, ground stations, teleports, receiving antennas, user terminals and other devices for link control.

- Intangibles assets include orbital positions and frequencies associated, laws, regulations, technologies, patents, licenses, concessions, trademarks and "know-how".

STRATEGY

a) National Development Plan

b) National Infrastructure Plan

c) Proposal of a Early Warning System as part of the space infrastructure for protection of the Mexicans and the natural resources and other public infrastructure.

Recent Developments

- AEM Board give the approval to request the funding for two EO satellites, one thermal-infrared (MIROS) and one high resolution (better that 1m)
- Also the PND (National Development Plan) consider the action line for the Early Warning System for Disasters

The disaster management in Mexico

- Several entities involved
  - Early Warning
  - Disaster relief
  - Remediation
- Challenges
  - Communication, coordination, timing information

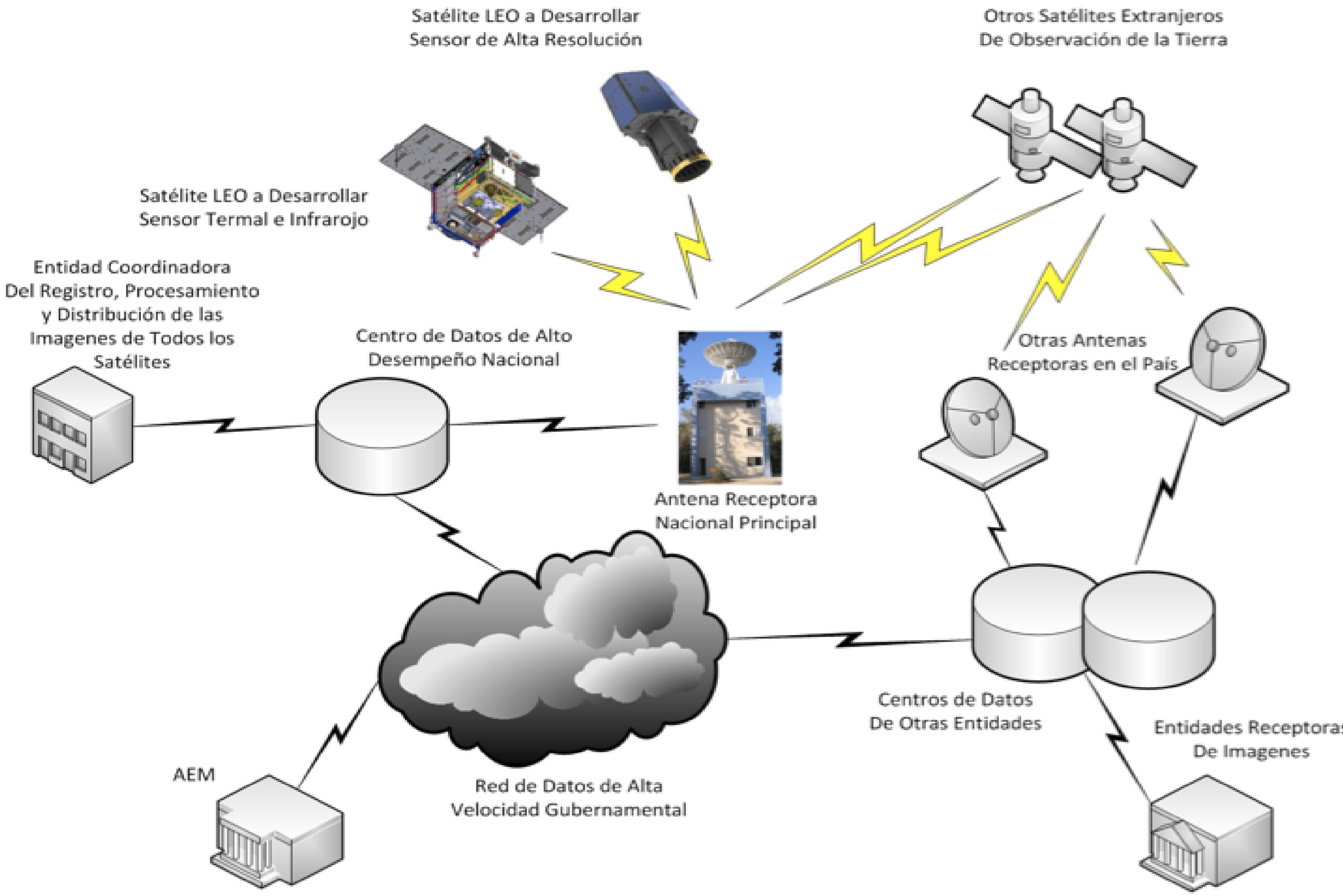
The BigData-BigNetwork Problem

- The space infrastructure for a early warning system, needs to develop the following:
  - National system for storage and data processing for Geomatics, Space and Astrophysicists Information (NSSDAP-DGSAI)
  - Connectivity
  - Applications

The NSSDAP-DGSAI

- A state-of-the-art storage facility (with backup) for a data flow of +1.5Petabytes, probably even more.
- A data processing facility to produce products for several entities initially part of the Government (Federal, State, Municipal)
- Connectivity between the main users, initially Federal Ministers.

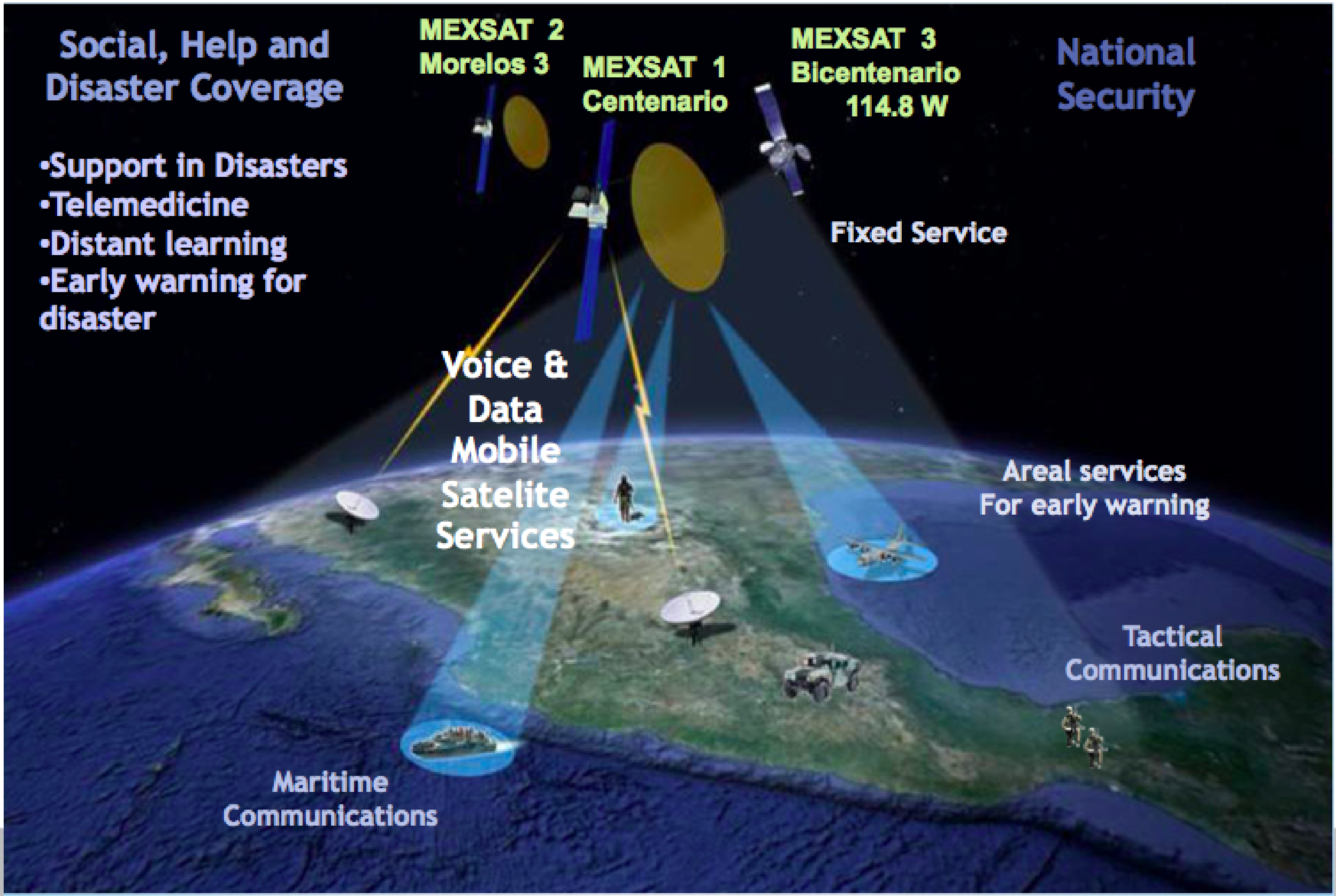
Space infraestructure for Earth Observation



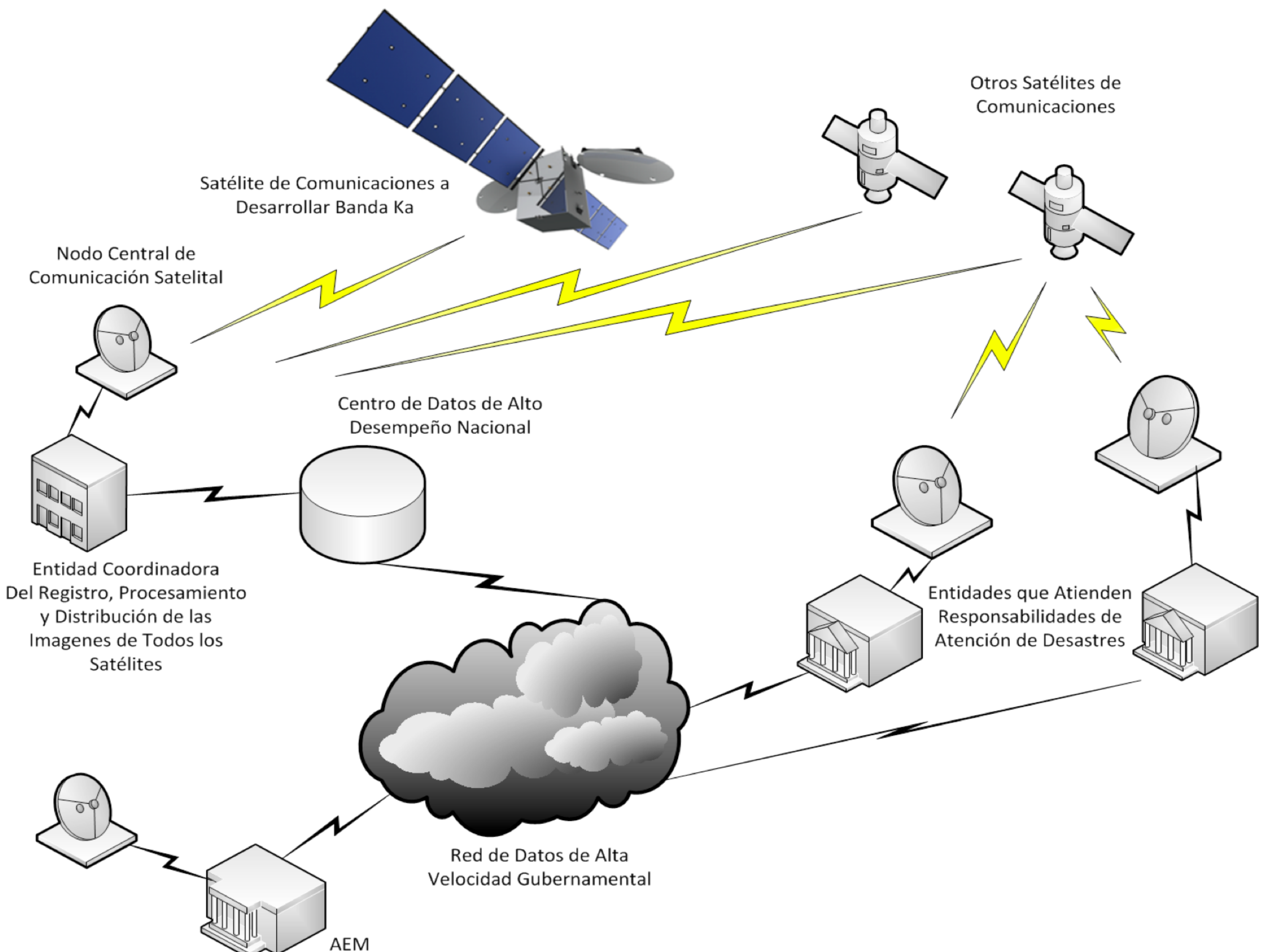
The satellite Last Mile

- The connectivity problem needs to address also remote rural areas where satellite is the only available option.
- New technologies and solutions based on satellite communications will also be needed.

MexSat Satellite System

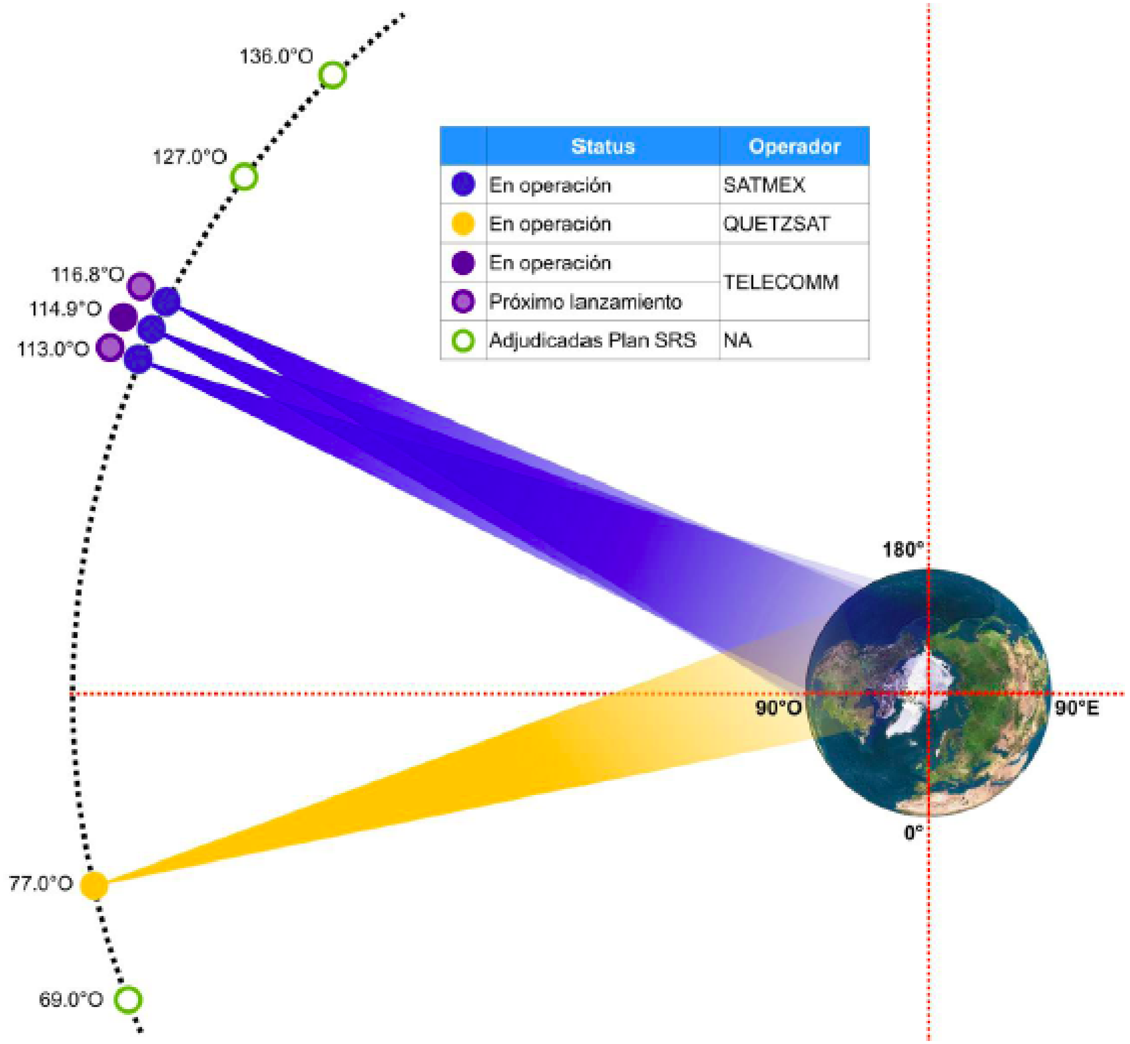


Space infraestructure for Satellite Comunications



Space infraestructure for Satellite Comunications

Resumen de recursos orbitales



ADDTIONAL ACTIONS

- Trust CONACYT-AEM as dedicated funding for space projects.
- Regional development supported by States
- International collaboration

SUMMARY

- Space infrastructure a change of vision
- Big-data problem associate with the use of space infrastructure.
- Capacity building base on the early warning system project.

Gracias

Click icon to add picture

pacheco.enrique

Integrating distributed LIDAR data repositories into the OpenTopography service infrastructure

Vishu Nandigam, Chaitan Baru

San Diego Supercomputer Center, UC San Diego

Big Data Big Network Workshop

Oct 10-11, 2013

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

ASU
ARIZONA STATE UNIVERSITY

NSF

LIDAR

- LIDAR = **LI**ght **D**etection **A**nd **R**anging (aka airborne laser swath mapping)
- Billions of of accurate distance measurements with a scanning laser rangefinder + GPS + Inertial Measurement Unit (IMU)

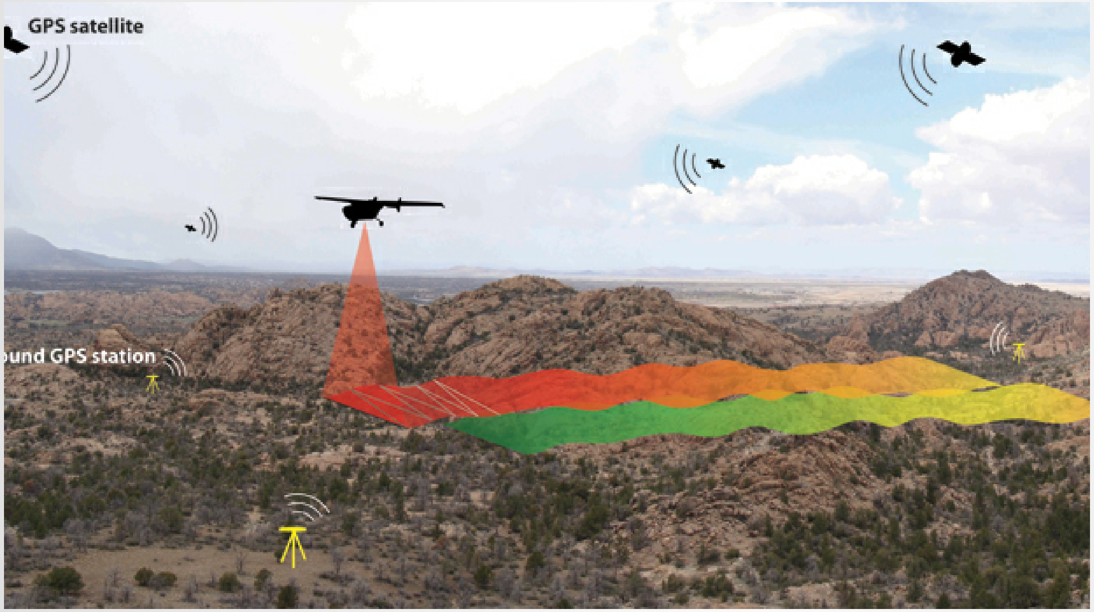
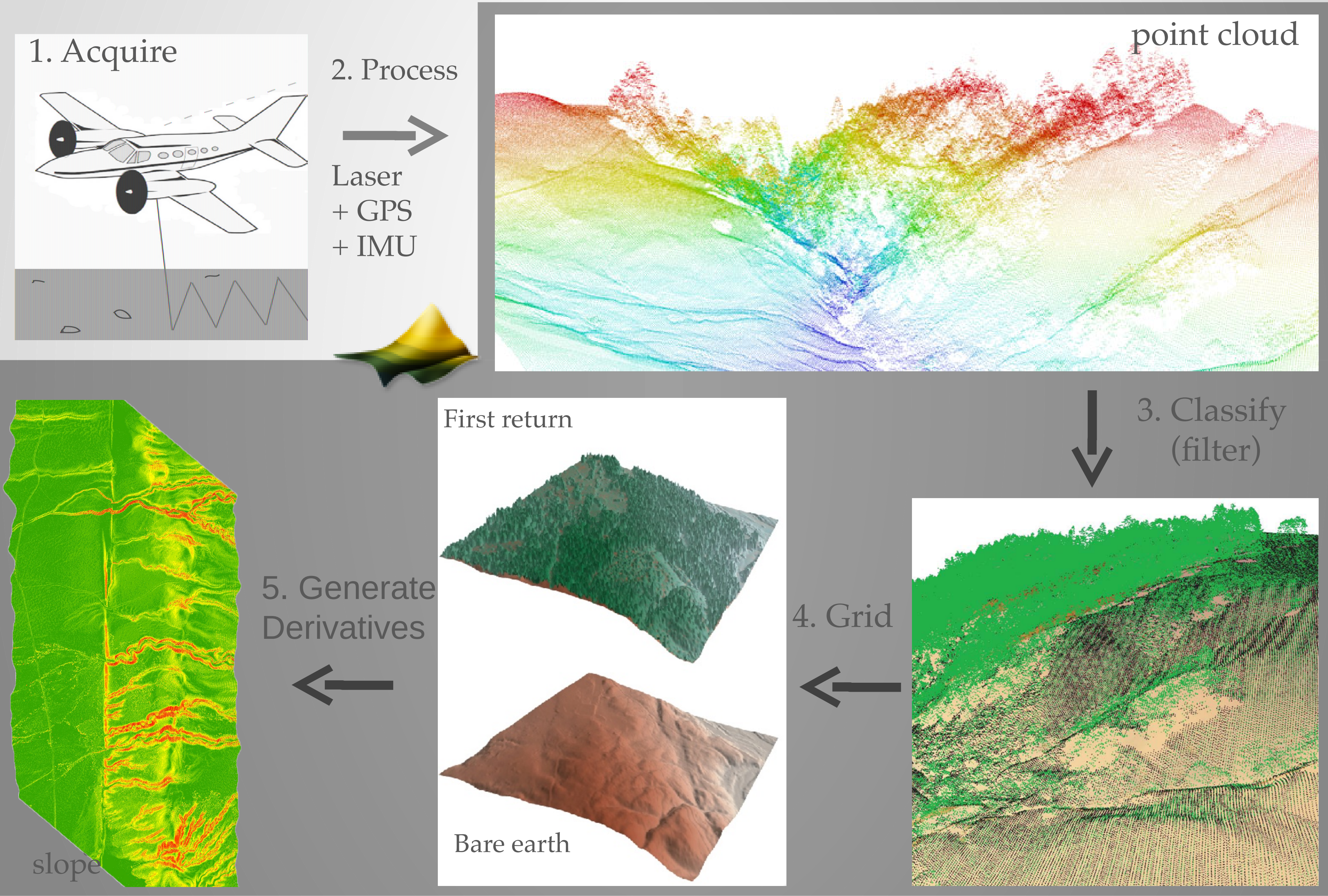


Image: David Haddad, AGS

- Beam Diameter 15-20 cm
- Vertical Accuracy ~15 cm
- Point cloud (x,y,z coordinates) = fundamental LIDAR data product

Airborne LIDAR Workflow



OpenTopography: Motivation

- Democratize access to data to achieve greatest scientific impact.

- Big data:
  - Treat data as an asset that can be used and reused
  - Co-locate data with on-demand processing

OpenTopography

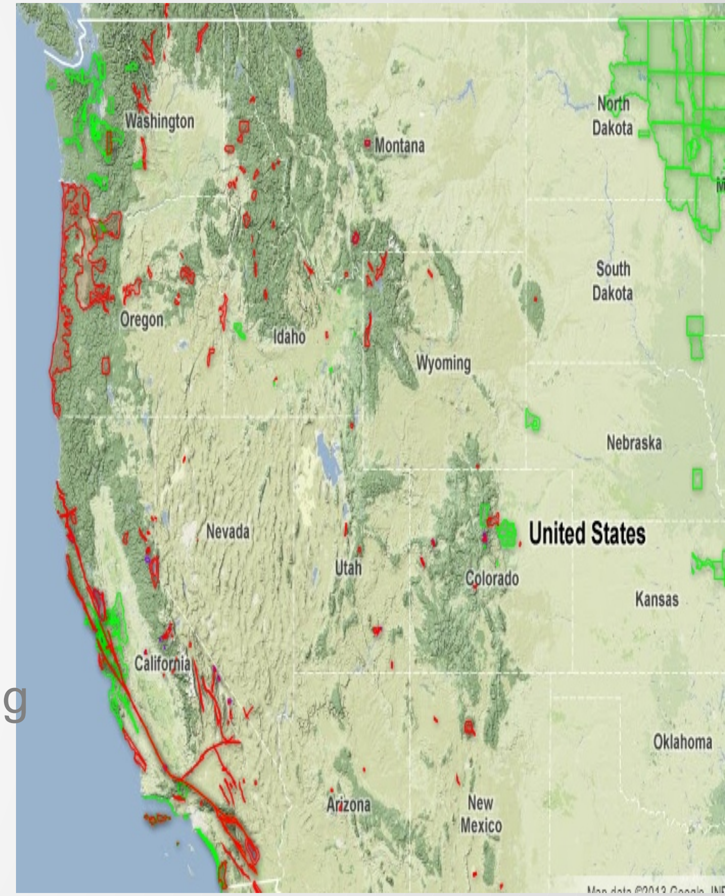
- Originally developed as part of GEON
- NSF Earth Science Facility: 3 year support in 2009. Renewed in 2012
- CI and Science Partnership
    - SDSC & ASU
- Related research efforts
    - NASA ROSES: Extend to Satellite-based LIDAR (waveform data)
    - NSF SI2 CyberGIS: OpenTopo as an exemplar of cyber GIS
    - NSF CluE: Investigate Computer Science issues in big data
- Partnerships with state and local agencies to support data hosting

OpenTopography: Tiered Access to Data

- Large user community with variable needs and levels of sophistication.
  - Processed and easy to access product for synoptic data browsing and education.
    - Google Earth
  - Vast majority of users want a standard gridded product:
    - Optimized ("standard") DEM
  - Co-locate "raw" point cloud data next to processing tools to allow users on-demand data access to optimize terrain representation
-

OpenTopography Data Status

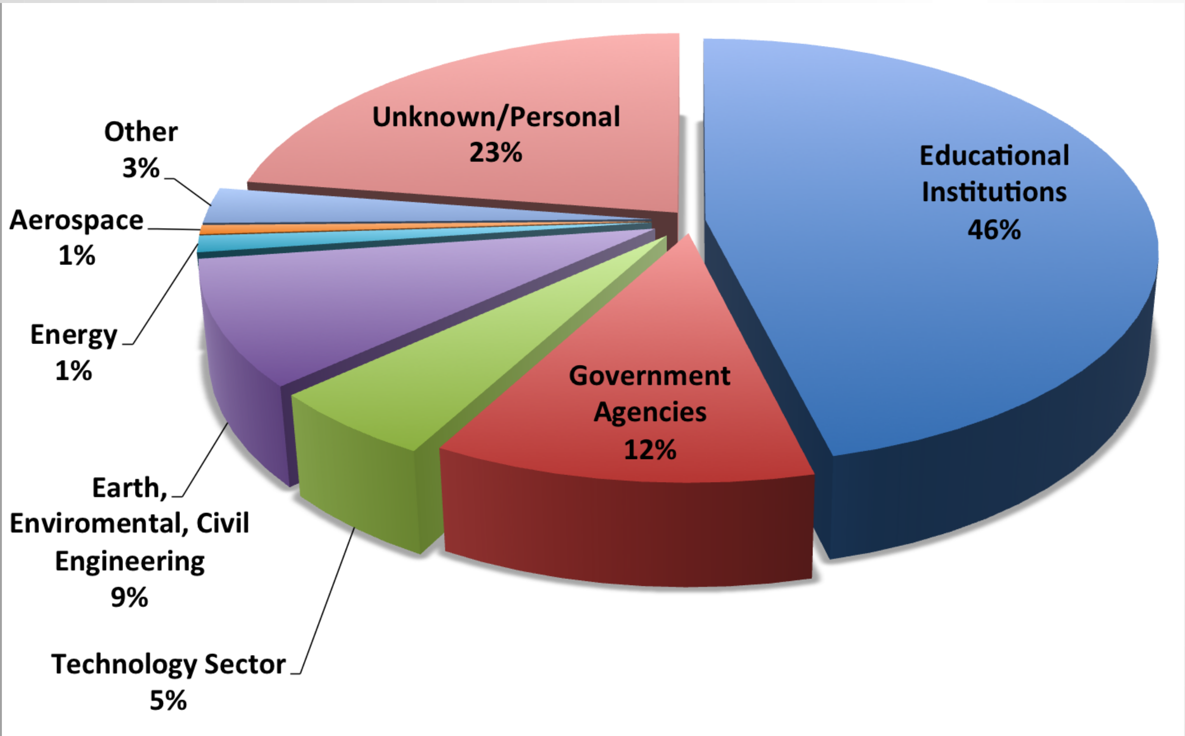
- Data Stats
  - LIDAR returns: ~597.5 Billion
  - Size on disk: 15.5+ TB (compressed PC)
  - Area covered: 121,055 km2
  - 160 datasets and growing

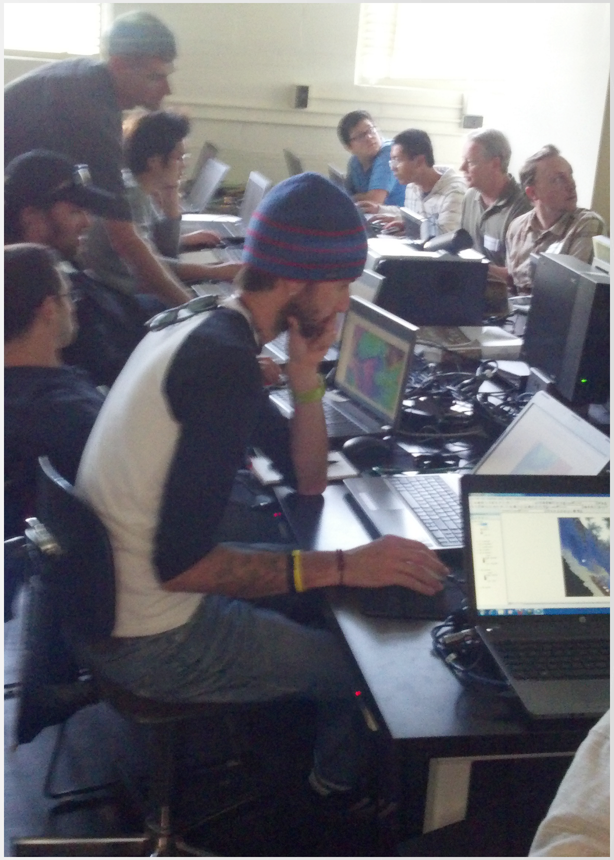


- MOUs and Partnerships
  - NSF: NCALM, UNAVCO, CZOs, LTER
  - Other: World Bank, Tahoe Regional Planning Authority, Teton Conservation District, Oregon LIDAR Consortium, Idaho LIDAR Consortium
  - Service Agreements: State of Indiana Watershed Sciences Inc (for PG&E)

Usage Metrics

- 24,617 jobs
- 701 billion points (currently >30 billion pts/month)
- 3500+ unique users





Significant # of student users

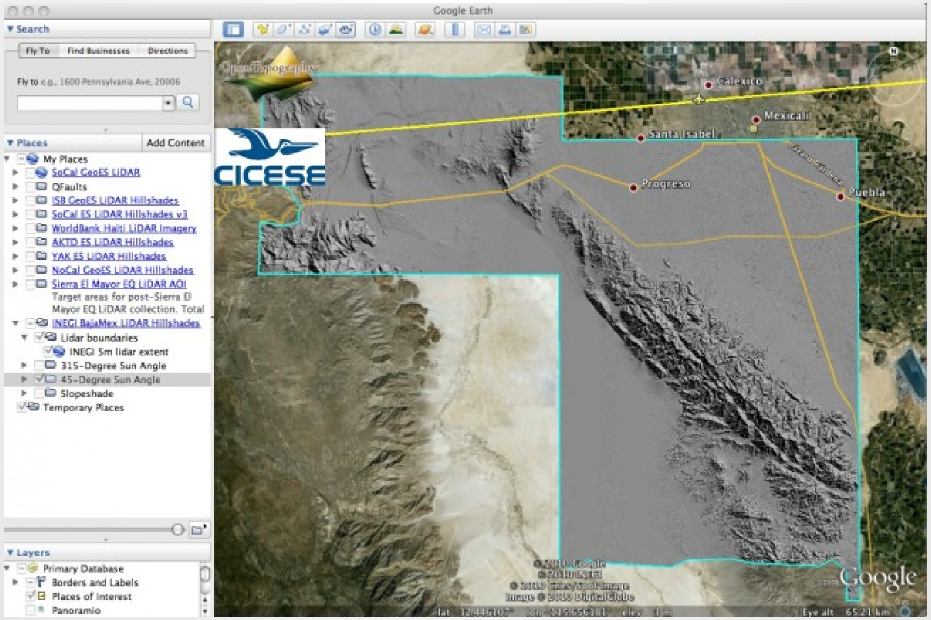
2010 El Mayor-Cucapah Earthquake

- 7.2 magnitude
- 16 miles south of Guadalupe Victoria, Baja California, Mexico

Scarp from 2010 El Mayor-Cucupah earthquake

Pre- El Mayor-Cucapah Earthquake Data

- **Alejandro Hinojosa** at **CICESE** in Ensenada, Mexico
- **5 meter resolution** LIDAR topography data for the epicentral region in northern Baja, Mexico. **2,000 km2** southwest of Mexicali, acquired in 2006 by the Instituto Nacional de Estadística y Geografía (INEGI)
- Available hillshades from OT portal on **Monday**.
- **Pre-event data** present an exciting opportunity for comparing pre- and post-event data to calculate near-field deformation along the rupture.



El Mayor-Cucapah Earthquake Rupture Data

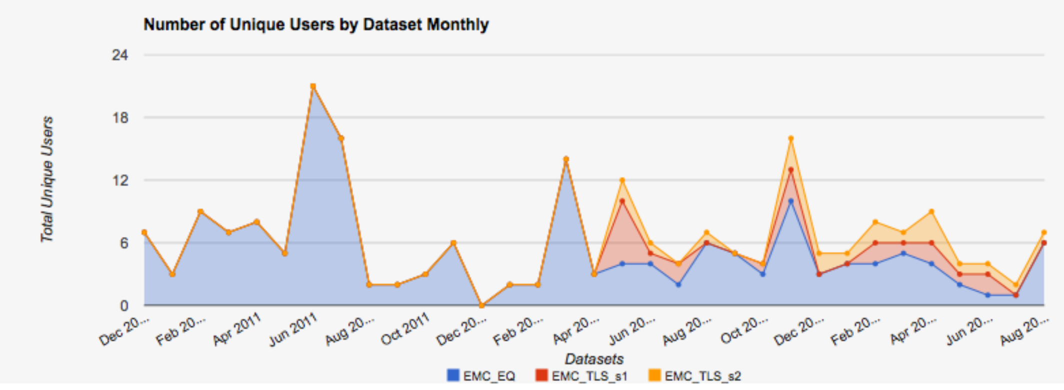
- Airborne LiDAR
  - El Mayor-Cucapah Earthquake (4 April 2010) Rupture LiDAR Scan
- Terrestrial Laser Scanner
  - El Mayor-Cucapah Earthquake Rupture Terrestrial Laser Scan-Site 1
  - El Mayor-Cucapah Earthquake Rupture Terrestrial Laser Scan-Site 2

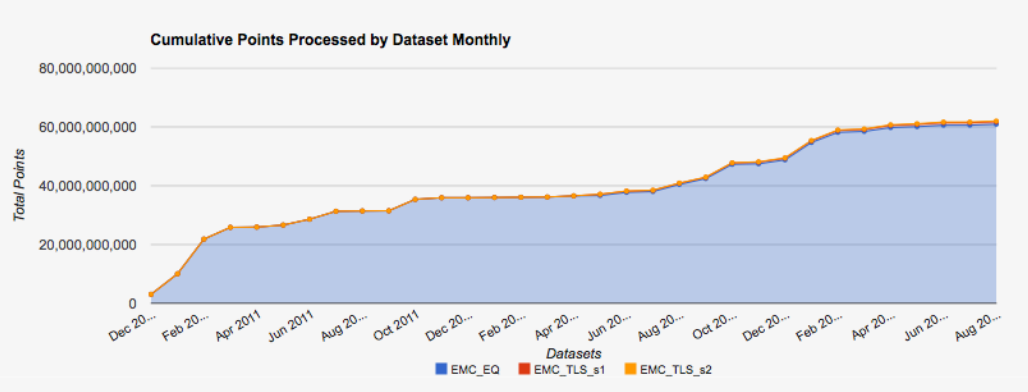
El Mayor-Cucapah Earthquake (4 April 2010) Rupture LIDAR Scan

- Full dataset freely available from OT - 372 km2
- 3.3 billion LIDAR Point Cloud Returns
- 818 custom jobs run by 115 unique users with 60 billion + returns
- DOI: 10.5069/G9TD9V7D

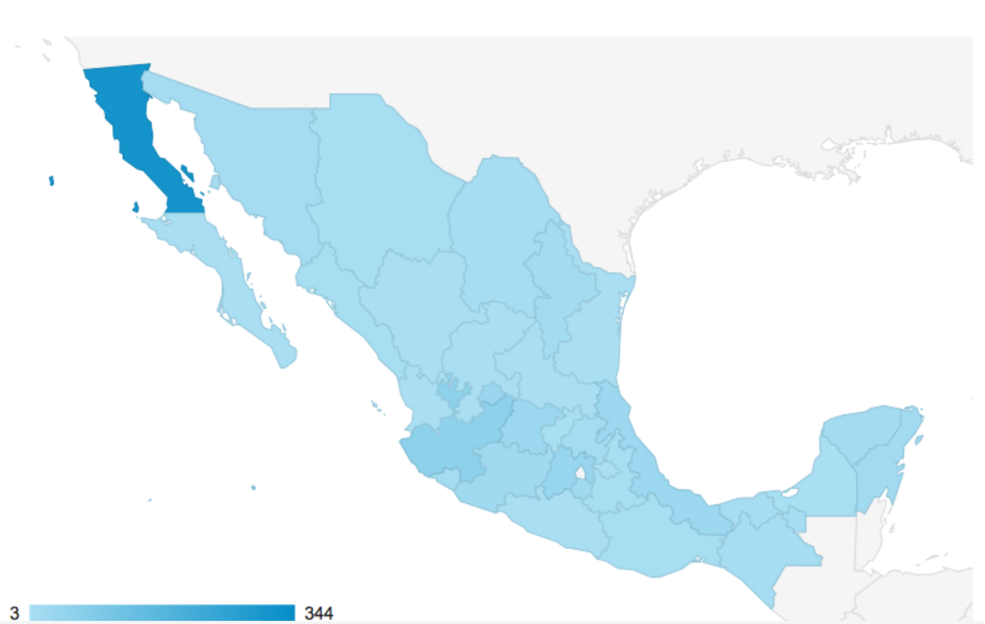
Data Usage Statistics

	Dataset	Number of Jobs ↓	Total Points	Number of Unique Users
1	El Mayor-Cucapah Earthquake (4 April 2010) Rupture LiDAR Scan	858 jobs	62,044,494,903 points	127 users
2	El Mayor-Cucapah Earthquake Rupture Terrestrial Laser Scan-Site 1	30 jobs	938,298,013 points	21 users
3	El Mayor-Cucapah Earthquake Rupture Terrestrial Laser Scan-Site 2	21 jobs	242,576,746 points	20 users
	All Datasets	909 jobs	63,225,369,662 points	157 users

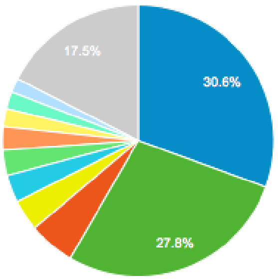




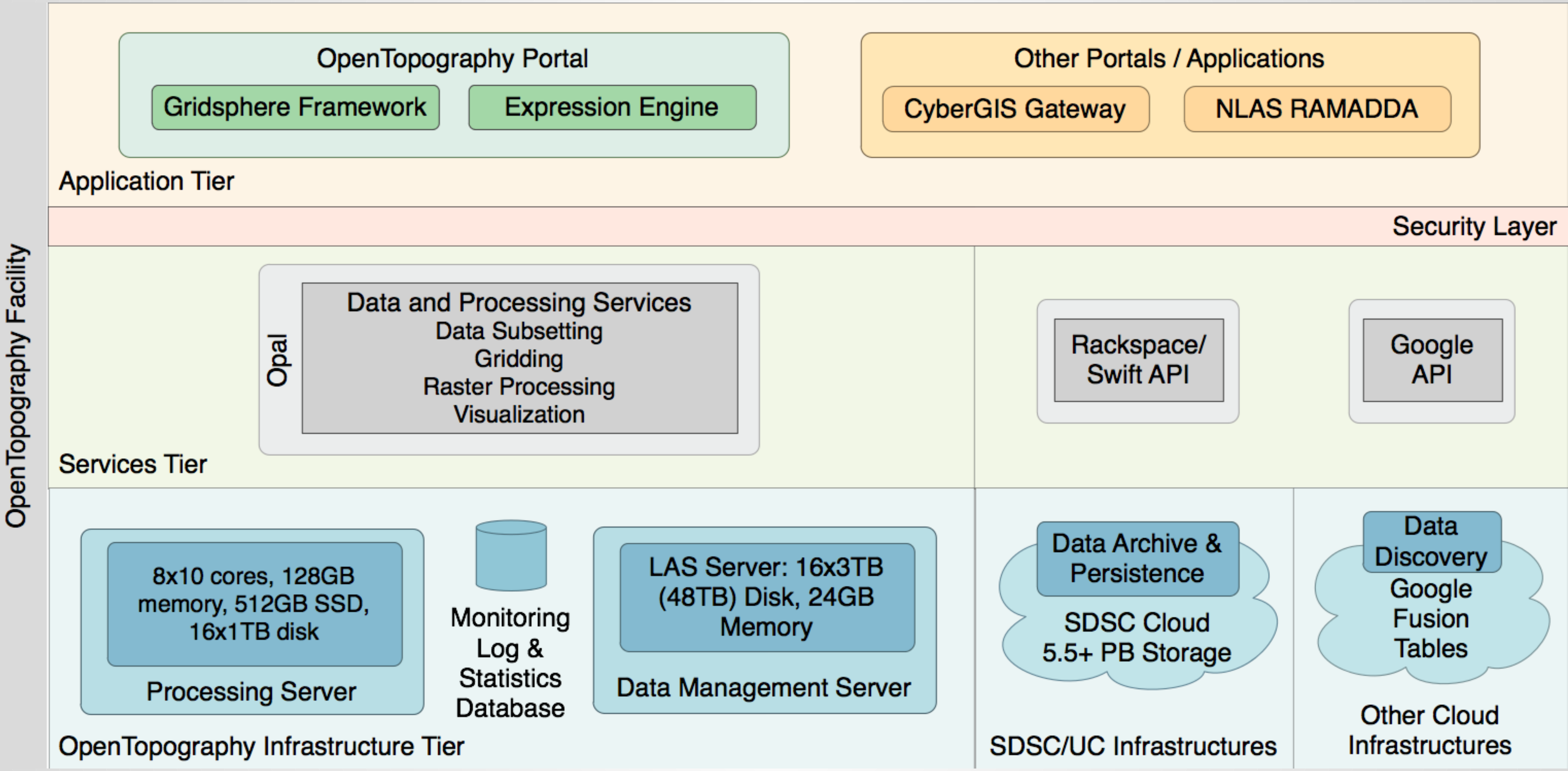
Site Visitor Statistics



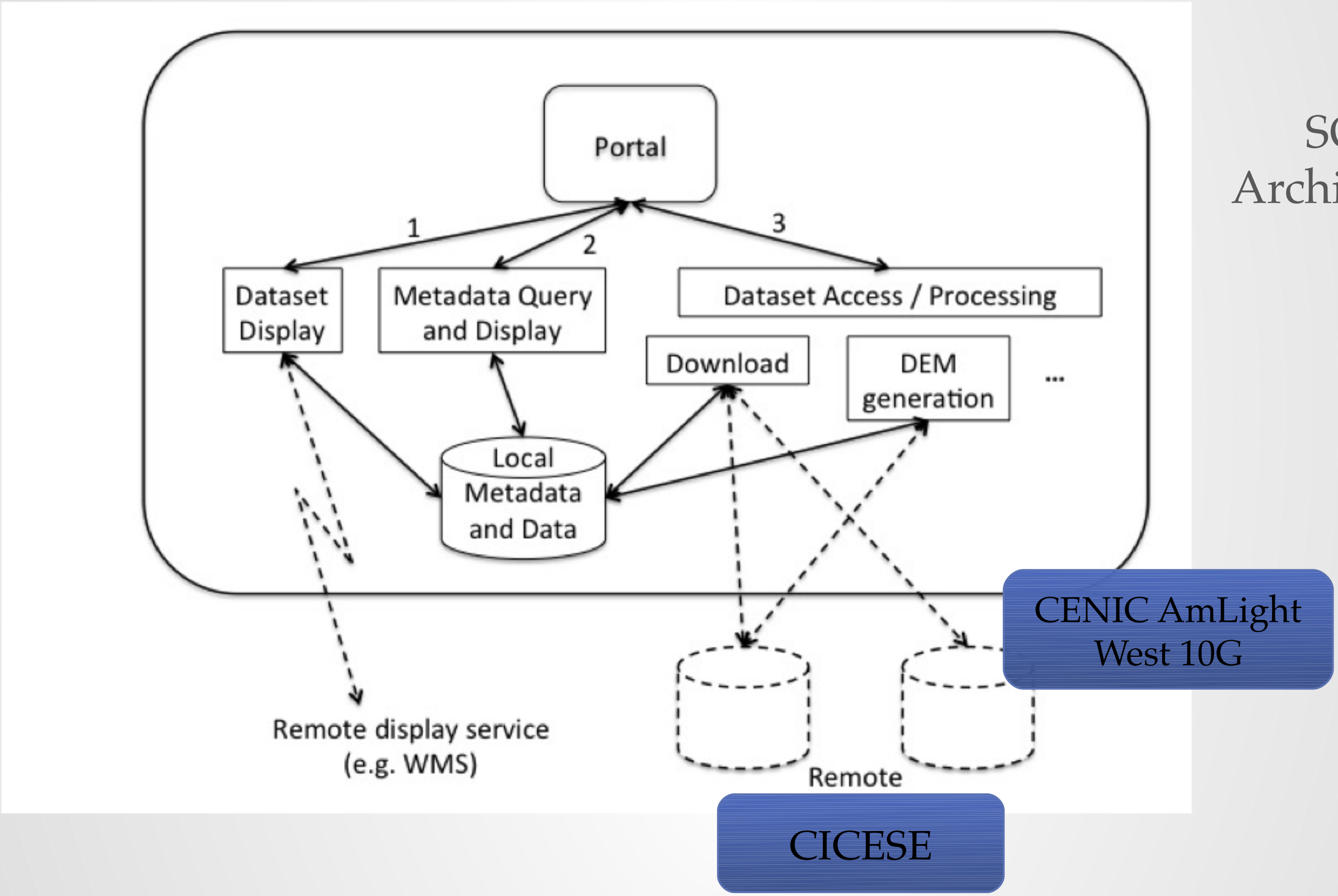
Region	Visits	Visits	Contribution to total: Visits
	1,126 % of Total: 0.94% (119,403)	1,126 % of Total: 0.94% (119,403)	
1. Federal District	344	30.55%	
2. Baja California	313	27.80%	
3. Jalisco	62	5.51%	
4. Mexico	42	3.73%	
5. Aguascalientes	37	3.29%	
6. Guanajuato	34	3.02%	
7. Veracruz	31	2.75%	
8. Michoacan	24	2.13%	
9. Morelos	23	2.04%	
10. Quintana Roo	19	1.69%	



OpenTopography Architecture



Distributed Data Repositories



SOA Architecture

Summary

- OT = excellent example of a mature Spatial Data Infrastructure system enabling access to challenging data for research, education and outreach.
- High value high demand datasets
- Growing partnerships with data providers in the academic, governmental and commercial domains
- High speed 10G link opens new opportunity for collaboration and partnerships with agencies and researchers in Mexico.

Thank You!

www.opentopography.org

Questions/Contact

info@opentopography.org

viswanat@sdsc.edu

baru@sdsc.edu

@OpenTopography

facebook.com/OpenTopography

The Observatorio Astronómico Nacional – San Pedro Mártir Network Bandwidth

Michael Richer
Observatorio Astronómico Nacional-San Pedro Mártir
Instituto de Astronomía - Ensenada
Universidad Nacional Autónoma de México

BIG DATA BIG NETWORK
Big Opportunities for Collaboration between Mexico and California
10-11 Oct 2013 • Ensenada, Mexico

What/where is the OAN-SPM

- The observatory is basically a small village:
  - 4.5 hours from Ensenada
  - 20-25 people
  - food, water brought in
  - generates its own power
  - microwave internet link

History


circa 1992
photo: Edgar Sáenz

- In the early 1990s, a satellite link was installed.
- The bandwidth was 64 kbits/second.
- Telephone service was half duplex.
- Later, the bandwidth was doubled with full duplex telephone service.

Today

unknown photographer, provided by Urania Ceseña

- In early 2003, a microwave link was installed by TelNor.
- The bandwidth is 2 Mbits/second.
- VOIP telephone service was installed.

Today

- Light use:
  - telephone service
  - general use
- Intensive use:
  - data transmission
- Only the RATIR instrument produces a large data volume.
- RATIR is a collaboration between UNAM, UC, ASU, and NASA/Goddard.

The condor's view…
unknown photographer,
provided by Urania Ceseña

Today, RATIR

- The RATIR instrument:
  - 4 channel imaging (2 in visible; 2 in the infrared)
  - robotic telescope (1.5m telescope)
  - data volume: 22 Gbytes/night
  - the data are sent to Ensenada:
    - **impossible** without compression
    - with compression, 100% of the bandwidth for 10 hours/day
- With the current microwave link, we cannot confront a substantially larger data volume.
- An upgrade to 10 Mbits/second is planned within the next few months.

Future (starting now)

- TAOS II (under construction; ASIAA/Taiwan, Harvard/CfA)
  - 3 robotic telescopes, advanced cameras (20 Hz frame readout)
  - 4 Tbytes/night – a fibre optic link is needed
- Boötes-5 (2014; Instituto de Astrofísica de Andalucía)
  - robotic telescope with wide-field camera
  - 35 Gbytes/night – feasible with upgraded microwave link and compression
- Flash (2015; French partners)
  - robotic telescope, similar instrumentation as RATIR
  - 25 Gbytes/night – feasible with upgraded microwave link and compression?
- SPMT (2014+?; INAOE and American partners):
  - 6.5m telescope, advanced instrumentation
  - 30-50 Gbytes/night, first generation instruments
  - probably requires a fibre optic link (SASIR will, > 1Gbit/second)

Future

- Cherenkov Telescope Array (2015+?; Mexican, European, and American partners)
  - array of 25 Cherenkov telescopes
  - 10 Tbytes/observation (¡!), only reduced data will be transmitted,
  - requires a fibre optic link.
- With CFE, we are planning to build an electricity line to connect the OAN-SPM to the national electricity grid.
- This connection will also include a fibre optic link to Ensenada.
- For now, the bandwidth specification is 100 Mbits/second, but can be upgraded with signal repeaters (TAOS-II will need 410 Mbits/second).

Ensenada's Connectivity

- CICESE and UNAM share a 10 Gbits/second link from Ensenada to Tijuana/San Diego for national/ international traffic.
- This link will be important for the instruments that will come online at the OAN-SPM in the next few years.
- This link is also important now for the computing needs of the theoretical astronomers in Ensenada, whose connectivity to computing resources in central Mexico have been limited.

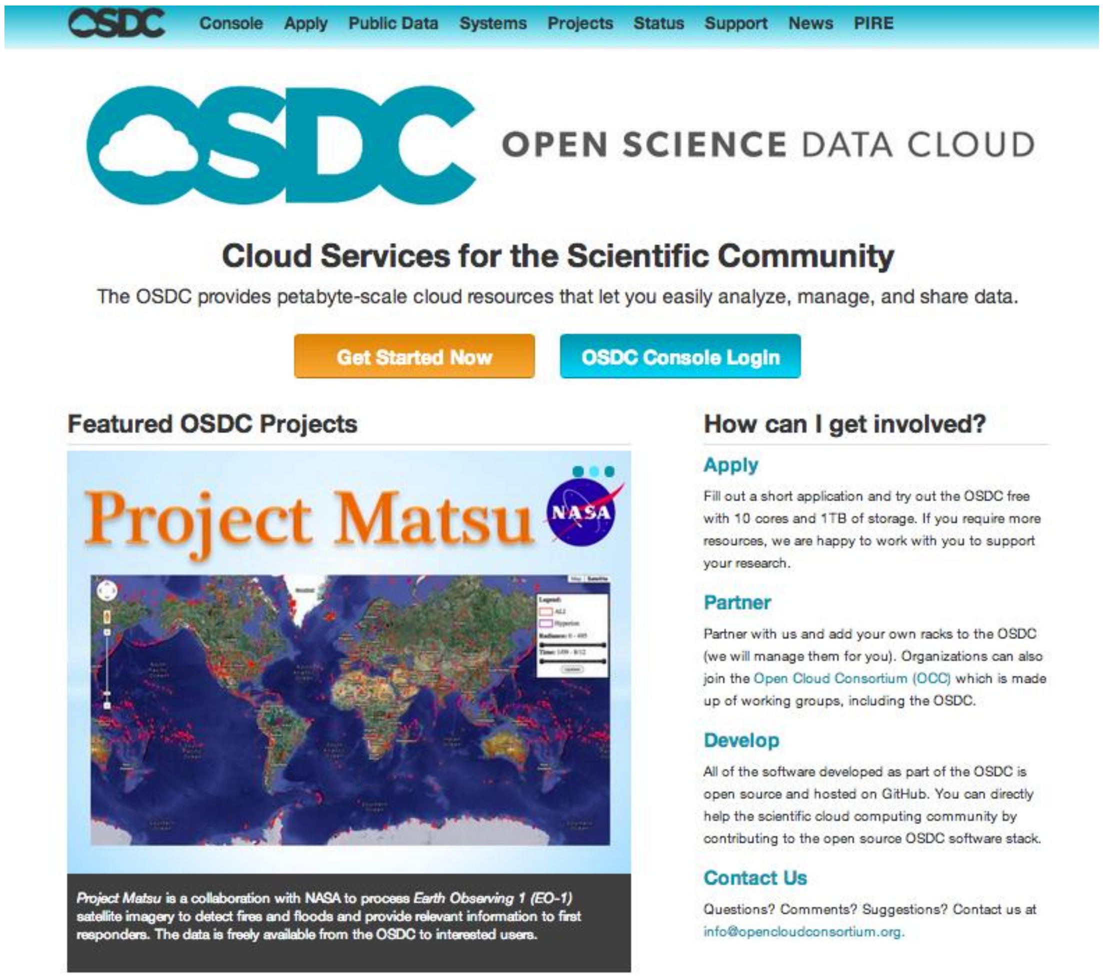
An Introduction to the Open Science Data Cloud

October 10, 2013

Heidi Alvarez
Florida International University

Robert L. Grossman
University of Chicago
Open Cloud Consortium

1. Open Science Data Cloud (OSDC)



Open Science Data Cloud (OSDC)

- OSDC is a Science Cloud Service Provider (CSP)
- Operated by not-for-profit Open Cloud Consortium
- OSDC is a 6 PB / 12,000 core science cloud
- 1 PB science data for the research community
- 1 PB of biomedical data for medical research
- We have been doubling in size each year
- We run production services for NASA and NIH researchers
- Interoperate with Amazon Web Services a.k.a. AWS (still rudimentary)
- Hundreds of users (not thousands)
- Typical job uses 1000s of core hours over 10-100's TB

OSDC OPEN SCIENCE DATA CLOUD

Cloud Services for the Scientific Community

Science Cloud

Biomedical Cloud

Open Science Data Cloud

PROTECTED DATA CLOUD

- Earth sciences
- Biological sciences
- Social sciences
- Digital humanities
- ACL, groups, etc.

Designed to hold Protected Health Information (PHI) e.g. genomic data, electronic medical records, etc. (HIPAA, FISMA)

What You Get with the OSDC

- Login with your university credentials via InCommon
- Launch virtual machines, virtual clusters, access to large Hadoop clusters, etc.
- Access PB+ of open and protected data
- Manage files, collections of files, collections of collections
- Manage users, groups of users
- Manage accounts, sub-accounts
- Efficient transfer of large data (UDT, UDR)

OPEN CLOUD CONSORTIUM

- U.S based not-for-profit corporation.
- Companies: Cisco, Yahoo!, Infoblox, …
- Universities: University of Chicago, Northwestern Univ., Johns Hopkins, Calit2, etc.
- Federal agencies and labs: NASA, LLNL, ORNL
- International university and government partners
- Manages cloud computing infrastructure to support scientific research: Open Science Data Cloud.
- Manages cloud computing testbeds: Open Cloud Testbed.

www.opencloudconsortium.org

Our Point of View

- We want to develop as little technology and software as possible – we want others to develop software and technology.
- We focus on providing researchers the ability to compute over large and very large datasets.
- We need open source solutions.
- Today it is difficult to interoperate with AWS for our protected data cloud, but we expect this to change (someday).
- Run lights out over multiple data centers connected with 10G (soon 100G) networks.

2. Challenges

OSDC Data Centers and Networks

- We have three data centers
  - Chicago with 100G to StarLight
  - FIU with 10G to StarLight
  - Livermore Valley Open Campus 10G to StarLight
- We're planning one more data center with 100G connection to StarLight
- We are looking to interoperate the OSDC with international partners over 10G and 100G networks

Challenges

- We are focusing on the following:
  - How do we authenticate, authorize and provide access controls to researchers at our international partners to data and to cloud based services (storage and compute)
  - We need open source implementations of these services
  - We need trust relationships with our peers
- We are running a series of interoperability workshops to try to get this right.

OSDC OPEN SCIENCE DATA CLOUD

PIRE

PARTNERSHIP FOR **INTERNATIONAL RESEARCH** AND EDUCATION

NSF Award # 1129076

NSF Partnerships for International Research and Education

THE UNIVERSITY OF CHICAGO

UIC

FIU FLORIDA INTERNATIONAL UNIVERSITY

CIARA

OPEN CLOUD CONSORTIUM

PIRE & OSDC

- National Science Foundation Partnership for International Research and Education 5 year program 2010 – 2014 at $3.5M.
- Prepares students to compete in the global cyberinfrastructure community
- Provides **international research and education experiences** around the world!
- The student/faculty/scientist research teams help develop large-scale distributed computing capabilities, data and, State-of-the-art services for integrating, analyzing, sharing and archiving scientific data.

International Collaborators

- Malcolm Atkinson – School of Informatics, Edinburgh University, Scotland, UK
- Paola Grosso & Cees de Laat – Faculty of Science, Informatics Institute, University of Amsterdam, The Netherlands
- Karen Langona and Tereza Cristina Carvalho - LARC – Laboratory of Computer Networks and Architecture at the University of Sao Paulo Brazil
- Satoshi Sekiguchi – National Institute of Advanced Industrial Science and Technology (AIST),Japan
- Chung-I Wu – Beijing Institute of Genomics (BIG), Chinese Academy of Sciences

Research Opportunities

- **What?**
  - Funded internships for US citizens and residents, which provide the chance to participate in sophisticated international research collaborations.
- **When?**
  - Summer of 2014
- **How long?**
  - 6 weeks
- **Where?**
  - At any of our international partners.

OSDC Community of Scholars

/OPEN-SCIENCE-DATA-CLOUD-PIRE-OSDC-PIRE

/OPEN CLOUD CONSORTIUM

- Websites: news.opensciencedatacloud.org , opensciencedatacloud.org
- Mailing list, Summer Workshops

NSF Partnerships for International Research and Education

THE UNIVERSITY OF CHICAGO

UIC

FIU FLORIDA INTERNATIONAL UNIVERSITY

CIARA

OPEN CLOUD CONSORTIUM

Questions?

Thank You!

Big data–aware scheduling with uncertainty in Cloud Computing

Andrei Tchernykh, Jose Lozano Rizk
http://usuario.cicese.mx/~chernykh/

Dzmitry Kliazovich
Johnatan E. Pecero
Pascal Bouvry
University of Luxembourg, Luxembourg

Samee U. Khan
North Dakota State University, U.S.A.

Albert Y. Zomaya
University of Sydney, Australia

Big Data, Big Network Workshop, Ensenada, Mexico. October 10, 2013

My first job

The Elbrus (Russian: Эльбрус) is a line of Soviet and Russian computer systems developed by Lebedev Institute of Precision Mechanics and Computer Engineering, Moscow.

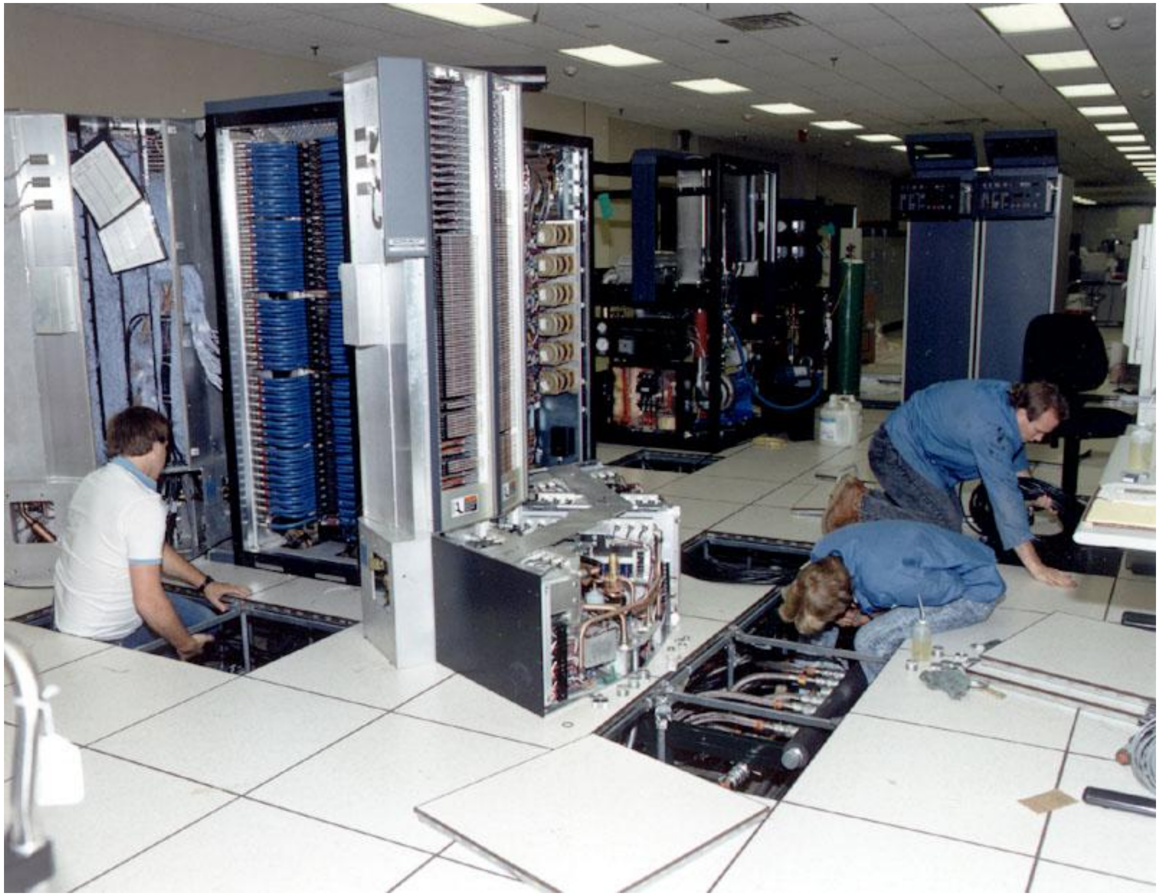
1975

Data flow

Models of Parallel Computation

Processing of Incolmplete Information

Scheduling



My Research Areas

CICESE

HPC

Resource optimization

Multiobjective Optimization

Computational Intelligence

Knowledge Free Scheduling with Uncertainty

Scheduling

online

offline

List Scheduling

Stealing

Scheduling with System Level Agreement

Approximation Algorithms

Workflow Orchestration

Cloud Computing

Real Time Systems

Grid Computing

Collaboration

Germany

Mexico

Universidad Autónoma de Baja California
Universidad Autónoma de Nuevo León
Tecnológico de Monterrey
Instituto Tecnológico de Morelia
Centro de Estudios Superiores del Estado de Sonora

Dortmund University
Prof. Uwe Schwiegelshohn

University of Göttingen
Prof. Ramin Yahyapour

CICESE

USA

Luxembourg

Ohio University, USA
Prof . Klaus Ecker
University of California Irvine, CA, USA
Prof. Isaac Scherson,
Prof. Jean Luc Gaudiot

University of Luxembourg
Prof. Pascal Bouvry
Dr. Johnatan E. Pecero
Dr. Dzmitry Kliazovich

Russia

France

Institute for System Programming, Russian Academy of Sciences
Dr. Nikolay Kuzurin

Institute of Informatics and Applied Mathematics of Grenoble
Prof. Denis Trystram

Institut für Informatik

TUC Technik - Umwelt Chancen

UCIrvine
University of California, Irvine

CEI Computer Engineering Institute

Computer Science Stony Brook University

Laboratoire Franco-Mexicain d'Informatique
LAFMI
Laboratorio Franco-Mexicano de Informática

UC MEXUS
Universitywide Headquarters

INP Grenoble

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

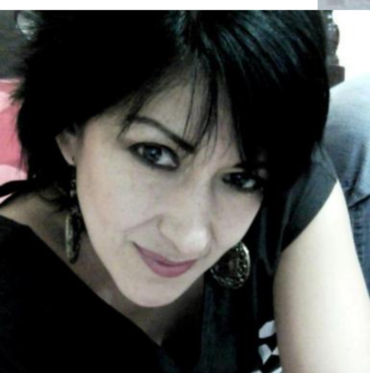
CONACYT

National Science Foundation
WHERE DISCOVERIES BEGIN

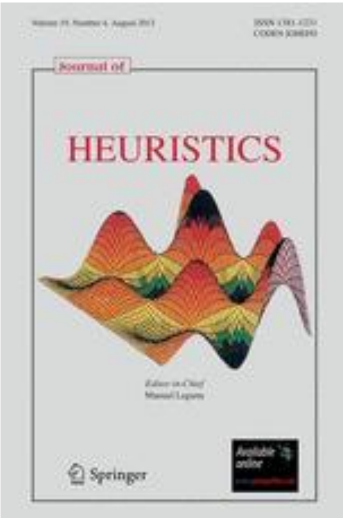
DAAD
Deutscher Akademischer Austausch Dienst
German Academic Exchange Service

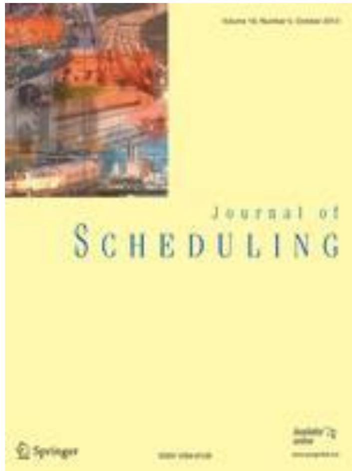
Team

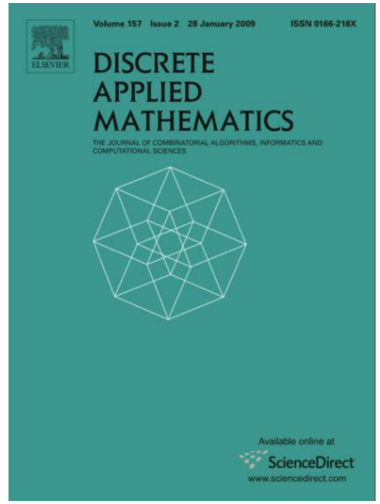

ISUM
rd INTERNATIONAL SUPERCOMPUTING
CONFERENCE IN MEXICO

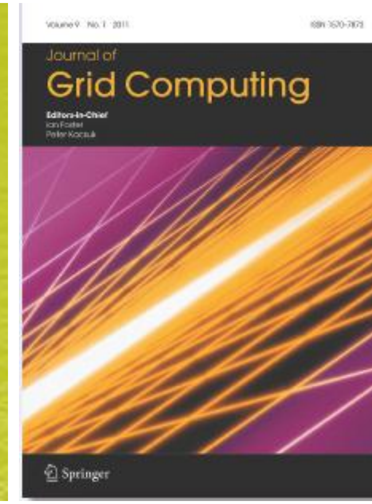


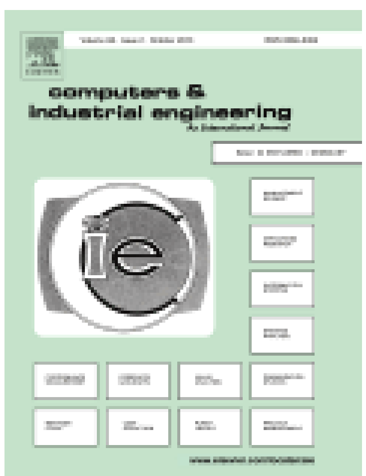
Publications





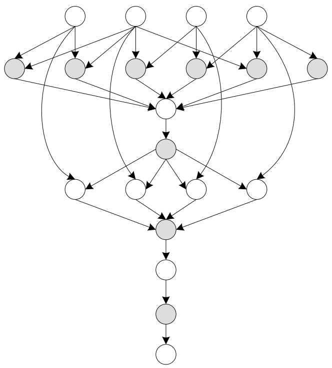




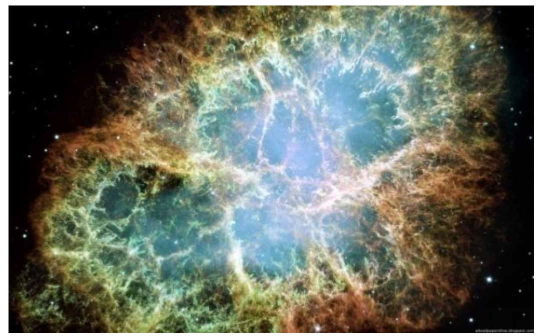


Modeling Applications

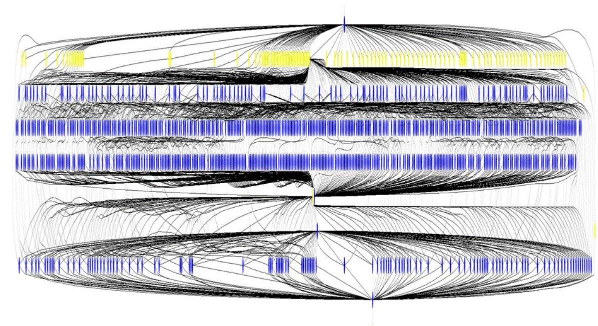
Scientific workflows



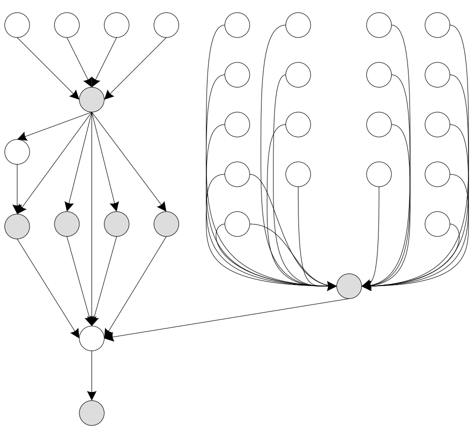
Montage



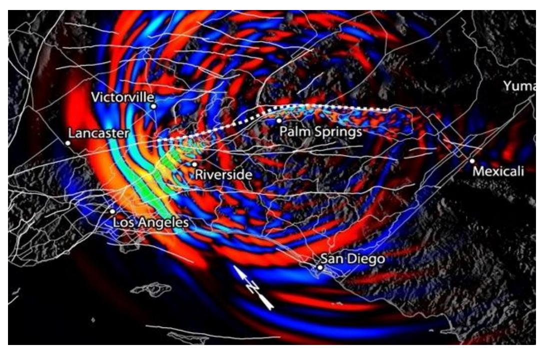
Space mosaics



Montage



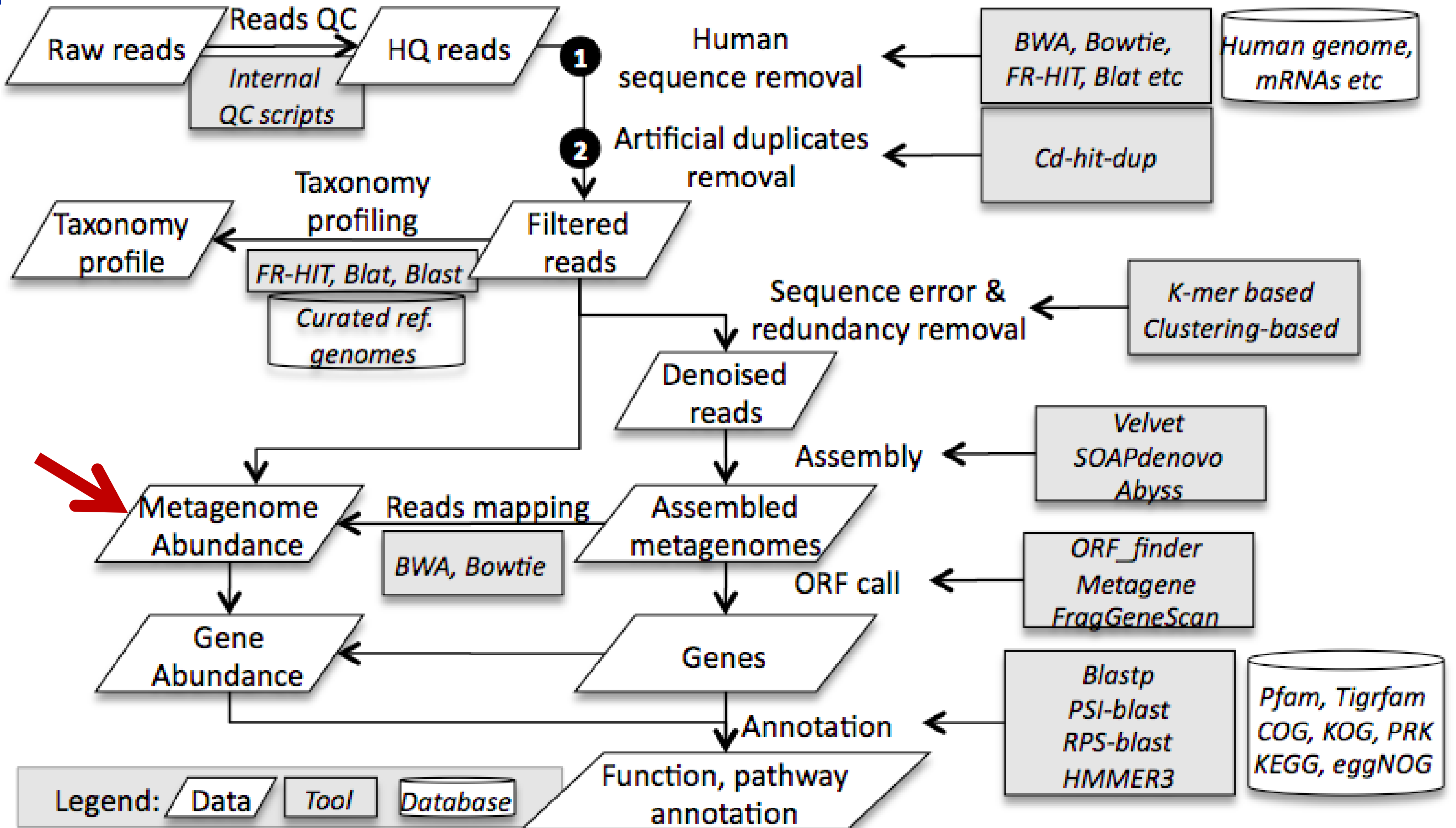
CyberShake



Earthquake

Other aplications:
-Epigenomics,
-Genome,
-LIGO
-SIPHT

Computational NextGen Sequencing Pipeline: From "Big Equations" to "Big Data" Computing



National Human Genome Research Institute

PI: (Weizhong Li, CRBS, UCSD):
NIH R01HG00

(by L. Smarr)

Modeling Applications

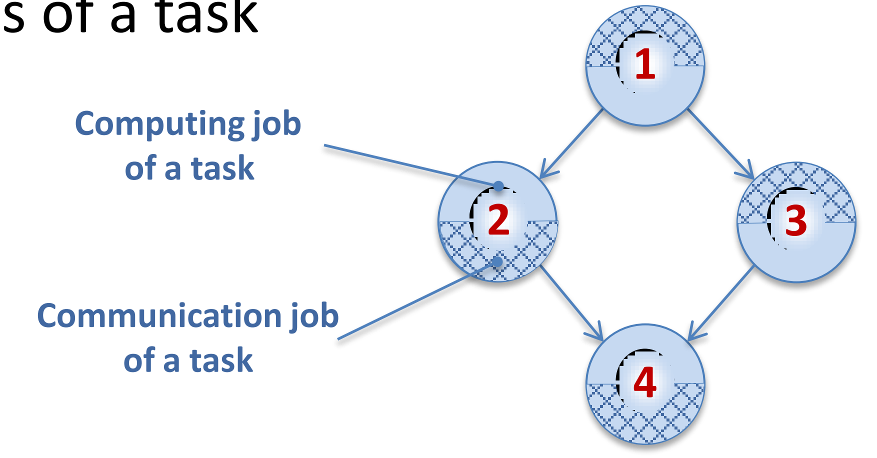
How to model communication processes?

Two known approaches!

– Communication-unaware model
– Edges-based model

Modeling Applications

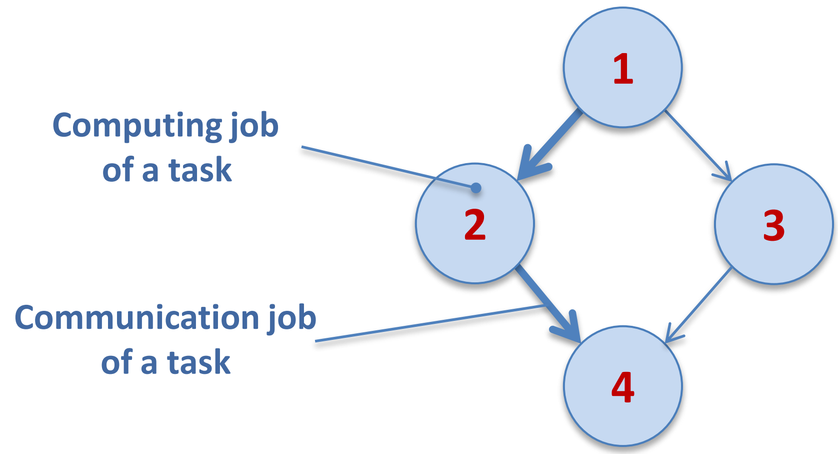
- Communication-unaware model
  - Each vertex represents both computing and communication processes of a task



- Main drawback
  - Having a single vertex for both computing and communications makes it impossible to make separate scheduling decisions

Modeling Applications

- Edge-based model
  - Vertex represents computing
  - Edges represent communication



- Main drawback
  - Two different computing tasks cannot have the same data transfer to receive input as a singe edge cannot lead to two different vertices

Key challenge

Uncertainty

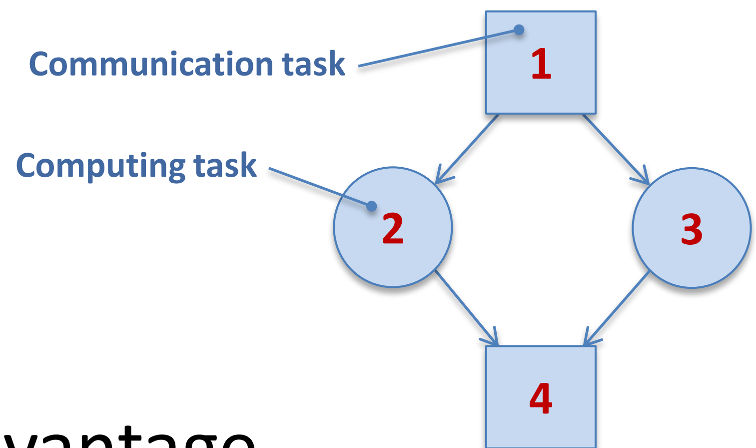
In real dynamic environment

Might not know
- quantity of data that can be managed,
- quantity of computation required by a task.
- exact knowledge about the system.
- effective processor speed,
- number of available processors,
- actual bandwidth,
- etc.

Parameters are changing over the time.

Modeling Applications

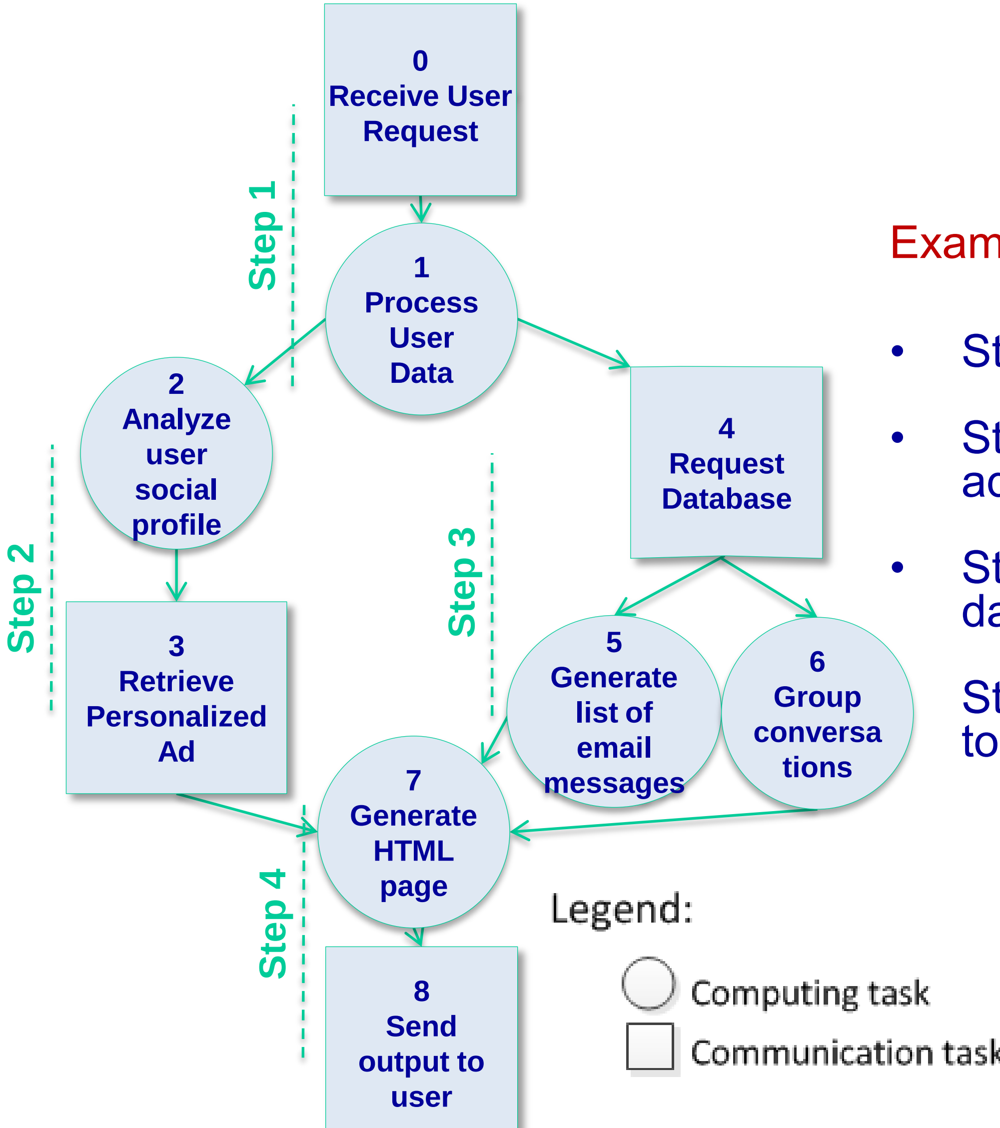
- Proposed CA-DAG: Communication-Aware DAG model
  - Two types of vertices: one for computing and one for communications
  - Edges define dependences between tasks and order of execution



- Main advantage
  - Allows separate resource allocation decisions, assigning processors to handle computing jobs and network resources for information transmissions

CA-DAG: Communication-Aware DAG

by Dzmitry Kliazovich IEEE Cloud’13



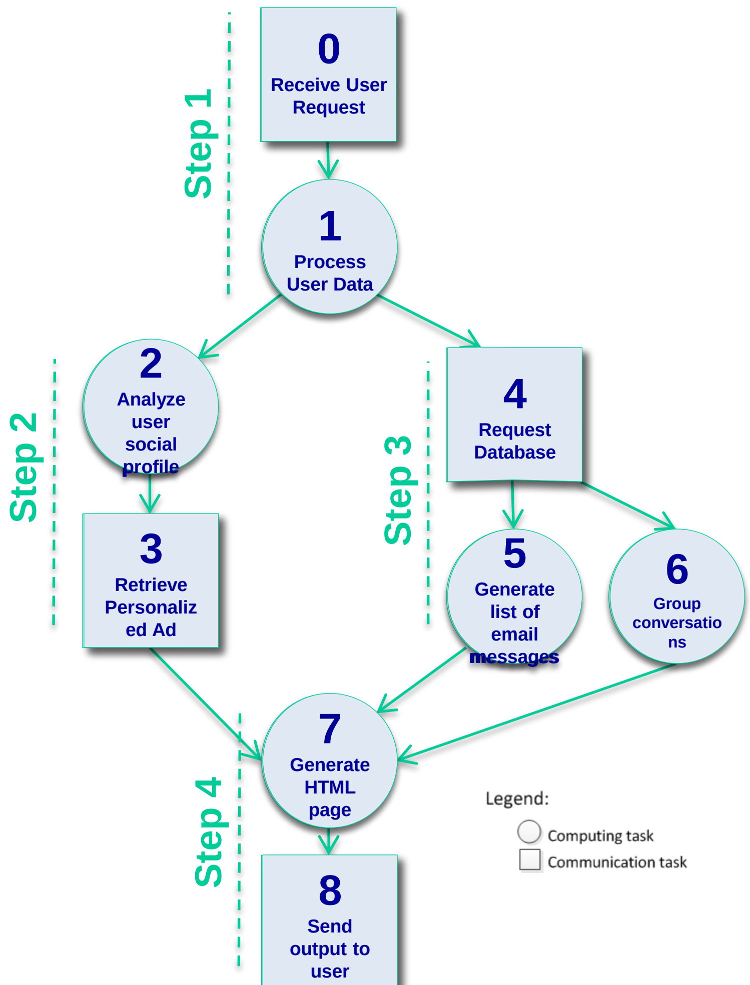
Example of webmail cloud application

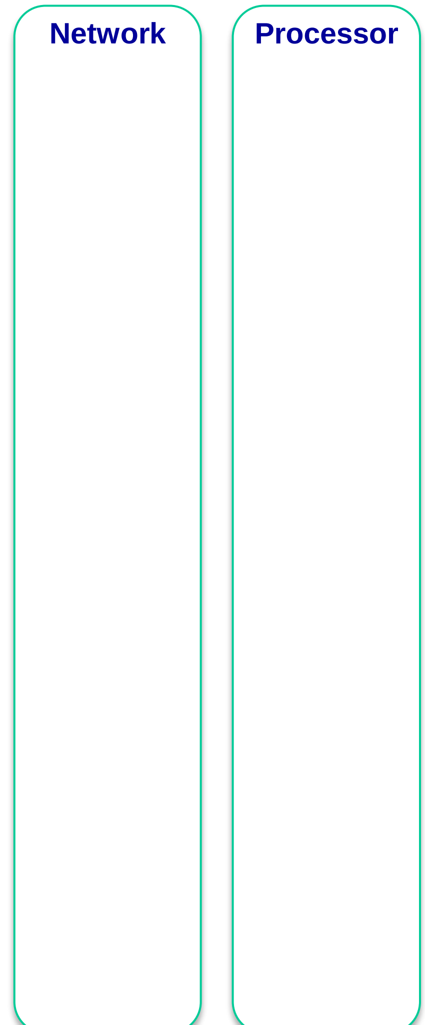
- Step 1: Receive user request and process it
- Step 2: Generate personalized advertisement
- Step 3: Request list of email messages from database

Step 4: Generate HTML pages and send it to the user

CA-DAG: Communication-Aware DAG

by Dzmitry Kliazovich IEEE Cloud’13

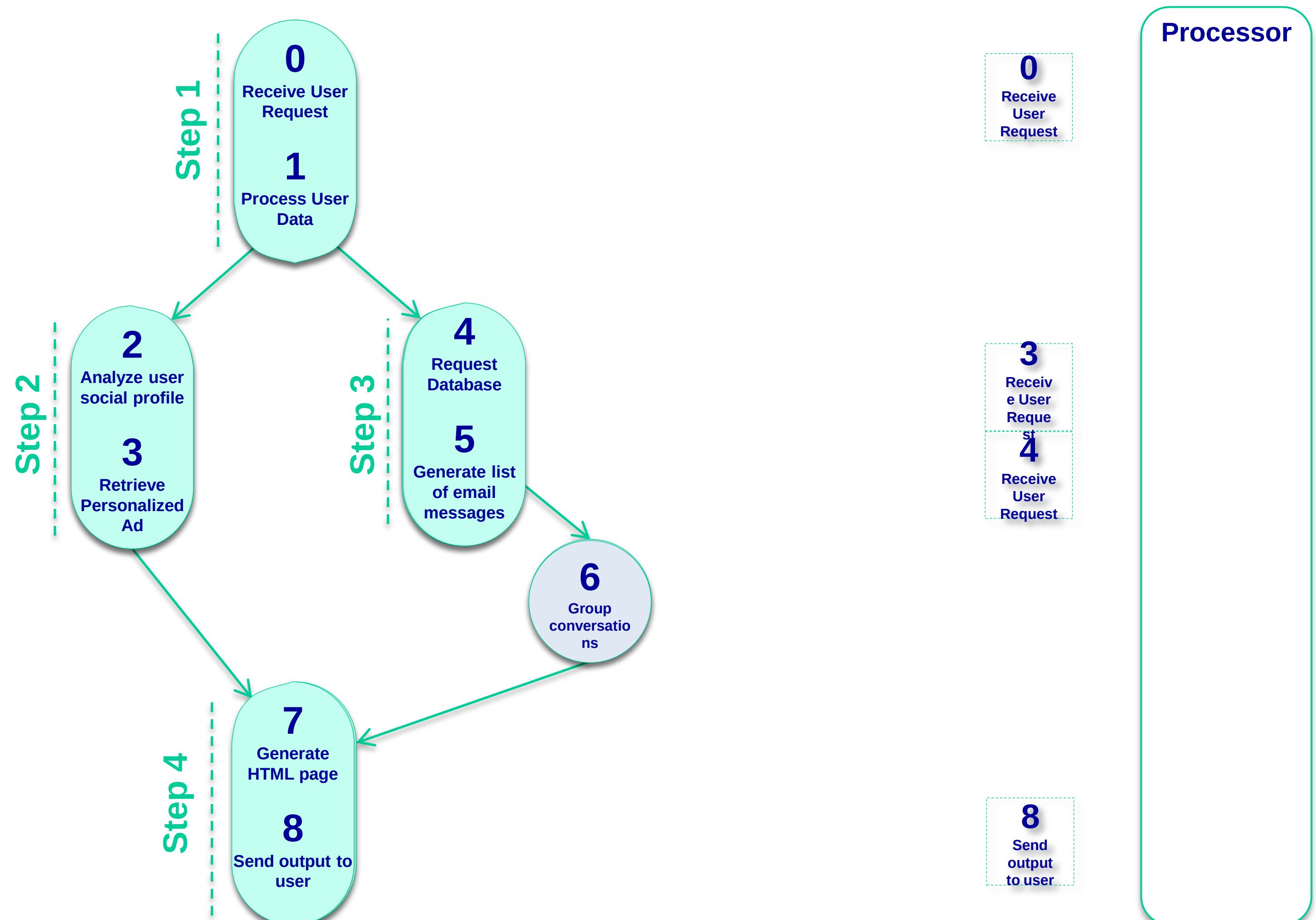




Makespan: 7

Communication-unaware model

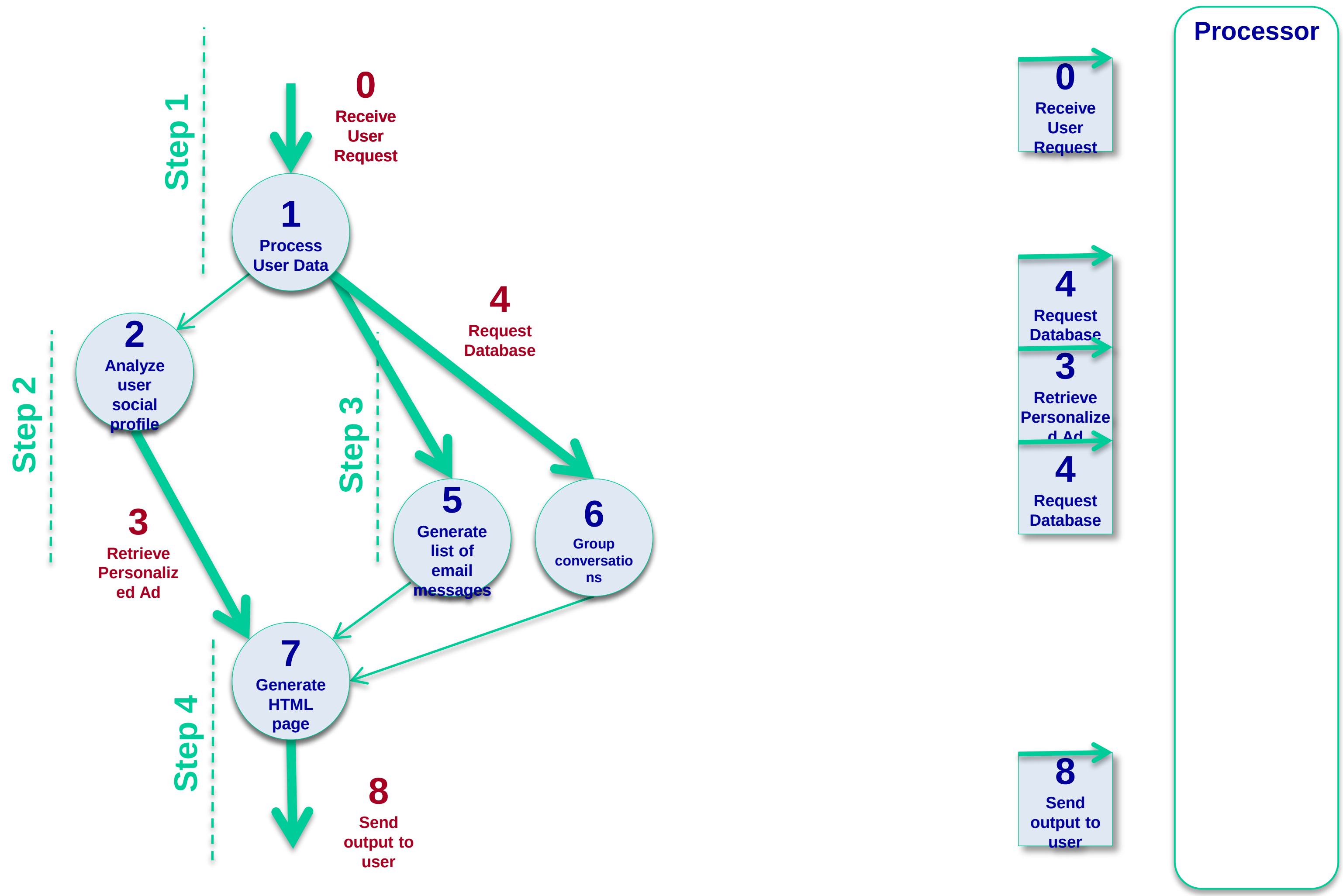
by Dzmitry Kliazovich IEEE Cloud'13



Makespan: 9

Edges-based model

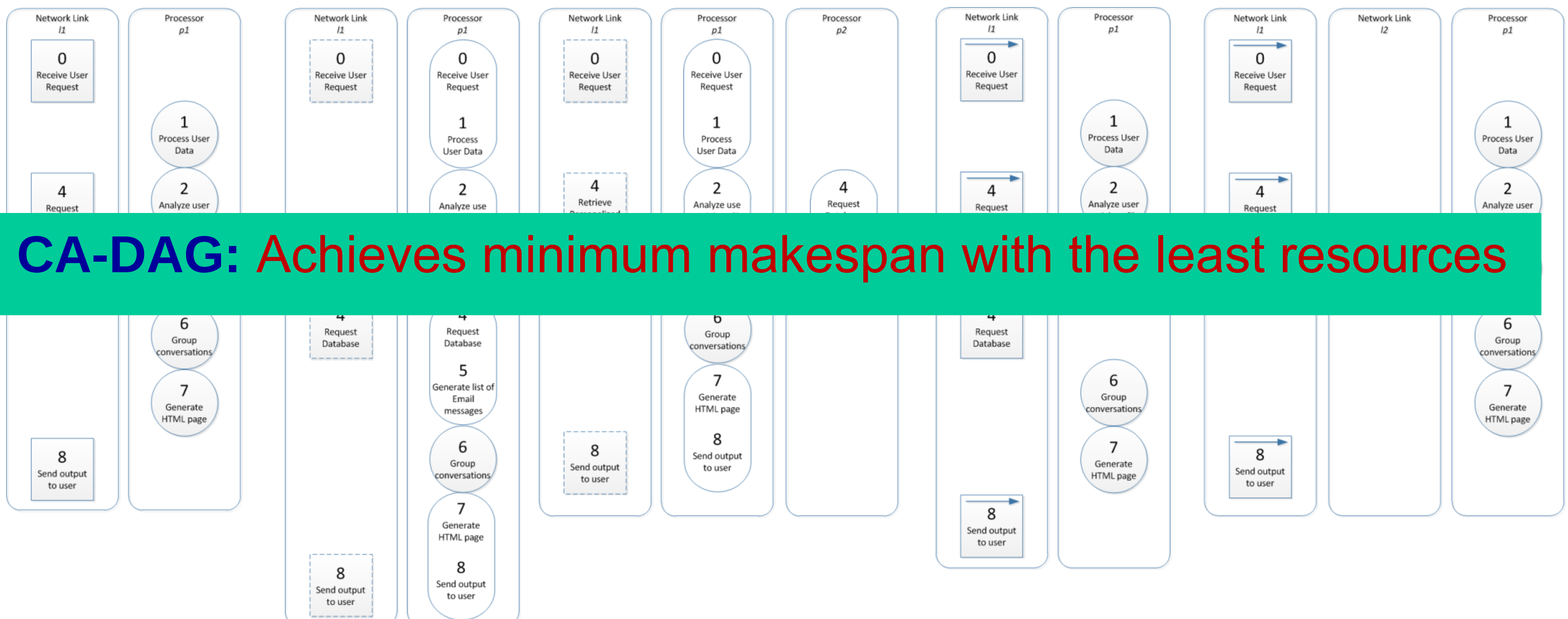
by Dzmitry Kliazovich IEEE Cloud'13



Makespan: 8

Comparison of schedules

CA-DAG model | **Communication-unaware model** | **Edges-based model**

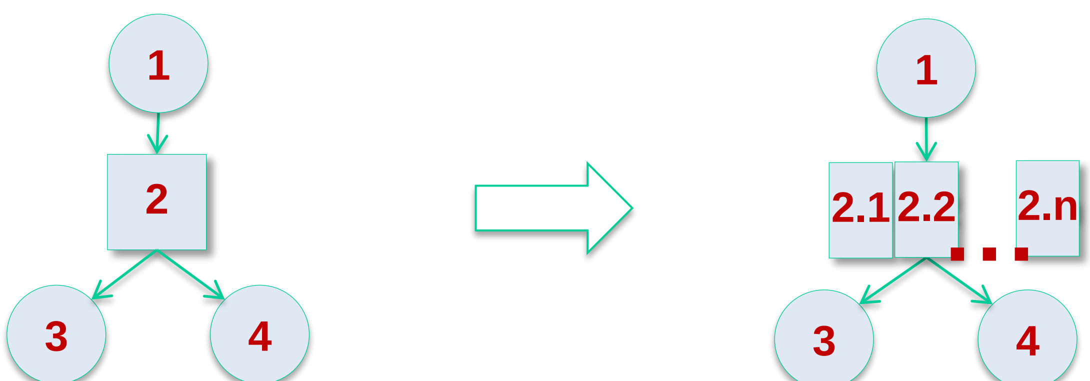


CA-DAG: Achieves minimum makespan with the least resources

# of Processors	# of Network links	Communication-unaware model	Edges-based model	**Proposed CA-DAG model**
1	1	9	8	**7**
1	2	9	7	**7**
2	1	7	8	**7**

Task Parallelization

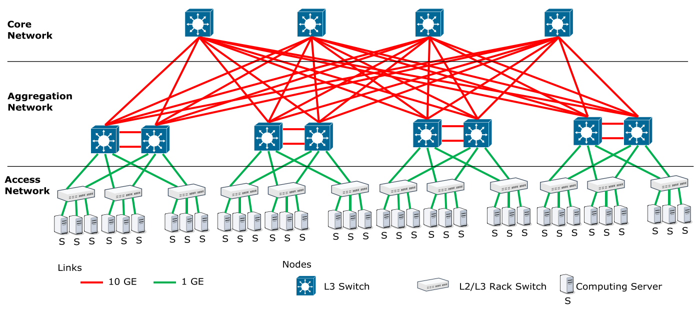
- Each communication task/vertex can be divided into n different independent communication tasks that can be executed in parallel



Multipath Routing

by Dzmitry Kliazovich IEEE Cloud'13

- Most of existing solutions rely on static network topology and fixed pre-allocation which implies circuit switching and pre-defined routing



- **In reality, datacenter networks are packet switched with routing decisions taken at every hop**

- **The availability of multiple paths is essential to benefit from parallelization property of communication tasks**

Thanks for your attention!

- Are there any
  - Questions ?
  - Comments/ Suggestions ?

Dzmitry Kliazovich, Johnatan E. Pecero, Andrei Tchernykh, Pascal Bouvry, Samee U. Khan, and Albert Y. Zomaya. CA-DAG: Modeling Communication-Aware Applications for Scheduling in Cloud Computing Data Centers. IEEE CLOUD 2013 - IEEE 6th International Conference on Cloud Computing. p. 277– 284, June 27-July 2, 2013, Santa Clara Marriott, CA, USA, DOI 10.1109/CLOUD.2013.40

Índice General

- Presentaciones Área: Oceanography

- Presentaciones Área: Nano-materials Science

- Presentaciones Área: Bioinformatics