# Prototyping Science As A Service with **Cloud Data Distribution and Tools**

Daniel Mandl - NASA\GSFC
May 17, 2012

# Team

Daniel Mandl – NASA/GSFC - lead

Robert Sohlberg – Univ. of Maryland/Dept. of Geography – SensorWeb scientist

Pat Cappelaere – Vightel Corporation – SensorWeb architecture

Stuart Frye – NASA/GSFC – SGT Inc. – systems engineer

Vuong Ly – NASA/GSFC – system engineer, software developer

Matthew Handy – NASA/GSFC – software developer

Steve Chien – NASA/JPL – algorithms

Tim Creech – Univ. of Maryland – NASA Science and Technology Research Fellow

Chris Flatley – NASA/GSFC summer intern

Neil Shaw – NASA/GSFC summer intern

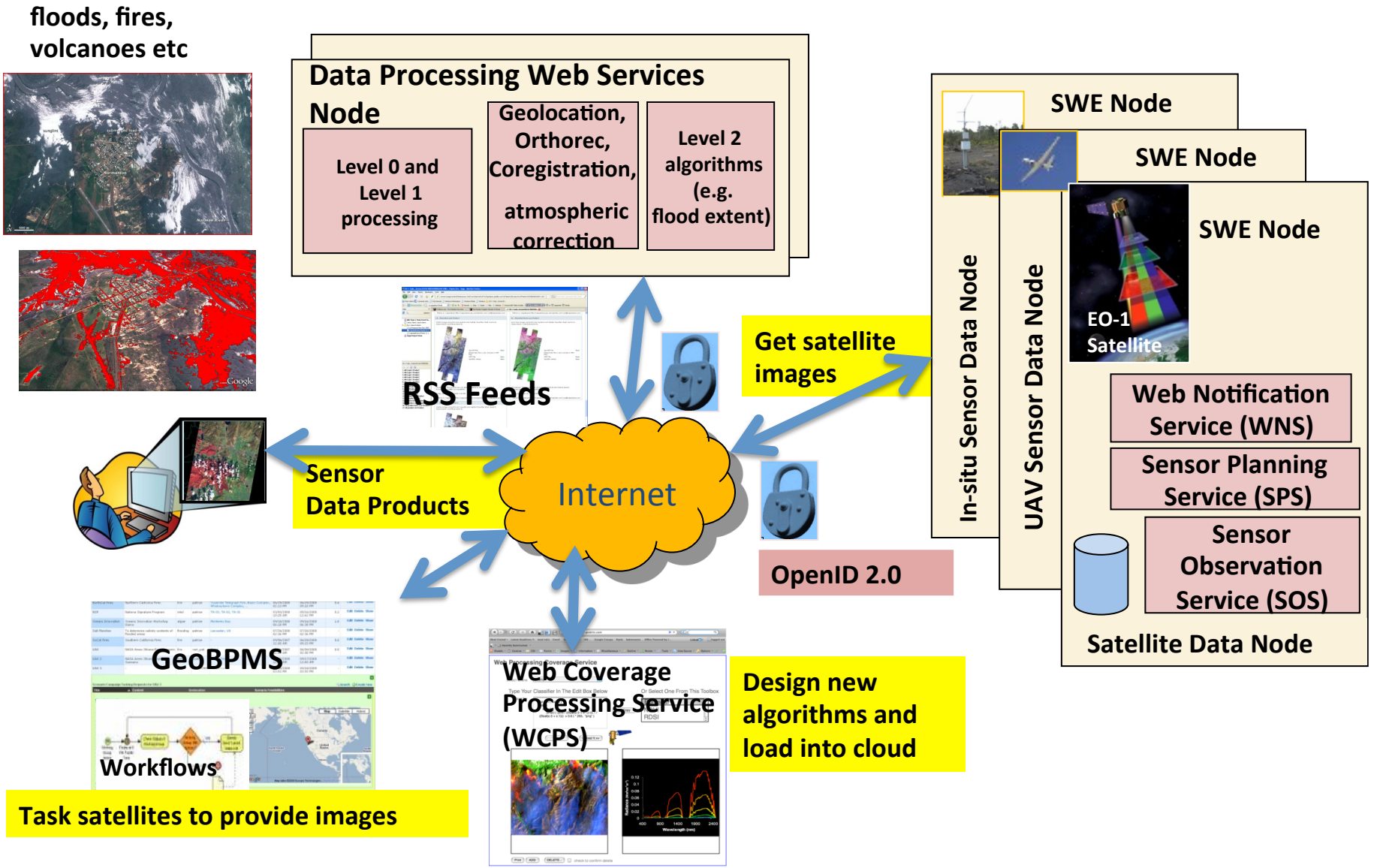
Joshua Bronston – NASA/GSFC co-op

# Objectives

- Experiment with Elastic Compute Cloud and demonstrate Science As A Service by virtualizing data processing pipeline processes
- Experiment and demonstrate automation and increased efficiency for large data set distribution and data product production
- Collaborate with Open Cloud Consortium and their Open Science Data Cloud
  - OCC provides Science as a Service (SAAS) or more accurately the infrastructure, platform and services to support science as a service.
- Using commercial cloud (Joyent) also
- Demonstrate various user run tools on the cloud

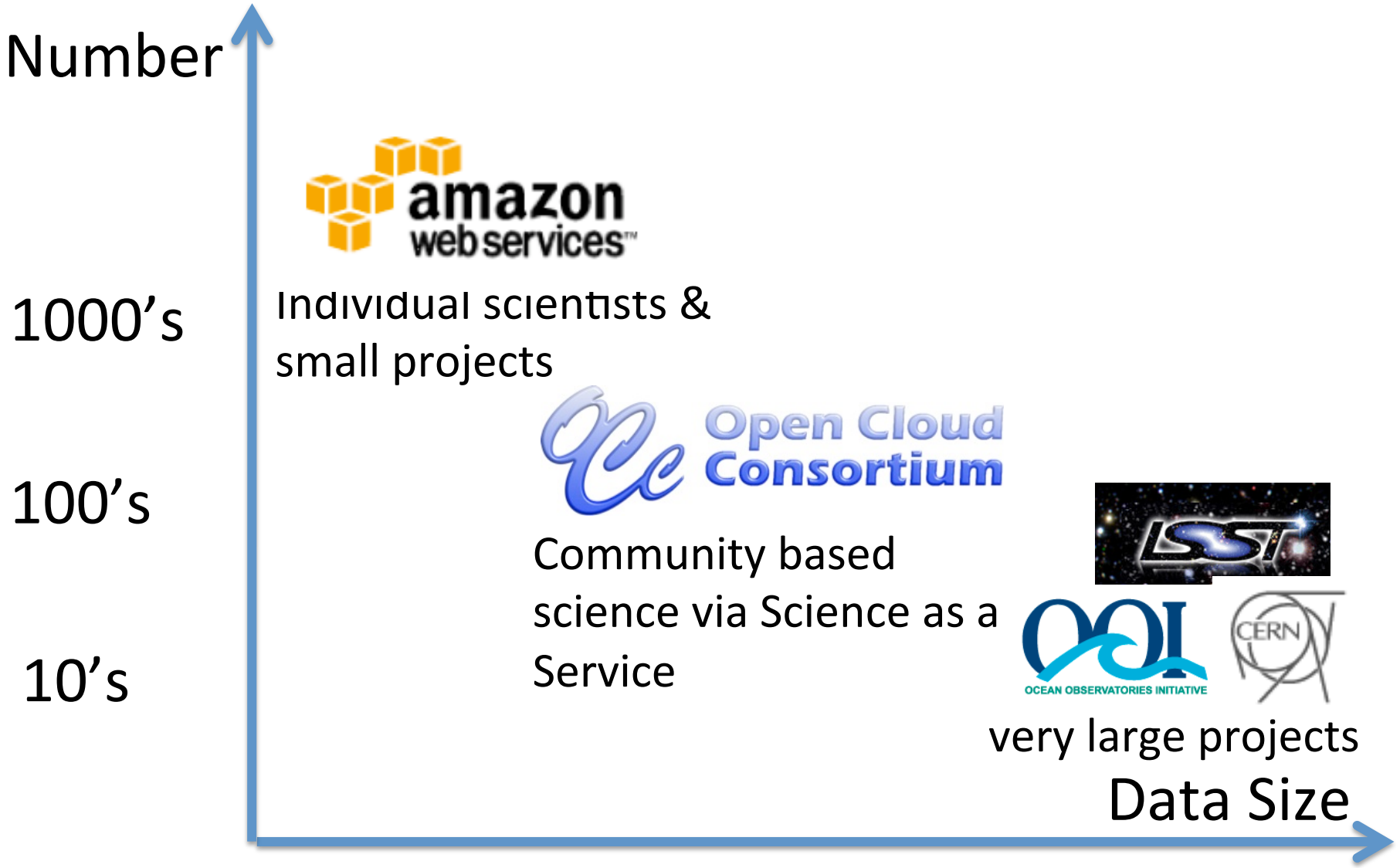
# SensorWeb High Level Architecture

Sensors, Algorithms and Models Wrapped in Web Services Provide Easy Access to Sensor Data and Sensor Data Products



# Integrate SensorWeb with Open Cloud Consortium Components


www.opencloudconsortium.org


Number
1000's
100's
10's
amazon web services
Individual scientists & small projects
Open Cloud Consortium
Community based science via Science as a Service
LSST
OOI
OCEAN OBSERVATORIES INITIATIVE
CERN
very large projects
Data Size
Small
Public infrastructure
Medium to Large
Shared community infrastructure
Very Large
Dedicated infrastructure

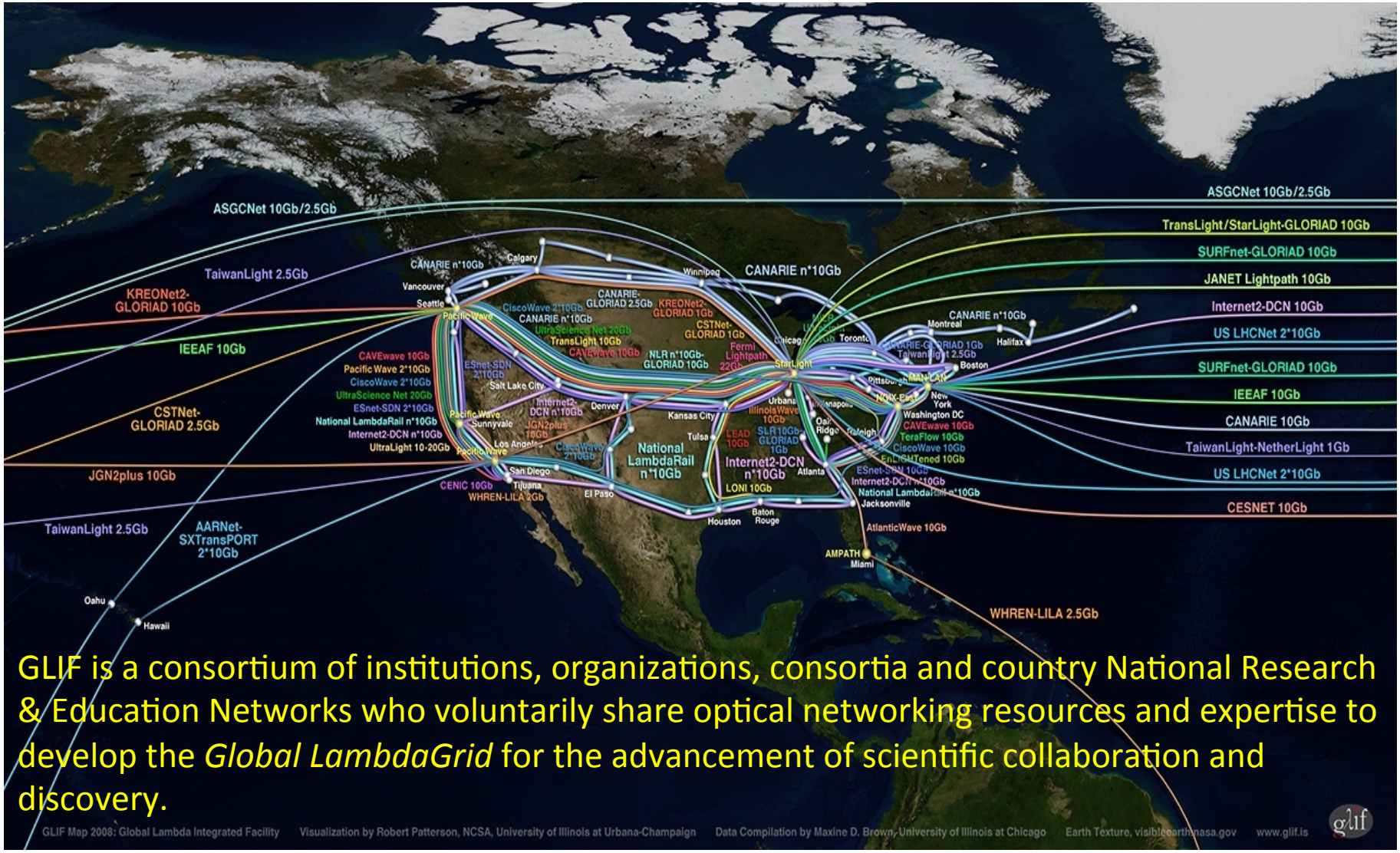

- U.S based not-for-profit corporation.
- Manages cloud computing infrastructure to support scientific, environmental, medical and health care research, such as the Open Science Data Cloud.
- Manages cloud computing testbeds such as the Open Cloud Testbed.
- Develops reference implementations of standards based software for clouds
- Engages in outreach and education to support cloud computing.

www.opencloudconsortium.org

# OCC Members

- Companies: Cisco, Yahoo!, Citrix, …
- Universities:  University of Chicago, Northwestern Univ., Johns Hopkins, Calit2, ORNL, University of Illinois at Chicago, …
- Federal agencies and labs: NASA, LLNL, ORNL
- International Partners: AIST (Japan)
- Partners: National Lambda Rail

# Global Lambda Integrated Facility (GLIF)
# OCC Collaboration with Starlight (part of GLIF)




GLIF is a consortium of institutions, organizations, consortia and country National Research & Education Networks who voluntarily share optical networking resources and expertise to develop the *Global LambdaGrid* for the advancement of scientific collaboration and discovery.

# OSDC Distribution of Scientific Data

- By the end of the 2012, the OSDC will make available approximately 1.5 PB of scientific and data to the research community.
- The data is available through both the commodity internet and through high performance research networks, including NLR and Internet2
- The data is from a variety of scientific disciplines:
  - 100+ TB of earth science through Project Matsu (joint with NASA)
  - 500 TB of genomic data, including the 1000 Genomes data set
  - 100 TB of astronomy data
  - 100 TB of web related data

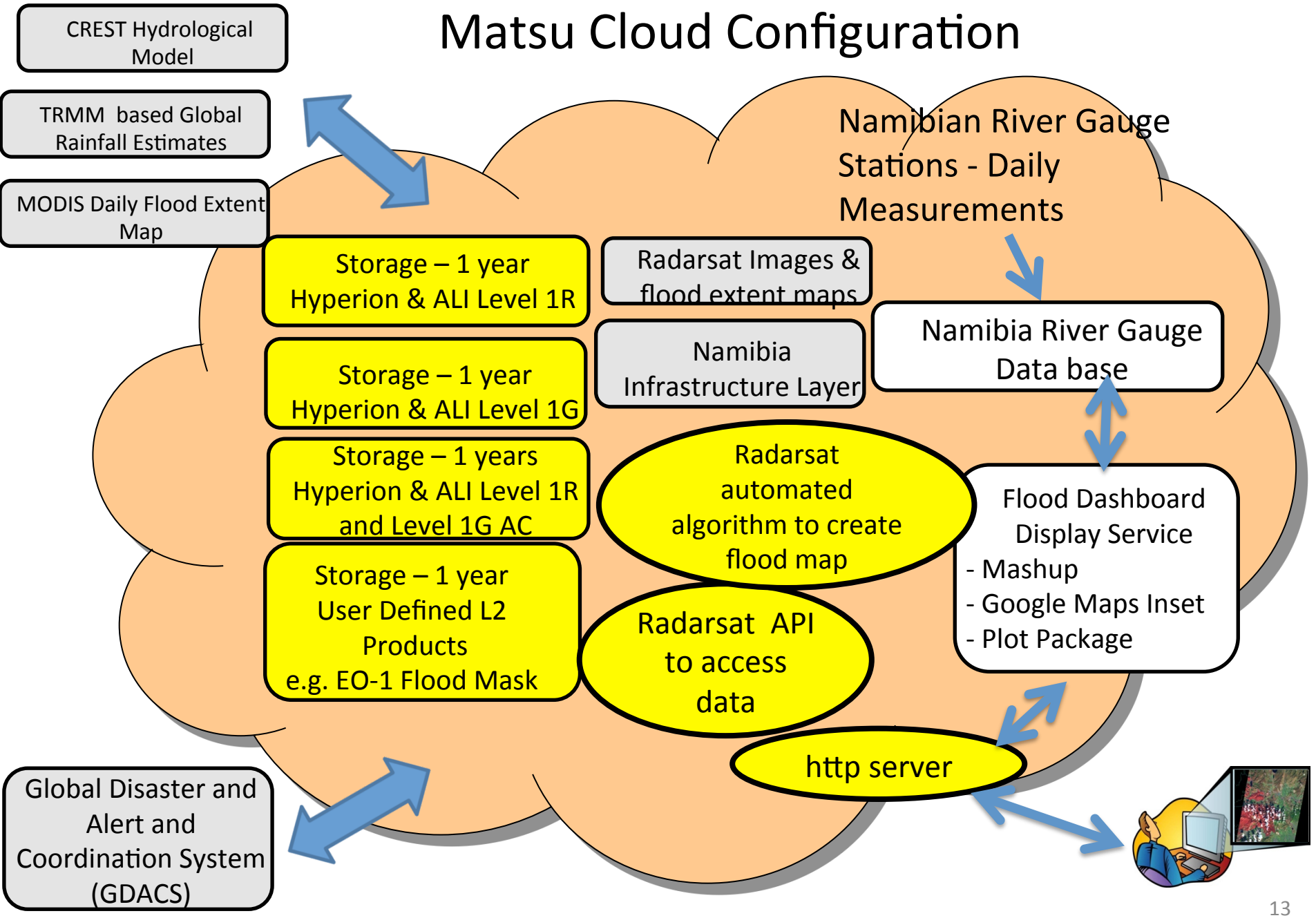
# OCC Investments

| Year | Organization | $M | Comment |
|---|---|---|---|
| 2009 | Cisco | c. $1M+ | Cisco provides access to the Cisco C-Wave to connect OCC data centers and partners with 10G wide area networks |
| 2009 | Yahoo! | c. 1.2M | Yahoo! donated a 2000 core cluster to the OCC |
| 2011 | NSF | $3.5M | NSF grant "Training and Workshops in Data Intensive Computing Using The Open Science Data Cloud" |
| 2011 | Yahoo! | c. $1M | Approximately $1M of equipment for OCC-Y Hadoop Cluster |
| 2011 | Moore Foundation | $2M | $1M per year for 2 years for equipment to support OCC-Adler & OCC-Sullivan |

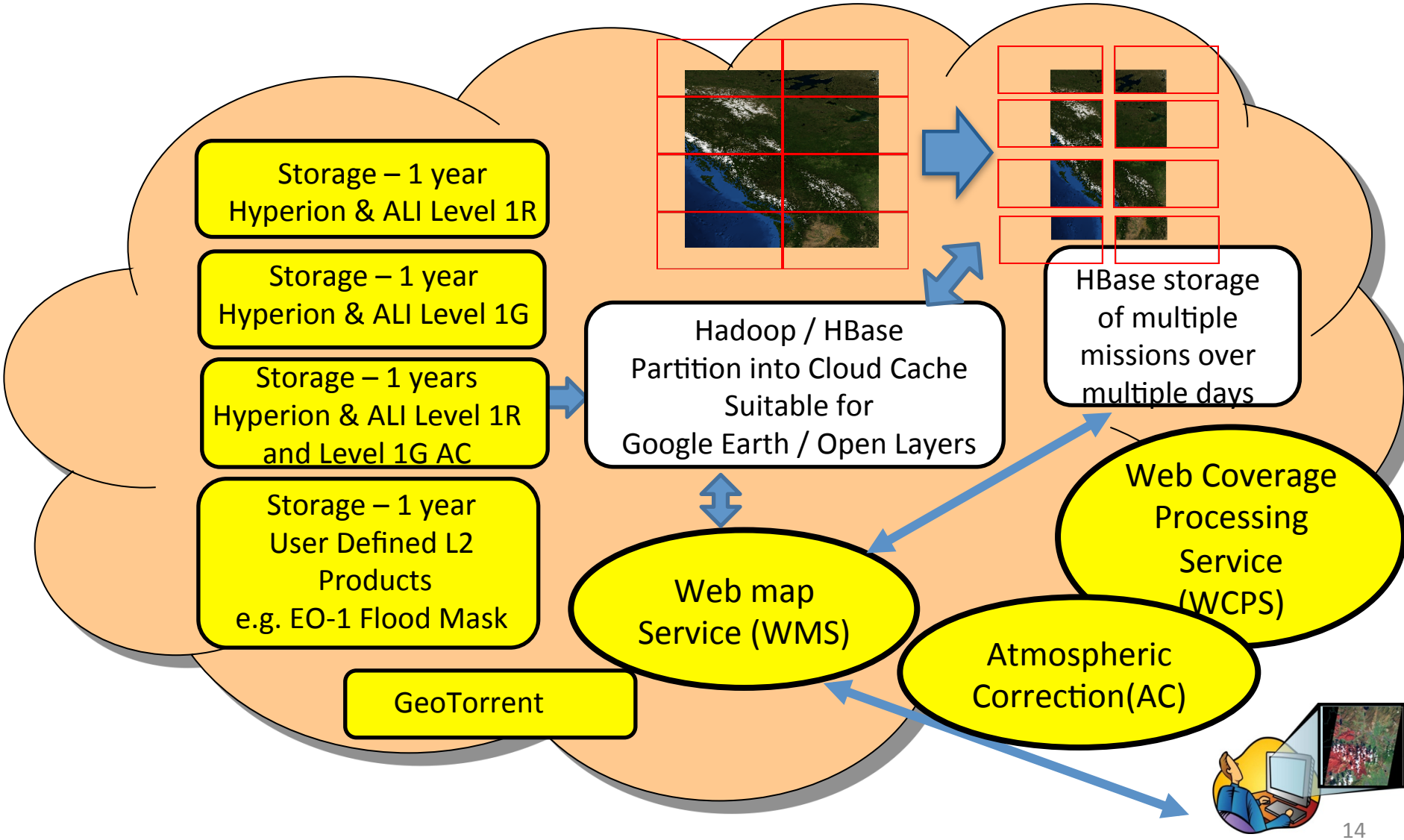
The Open Science Data Cloud (OSDC) is a hosted distributed facility managed by the OCC that:

- Manages & archives medium and large size datasets.
- Provides computational resources to analyze them.
- Provides networking to share the datasets with your colleagues and with the public.

# Matsu Cloud Configuration

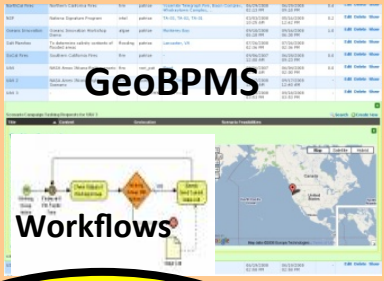


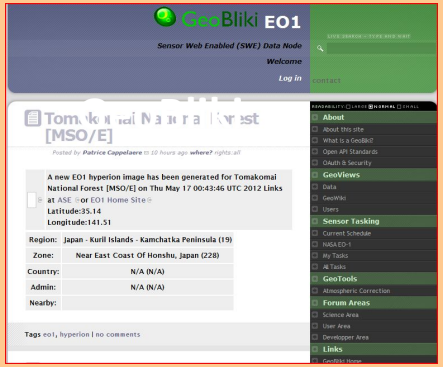
# Matsu Cloud (In process)
# Hadoop and Tiling Handles Large Dataset Displays



# Joyent Cloud (commercial)

GeoTorrent




Radarsat Tasking API extension to GeoBPMS

# Namibia Flood Dashboard

## Namibia Flood Dashboard

**SensorWeb enabled for early flood warning**

Daily Report

April 12

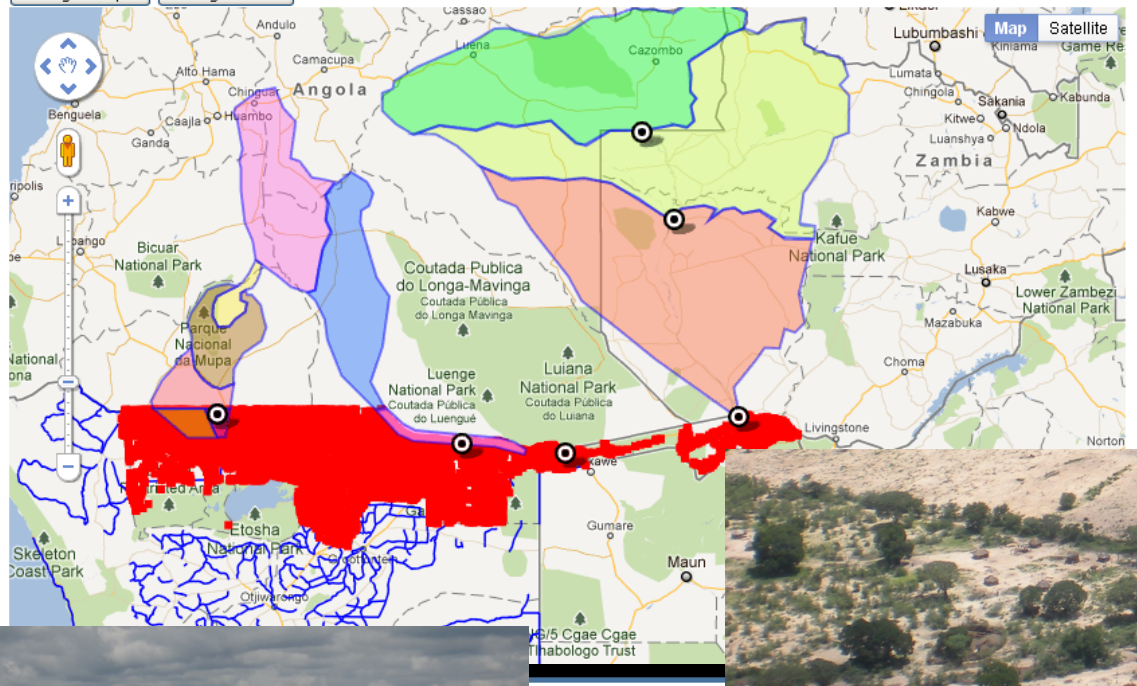
Daily Bulletin:

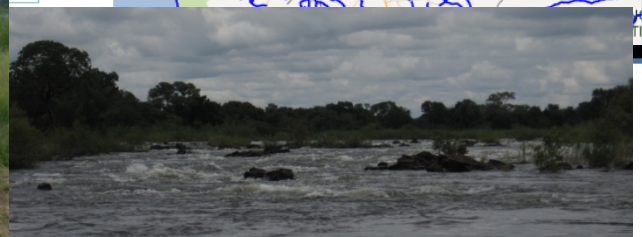
HYDROLOGICAL SERVICES NAMIBIA – DAILY FLOOD BULLETIN 03 APRIL 2012

View Complete Current Bulletin
View Bulletin Records
Search Bulletin Records
New Bulletin

- River Stations
- SensorWeb Layers
- Water Lines and Areas
- Satelite Overlays
- Ground Pics
- Kavango Radarsat Data
- Cuvelai Radarsat Data
- TRMM Rainfall Accumulation and Flood Forecast
- Global Scene Counts

Google Maps | Google Earth



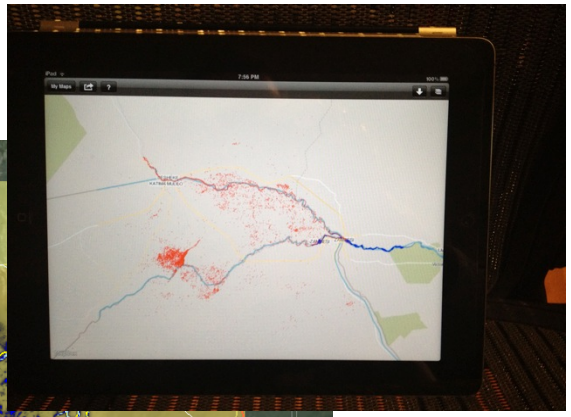



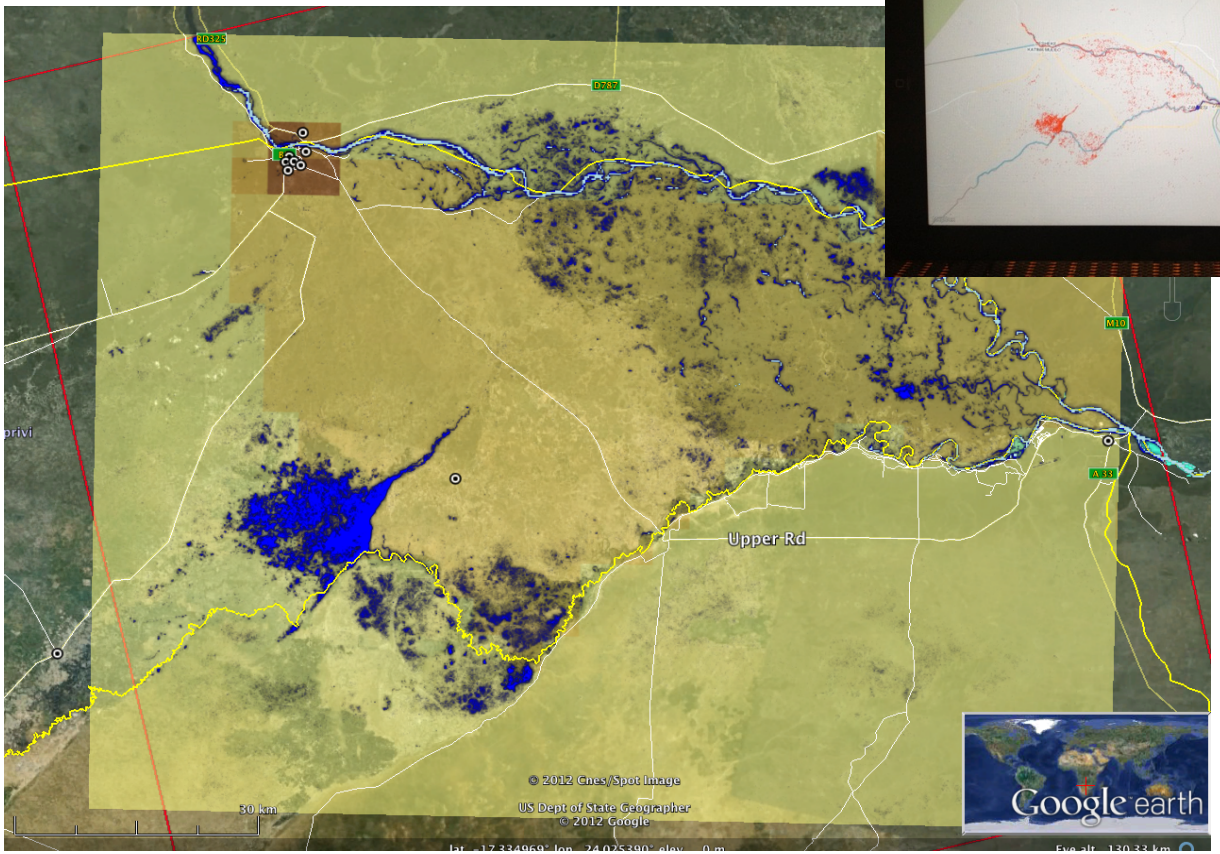
# Matsu Cloud Functionality Enhancements

- Began developing method for automated co-registration using WCPS in Matsu Cloud
- Radarsat tasking Application Processing Interface (API) integrated with GeoBPMS (due in summer or fall 2012)
  - Funded by AIST ESTO QRS 2011
- Radarsat processing, and data distribution API
  - Display Radarsat data on IPAD
  - Funded by AIST ESTO QRS
- Waterpedia for Architecture Implementation Pilot 2 (AIP-5) (due summer/fall 2012)
  - High resolution water mask on map
  - Open street format
  - Crowd sourcing to validate and calibrate
  - Funded by AIST QRS 2011
- EO-1 tasking API for Pacific Data Center (due summer/fall 2012)
  - Funded by AIST ESTO QRS 2011

# Matsu Cloud Functionality Enhancements

- Data distribution API for SERVIR nodes
  - CATHALAC (SERVIR in Caribbean)
  - RCMRD (SERVIR in Africa)
  - ICIMOD (SERVIR in Asia)






Radarsat data processed in cloud and tiled displayed on Ipad for field work and in preparation for Waterpedia with crowdsourcing

# Conclusion: Relevance to HyspIRI

- HyspIRI will have large data sets
- Experiment with managing, processing and distributing large data sets in cloud on a “do-it-yourself” basis
- Prototype Science As A Service for users of the HyspIRI large data sets

# Backup

# StarLight – "By Researchers For Researchers"

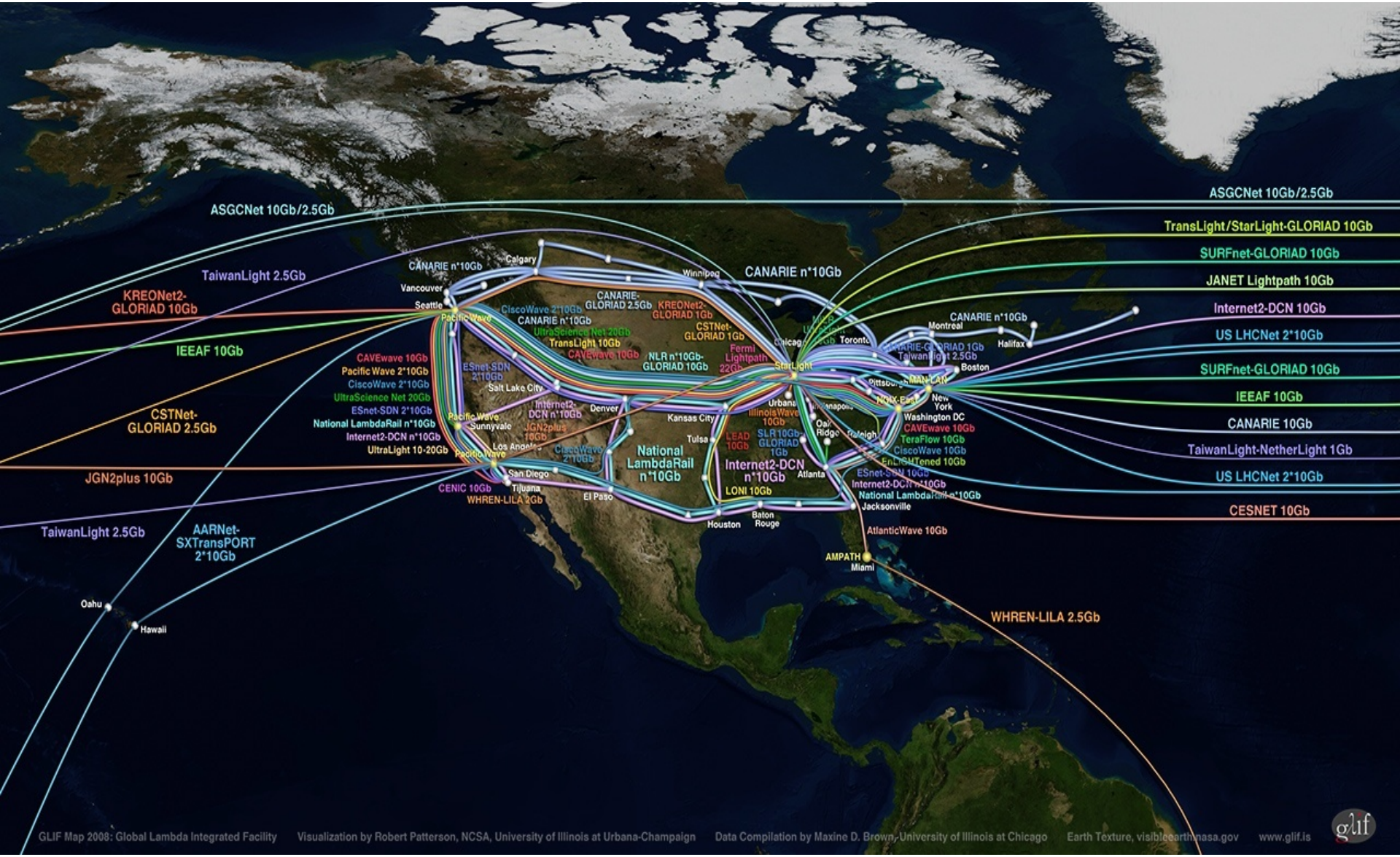
StarLight is an experimental optical infrastructure and
proving ground for network services optimized for high-performance applications
GE+2.5+10GE
Exchange
Soon:
Multiple 10GEs
Over Optics –
World's "Largest"
10GE Exchange
First of a Kind
Enabling Interoperability
At L1, L2, L3

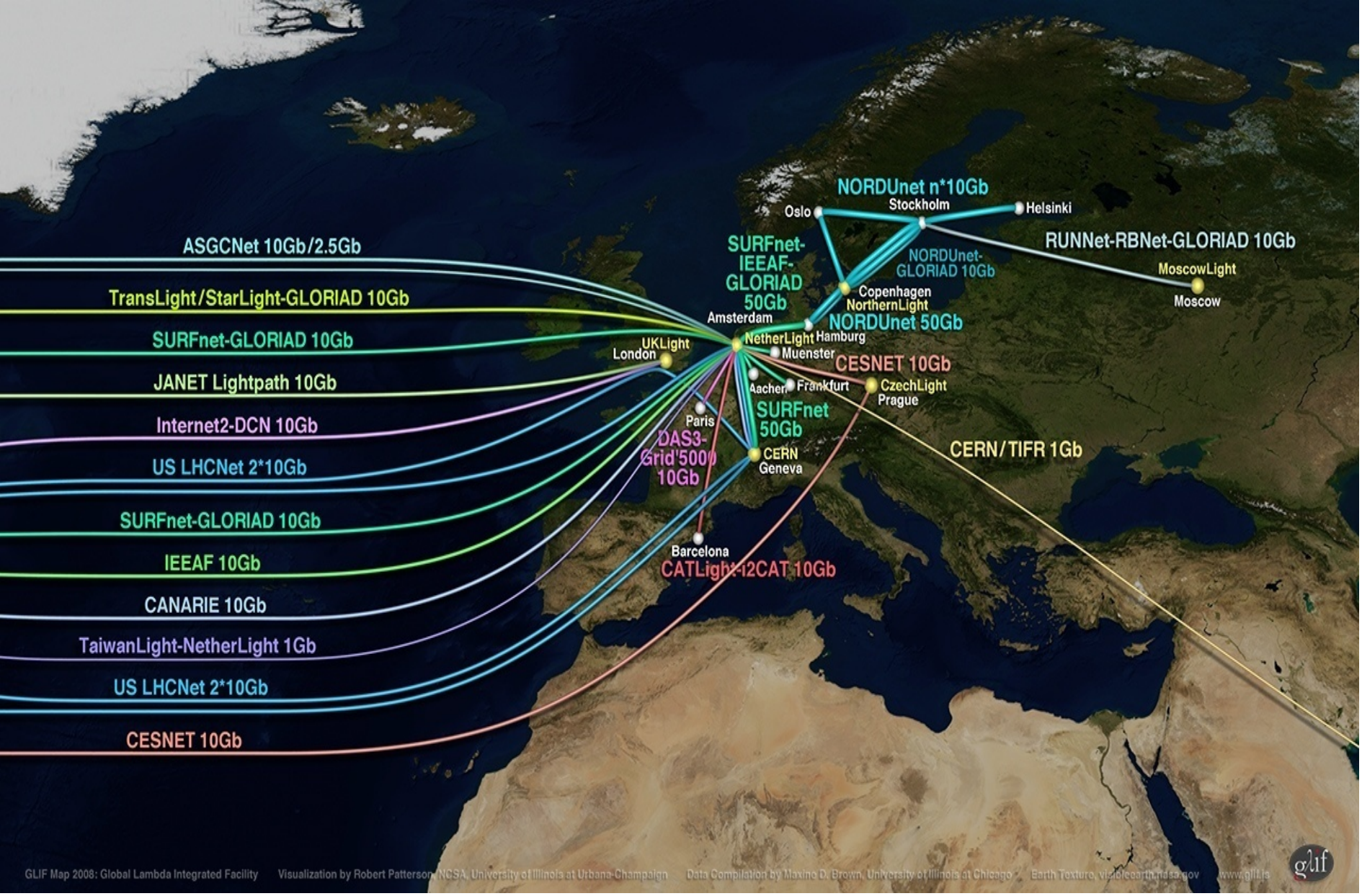
View from StarLight

Abbott Hall, Northwestern University's Chicago downtown campus

# iCAIR: Founding Partner of the Global Lambda Integrated Facility
# Available Advanced Network Resources

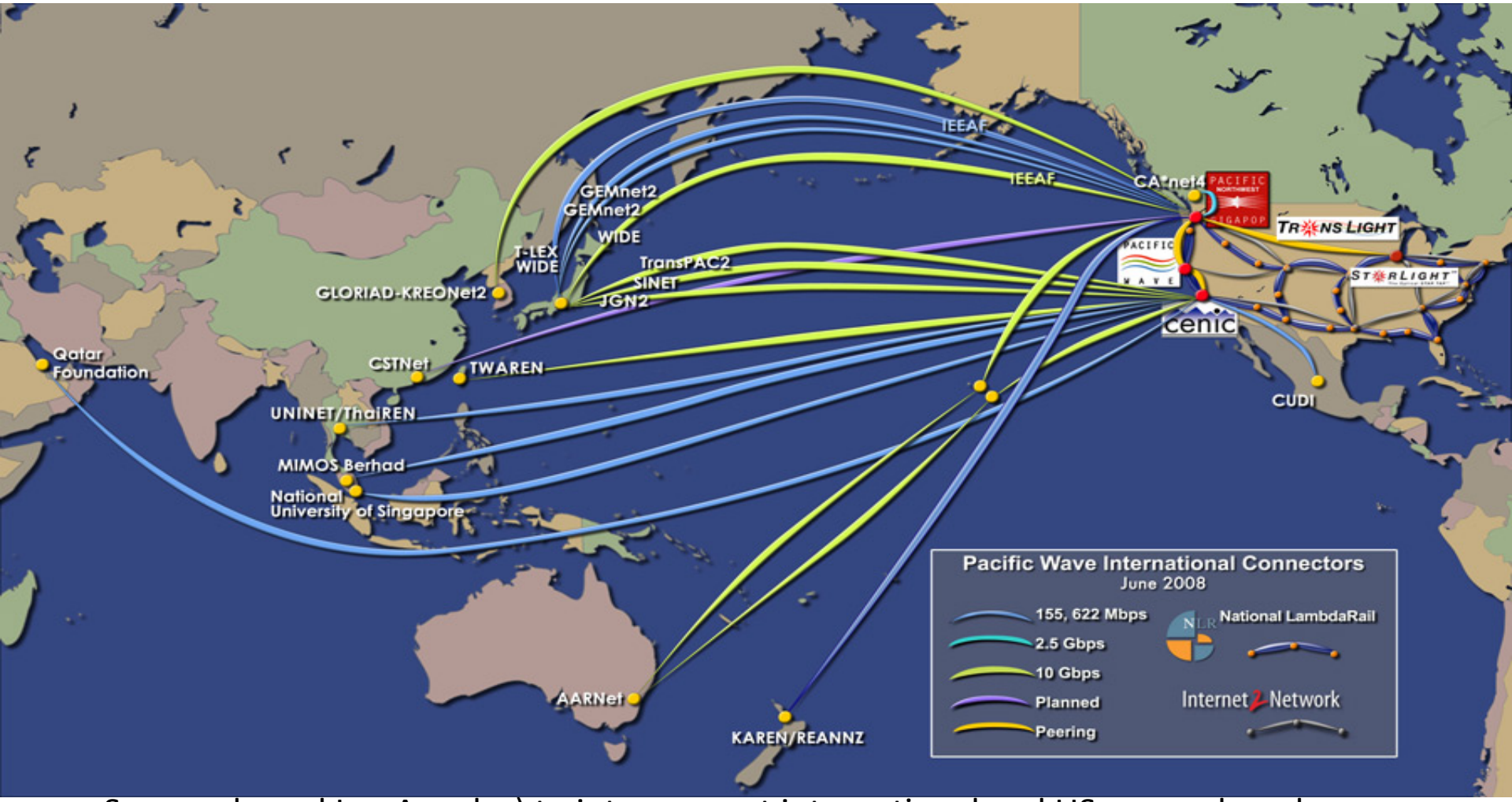

www.glif.is


ASGCNet 10Gb/2.5Gb
TaiwanLight 2.5Gb
KREONet2-GLORIAD 10Gb
IEEAF 10Gb
CSTNet-GLORIAD 2.5Gb
JGN2plus 10Gb
TaiwanLight 2.5Gb
AARNet-SXTransPORT 2*10Gb
Oahu
Hawaii
CANARIE n*10Gb
Calgary
Vancouver
Seattle
Pacific Wave
CiscoWave 2*10Gb
CANARIE n*10Gb
UltraScience Net 20Gb
TransLight 10Gb
CAVEwave 10Gb
Pacific Wave 2*10Gb
CiscoWave 2*10Gb
UltraScience Net 20Gb
ESnet-SDN 2*10Gb
National LambdaRail n*10Gb
Internet2-DCN n*10Gb
UltraLight 10-20Gb
ESnet-SDN 2*10Gb
Salt Lake City
Pacific Wave
Sunnyvale
Los Angeles
Pacific Wave
San Diego
Tijuana
CENIC 10Gb
WHREN-LILA 2Gb
Internet2-DCN n*10Gb
Denver
JGN2plus 10Gb
CiscoWave 2*10Gb
El Paso
CANARIE-GLORIAD 2.5Gb
KREONet2-GLORIAD 1Gb
CSTNet-GLORIAD 1Gb
NLR n*10Gb-GLORIAD 10Gb
Fermi Lightpath 22Gb
Kansas City
Tulsa
National LambdaRail n*10Gb
Houston
Baton Rouge
Winnipeg
CANARIE n*10Gb
Chicago
StarLight
Toronto
Urbana
IllinoisWave 10Gb
LEAD 10Gb
SLR 10Gb-GLORIAD 1Gb
Internet2-DCN n*10Gb
LONI 10Gb
Atlanta
Oak Ridge
Raleigh
Montreal
Halifax
Boston
MAN LAN
New York
Washington DC
CAVEwave 10Gb
TeraFlow 10Gb
CiscoWave 10Gb
EnLightened 10Gb
ESnet-SDN 10Gb
Internet2-DCN n*10Gb
National LambdaRail n*10Gb
Jacksonville
AtlanticWave 10Gb
AMPATH
Miami
ASGCNet 10Gb/2.5Gb
TransLight/StarLight-GLORIAD 10Gb
SURFnet-GLORIAD 10Gb
JANET Lightpath 10Gb
Internet2-DCN 10Gb
US LHCNet 2*10Gb
SURFnet-GLORIAD 10Gb
IEEAF 10Gb
CANARIE 10Gb
TaiwanLight-NetherLight 1Gb
US LHCNet 2*10Gb
CESNET 10Gb
WHREN-LILA 2.5Gb
GLIF Map 2008: Global Lambda Integrated Facility
Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign
Data Compilation by Maxine D. Brown, University of Illinois at Chicago
Earth Texture, visibleearth.nasa.gov
www.glif.is
glif


ASGCNet 10Gb/2.5Gb
TransLight/StarLight-GLORIAD 10Gb
SURFnet-GLORIAD 10Gb
JANET Lightpath 10Gb
Internet2-DCN 10Gb
US LHCNet 2*10Gb
SURFnet-GLORIAD 10Gb
IEEAF 10Gb
CANARIE 10Gb
TaiwanLight-NetherLight 1Gb
US LHCNet 2*10Gb
CESNET 10Gb
NORDUnet n*10Gb
Oslo
Stockholm
Helsinki
SURFnet-IEEAF-GLORIAD 50Gb
NORDUnet-GLORIAD 10Gb
RUNNet-RBNet-GLORIAD 10Gb
MoscowLight
Moscow
Copenhagen
NorthernLight
Amsterdam
NORDUnet 50Gb
NetherLight
Hamburg
UKLight
London
Muenster
CESNET 10Gb
Aachen
Frankfurt
CzechLight
Prague
Paris
SURFnet 50Gb
DAS3-Grid'5000 10Gb
CERN
Geneva
CERN/TIFR 1Gb
Barcelona
CATLight-I2CAT 10Gb
GLIF Map 2008: Global Lambda Integrated Facility
Visualization by Robert Patterson, NCSA, University of Illinois at Urbana-Champaign
Data Compilation by Maxine D. Brown, University of Illinois at Chicago
Earth Texture, visibleearth.nasa.gov
www.glif.is
glif

# TransLight/Pacific Wave/StarLight

10GE Wave Facilitates US West Coast Connectivity




Sunnyvale and Los Angeles) to interconnect international and US research and education networks

www.pacificwave.net/participants/irnc/


CANARIE
GLORIAD 2.5Gb
KREONet2-
GLORIAD 1Gb
CSTNet-
GLORIAD 1Gb
NLR n*10Gb-
GLORIAD 10Gb
Fermi
CANARIE n*10Gb
MiLR-
UltraLight
10Gb
Chicago
StarLight
Kansas City
Urbana
IllinoisWave
10Gb
LEAD
10Gb
Tulsa
SLR 10Gb-
GLORIAD
1Gb
Internet2-DCN
n*10Gb
LONI 10Gb
Atlanta
Indianapolis
Oak
Ridge
Raleigh
Pittsbu
NGIX