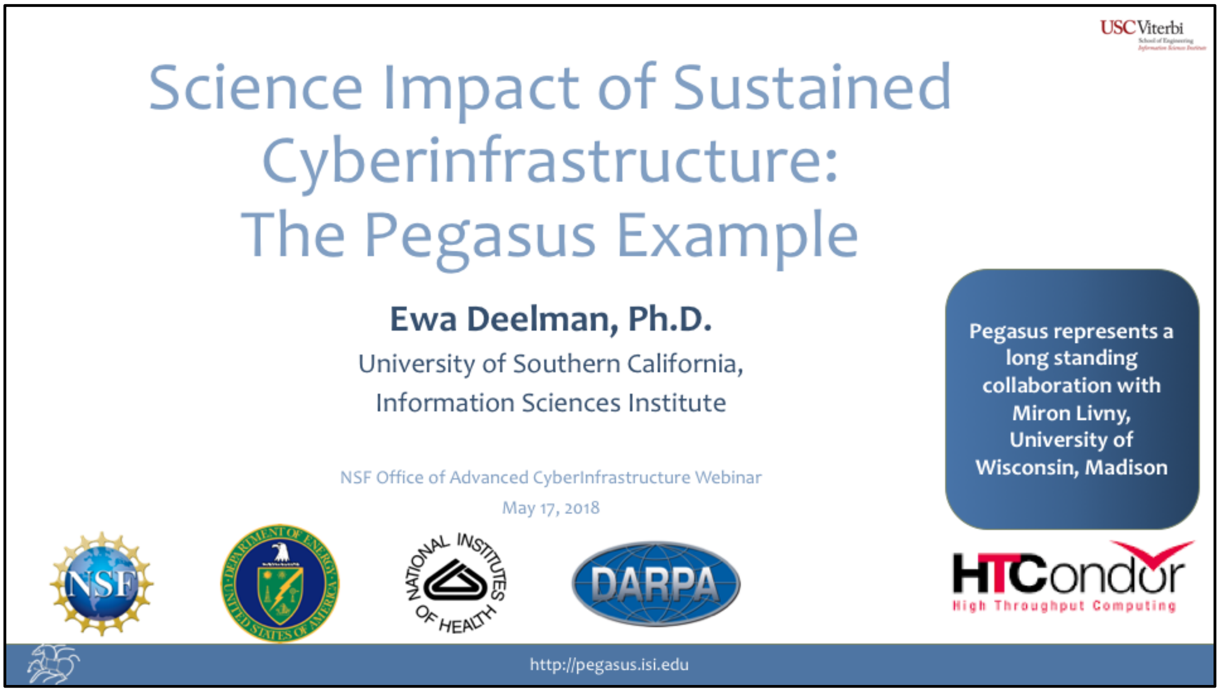
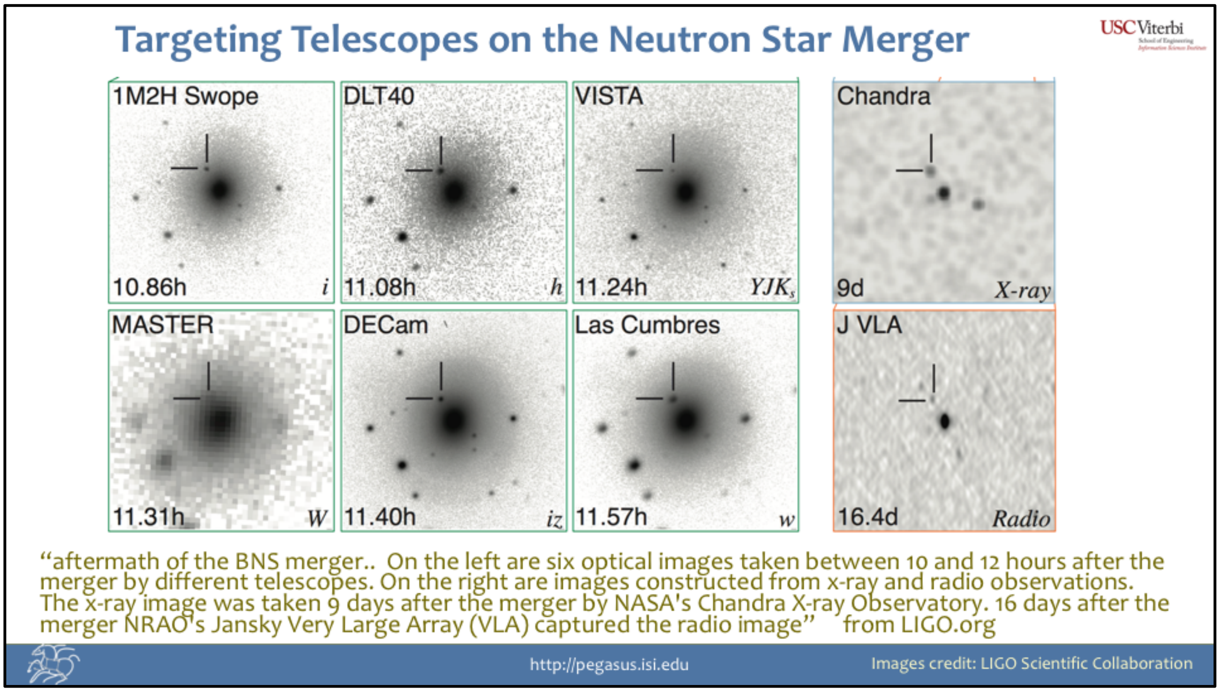


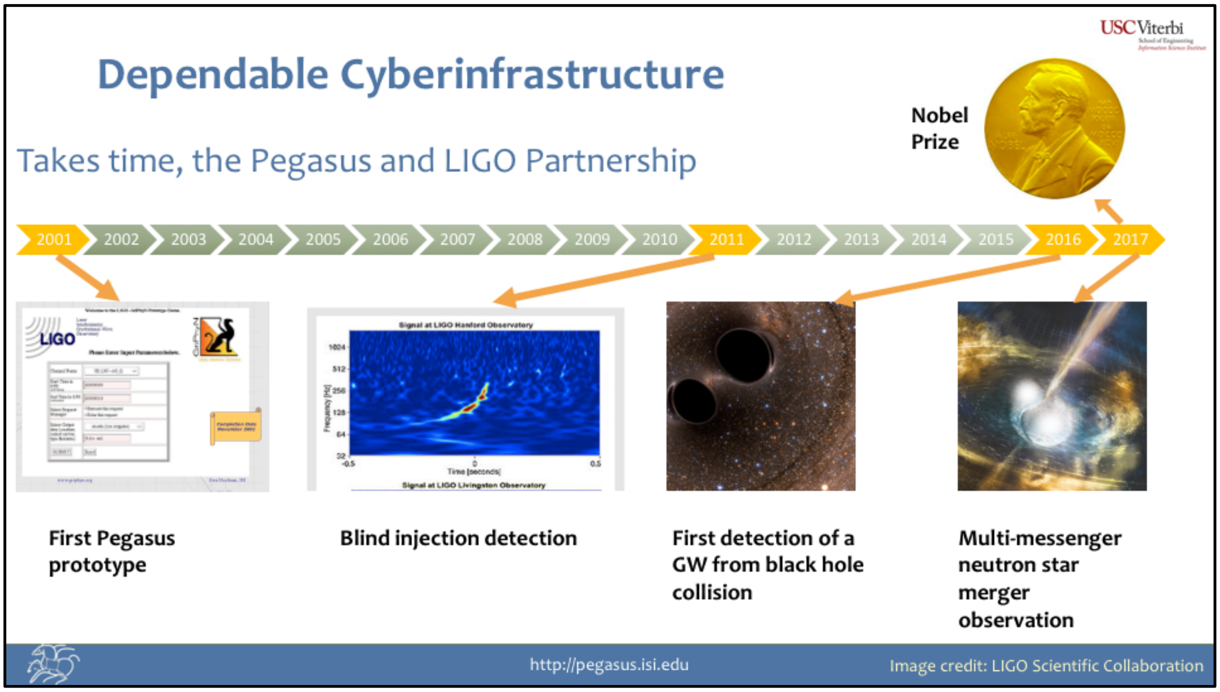
Slide 1: Title slide. Science Impact of Sustained Cyberinfrastructure: The Pegasus Example. Ewa Deelman, USC. Title slide also indicates that Pegasus represents a long standing collaboration with Miron Livny, University of Wisconsin, Madison

Slide 2: Pictures depicting colliding neutron stars, before and during the merger

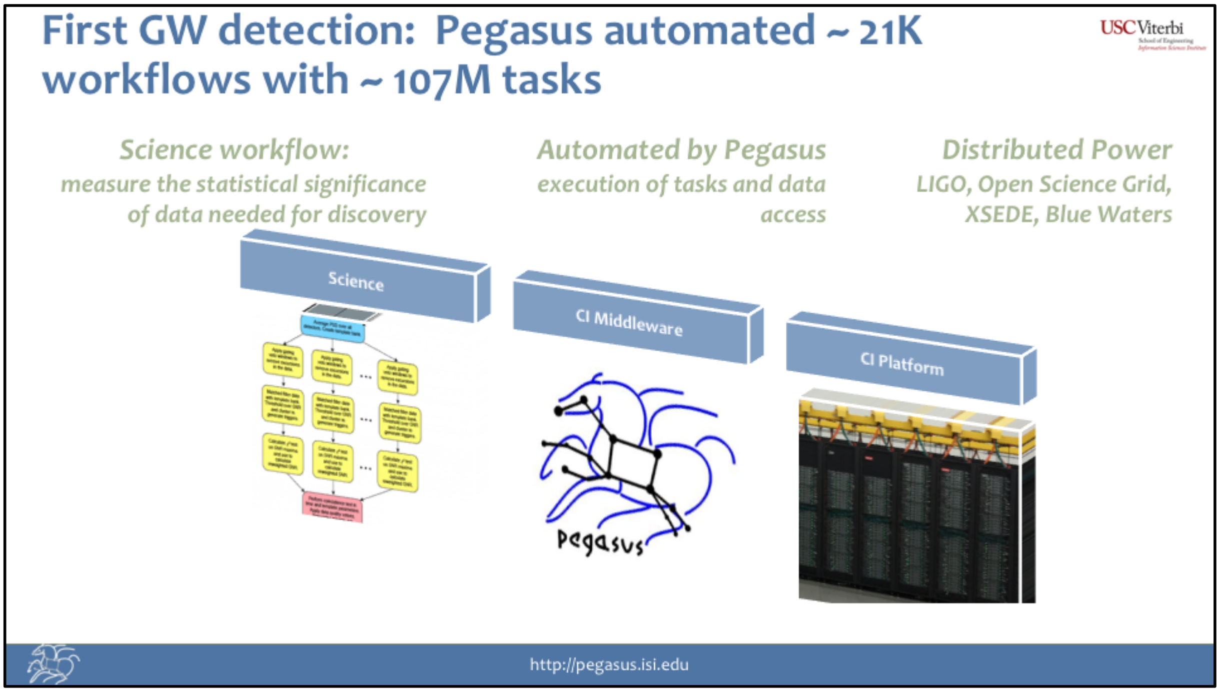
The caption for the picture is
The inspiral and merger of two neutron stars, as illustrated here, should produce a very specific gravitational wave signal, but the moment of the merger should also produce electromagnetic radiation that's unique and identifiable as such.", credit LIGO



Slides 3: Pictures depict the colliding neutron stars as seen by various telescopes hours and days after the merger.

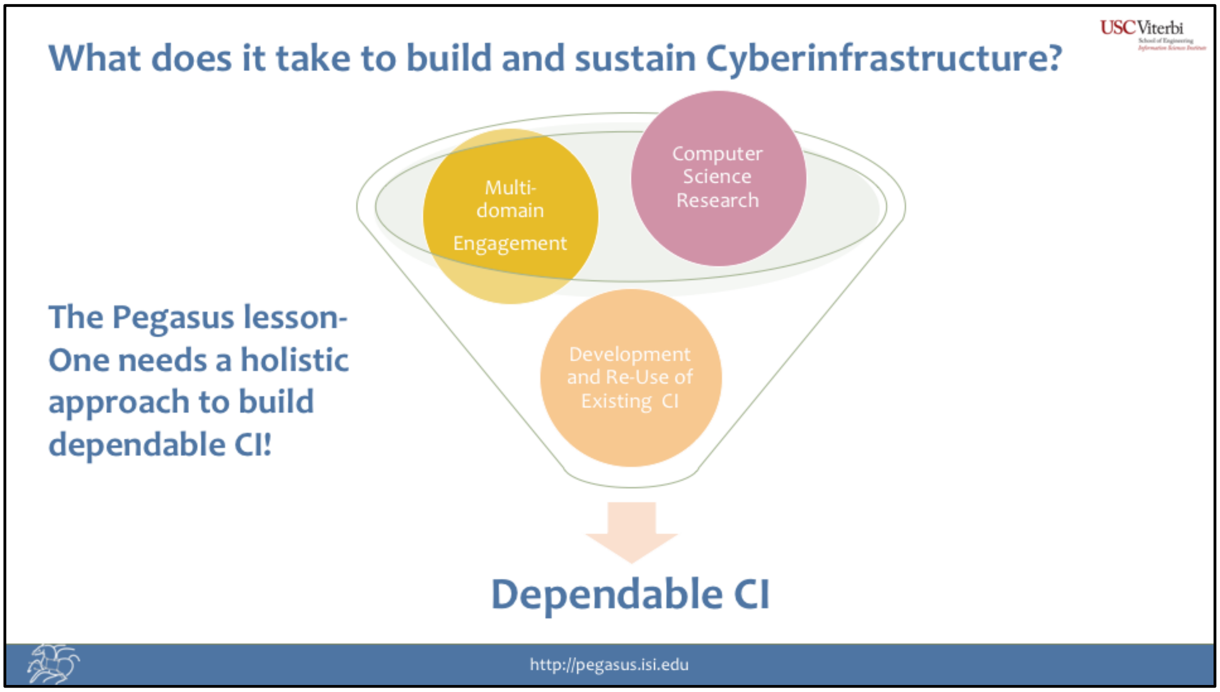


Slide 4: Shows the timeline of the Pegasus and LIGO partnership since 2001 to 2017, includes the first prototype in 2001, the detection of the blind injection in 2011, the first ever direct detection of the gravitational wave from colliding black holes in 2016, and the Nobel Prize in 2017.



Slide 5: Description of the Pegasus automated worflows.

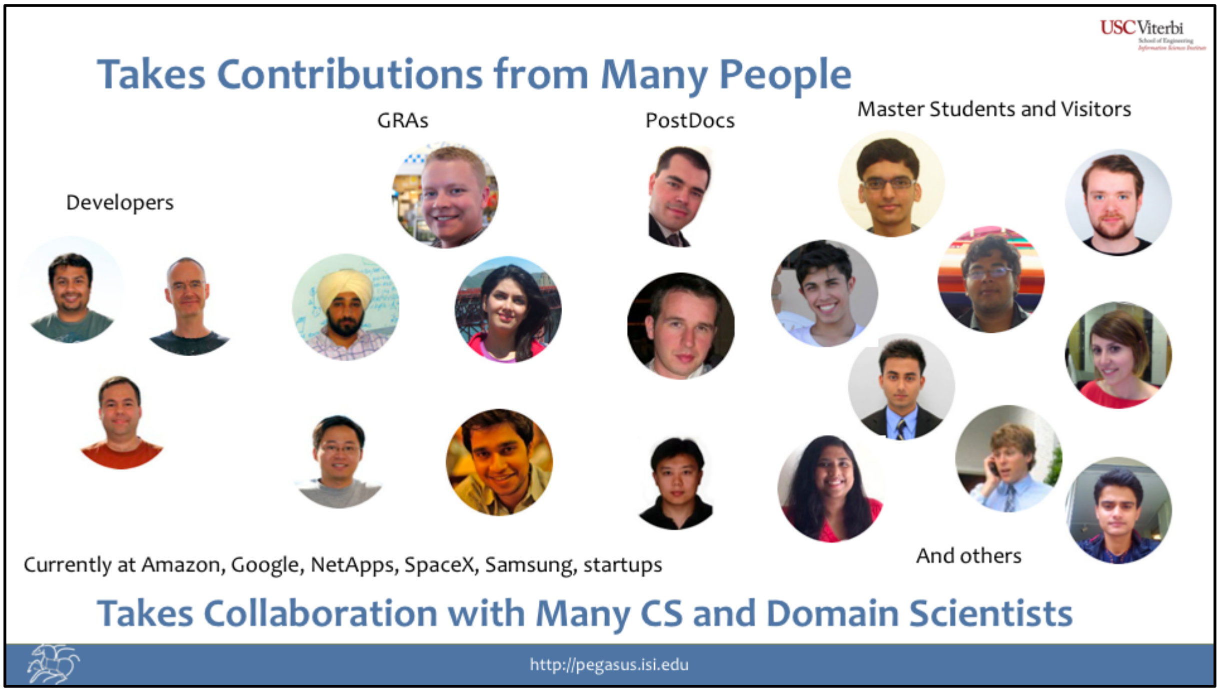
Peter Couvares, data analysis computing manager for the Advanced LIGO project at Caltech: “When a workflow might consist of 600,000 jobs, we don’t want to rerun them if we make a mistake. So we use DAGMan(Directed Acyclic Graph Manager, a meta-scheduler for HTCondor) and Pegasus workflow manager to optimize changes,”



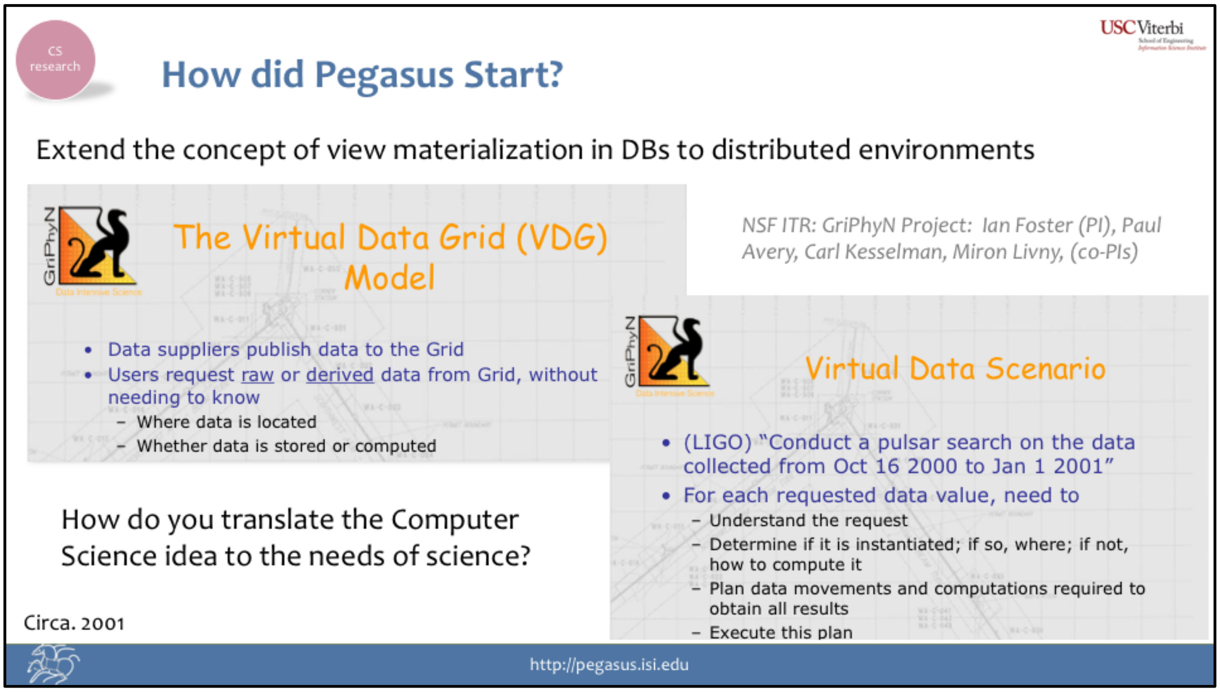
Slide 6: Pictures depict three important aspects of CI. 1) Multi- domain engagement, 2) Computer science research, 3) development and re-use of existing CI

Slide 7: Timeline from 2000 to 2018 with Pegasus team members. Karan Vahi, the main Pegasus developer has been working on the project for 17 years.

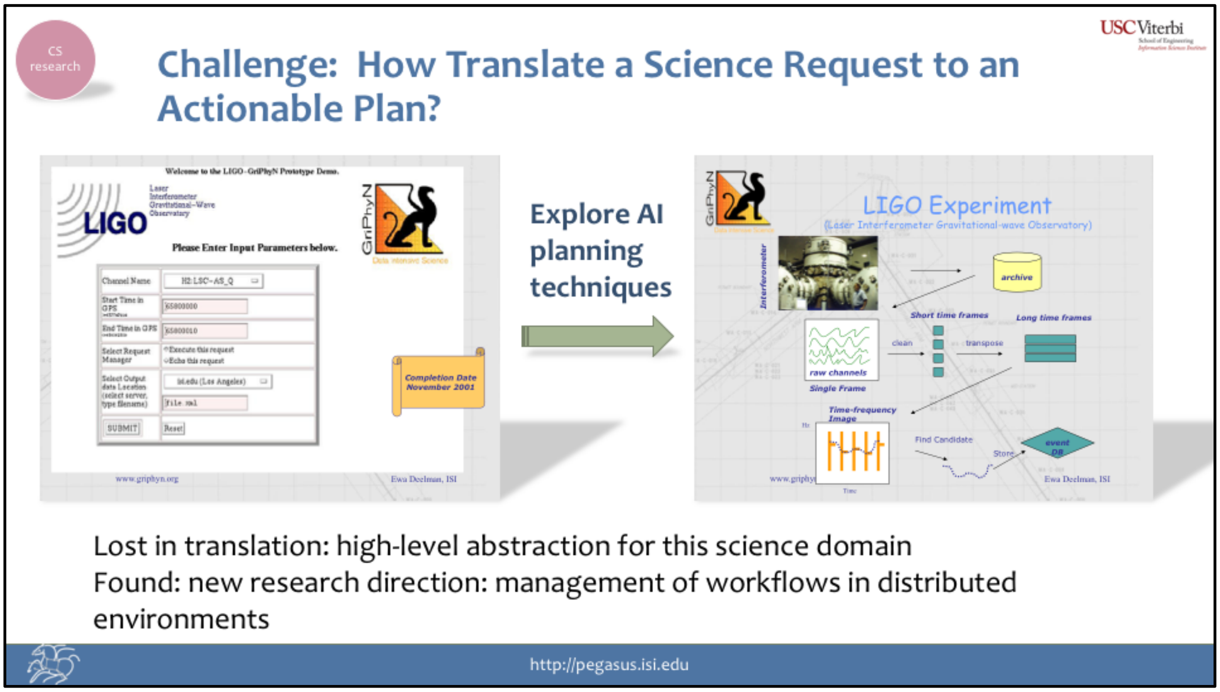
Back Row: Tu Mai Anh Do, Mats Rynge, Karan Vahi, George Papadimitriou
Front Row: Rosa Filgueira, Ewa Deelman, Rajiv Mayani
Missing: Rafael Ferreira da Silva, Ashwin Venkatesha



Slide 8: Takes Collaboration with Many CS and Domain Scientists



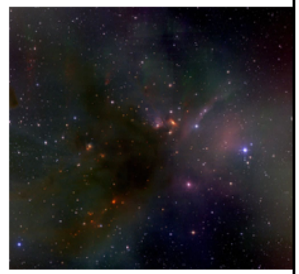
Slide 9: Pictures describes the idea of Virtual Data Grid and Virtual Data Scenario.



Slide 10: On the left a picture of the user interface for the first LIGO prototype, on the right the resulting scientific workflow.

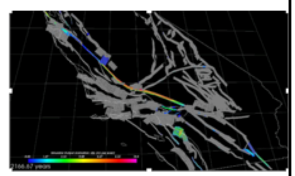
Challenges of Workflow Management

- Working with LIGO and other applications (astronomy, earthquake science), found common challenges:
  - Need to describe complex workflows in a simple way
  - Need to access distributed, heterogeneous data and resources
  - Need to deal with resources/software that change over time



Sky mosaic, IPAC, Caltech

- Our focus:
  - Separation between workflow description and workflow execution
  - Workflow planning and scheduling (scalability, performance)
  - Task execution (monitoring, fault tolerance, debugging)



Earthquake simulation, SCEC, USC

http://pegasus.isi.edu

Slide 11: Describes workflow management challenges. The images show sample results from astronomy and earthquake science.

Multi-domain

USC Viterbi School of Engineering Information Sciences Institute

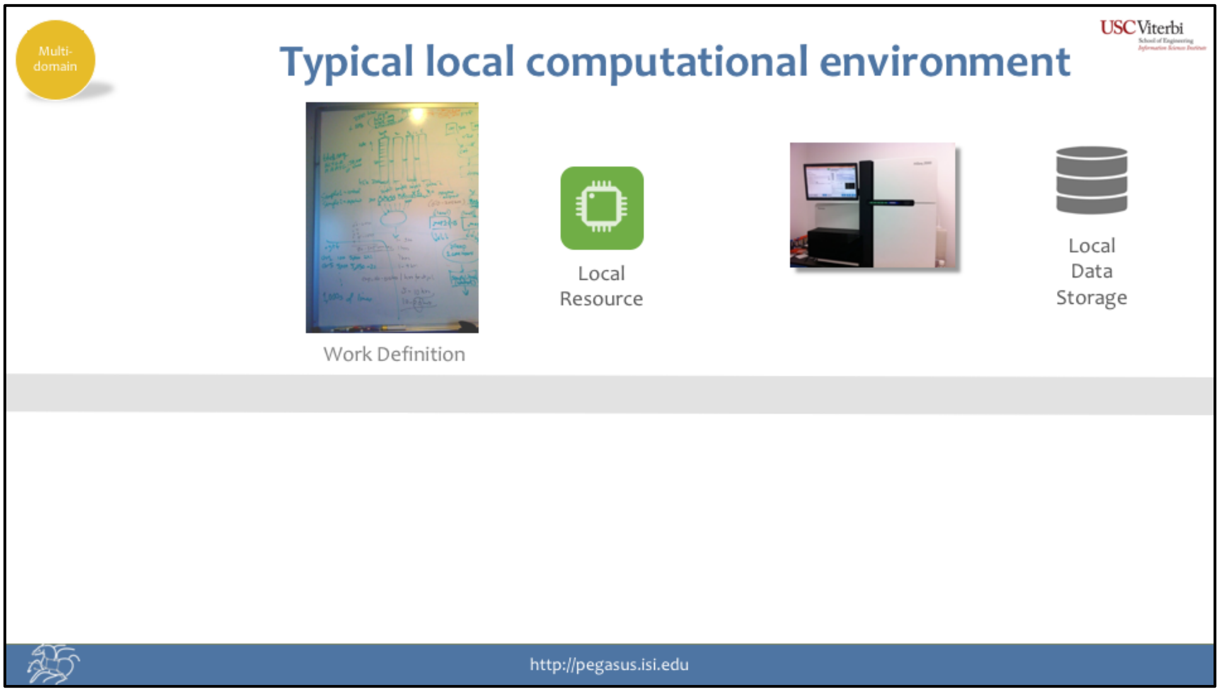
Benefits of Scientific Workflows (from the point of view of an application scientist)

- Conducts a series of computational tasks
- Chaining (outputs become inputs) replaces manual hand-offs
- Ease of use: gives non-developers access to sophisticated codes
- Provides framework to host or assemble community set of applications, can be multi-disciplinary
- Framework to define common formats or standards when useful

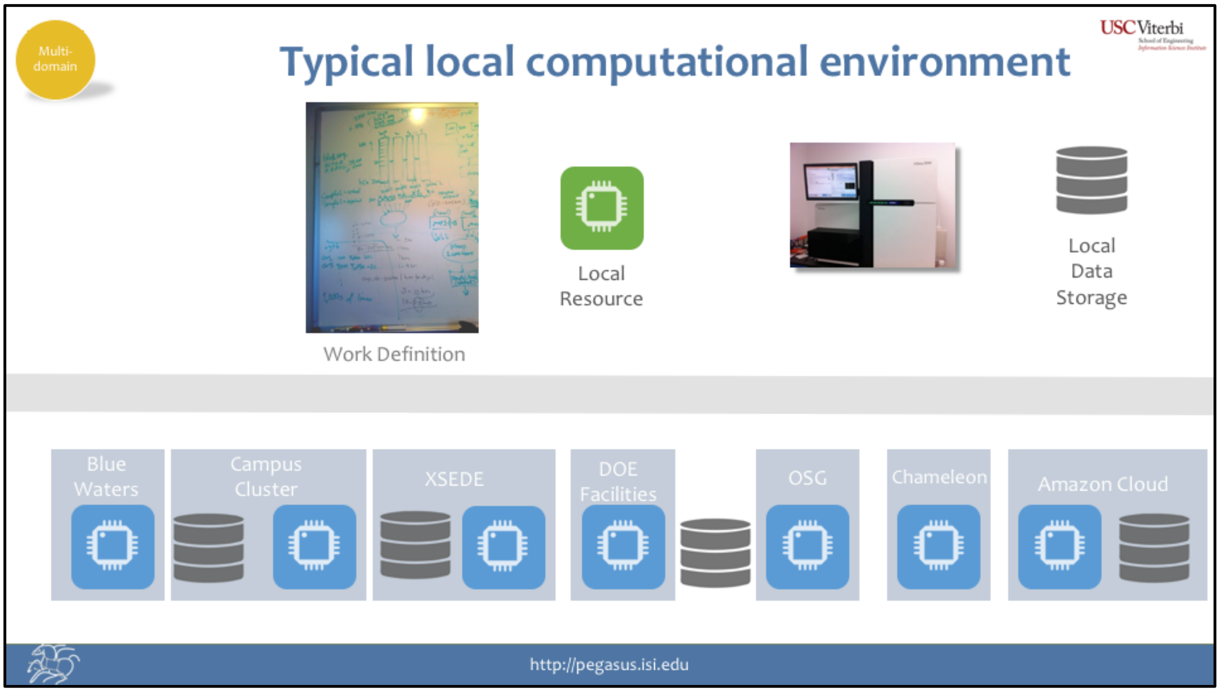
http://pegasus.isi.edu

Slide credit: David Okaya, SCEC, USC

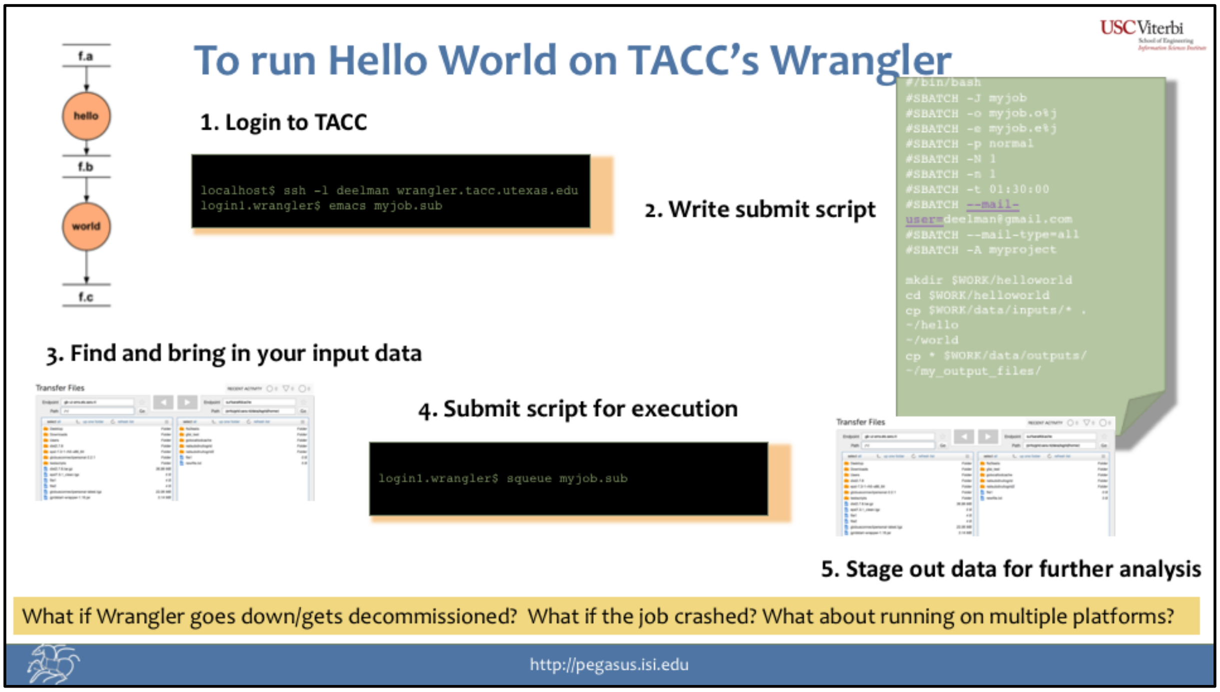
Slide 12: Describes the benefit of science workflow.



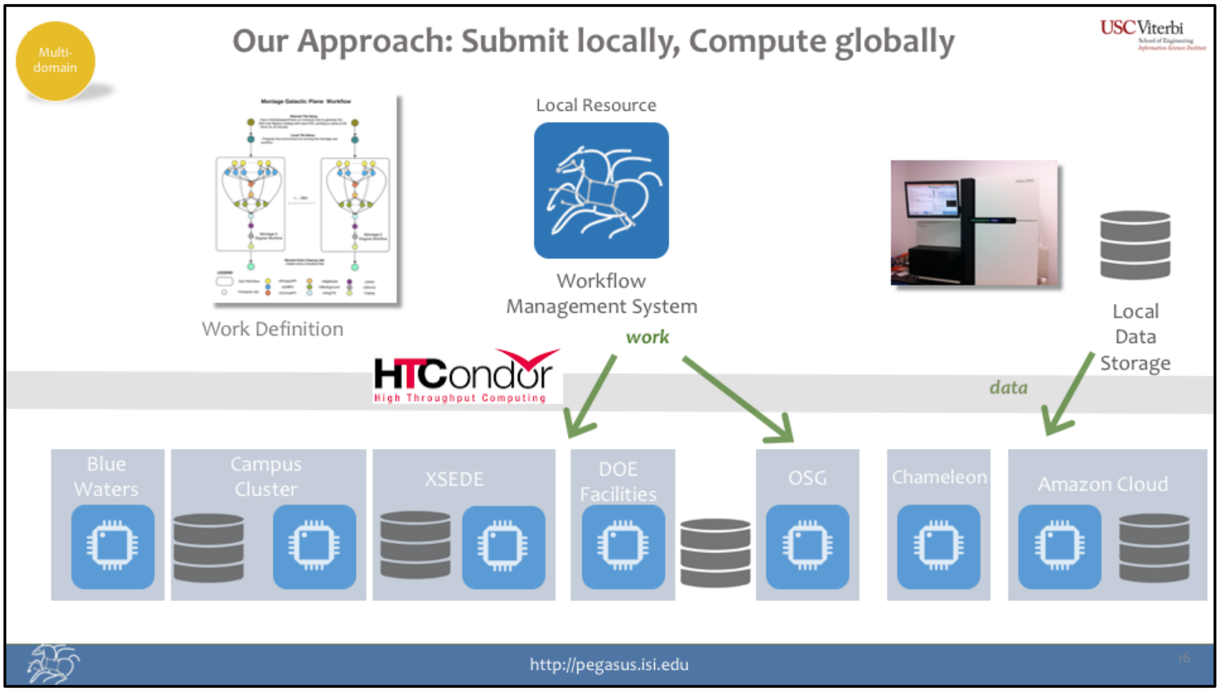
Slide 13: Pictures describe typical computational workflow environment with local resources and storage.



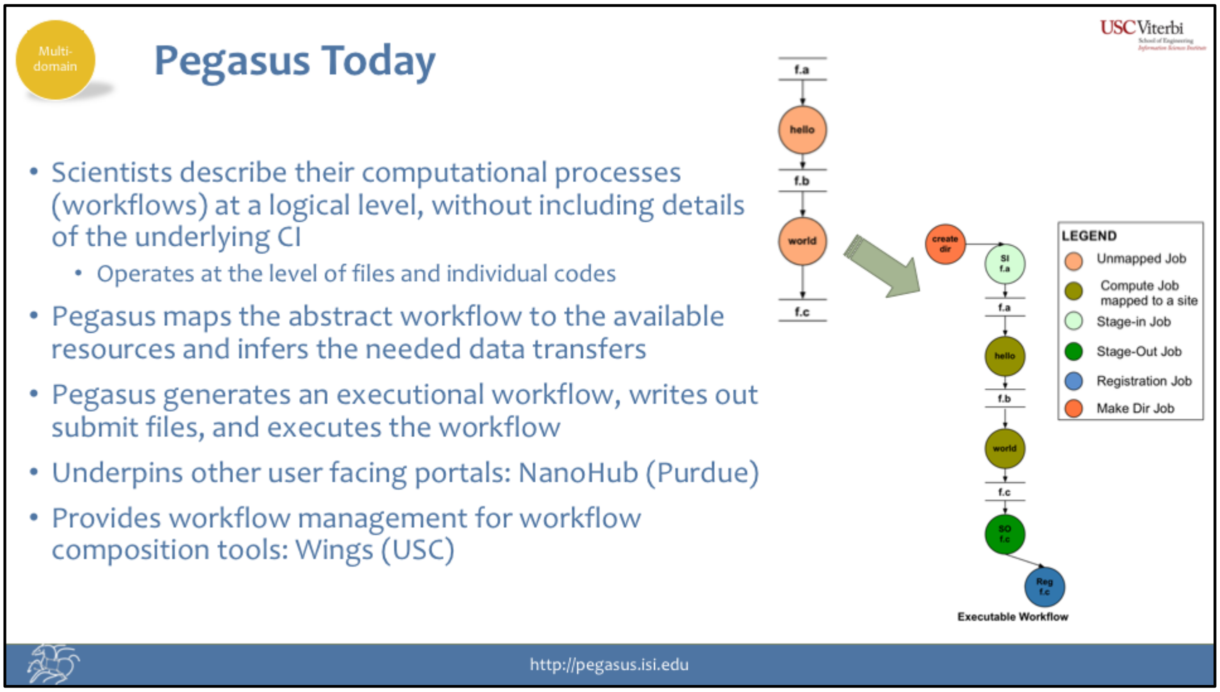
Slide 14: Additional description of computational workflow environment that includes campus resources, national cyberinfrastructure, NSF and commercial clouds.



Slide 15: Pictures walk the viewer through a programming workflow. 1) Login to computer, 2) Write script, 3) Find and download data, 4) Submit script to execute on system.



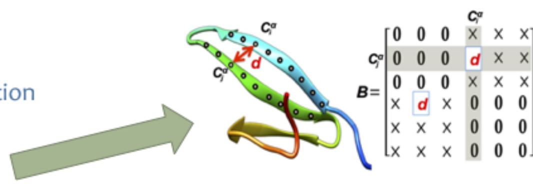
Slide 16: Pictures describe typical computational workflow environment with local resources and storage that accesses global resources to execute.



Slide 17: add at the end: The picture on the left shows the abstract workflow that is devoid of resource description. The picture on the right shows the executable workflow generated by Pegasus and that include compute jobs mapped to an execution site, data stage-in and stage out jobs, data registration jobs, and make directory jobs.

CS Principles Help in Cyberinfrastructure Development

- Structure workflows as **directed acyclic graphs (DAGs)**
  - Re-use of graph traversal algorithms, node clustering, pruning, other complex graph transformation
- Use hierarchical structures in DAGs
  - To achieve scalability, recursion, dynamic behavior
- Develop new algorithms:
  - Task clustering
  - Data placement
  - Data re-use
  - Resource usage estimation
  - Resource provisioning
  - ***In situ* workflows**

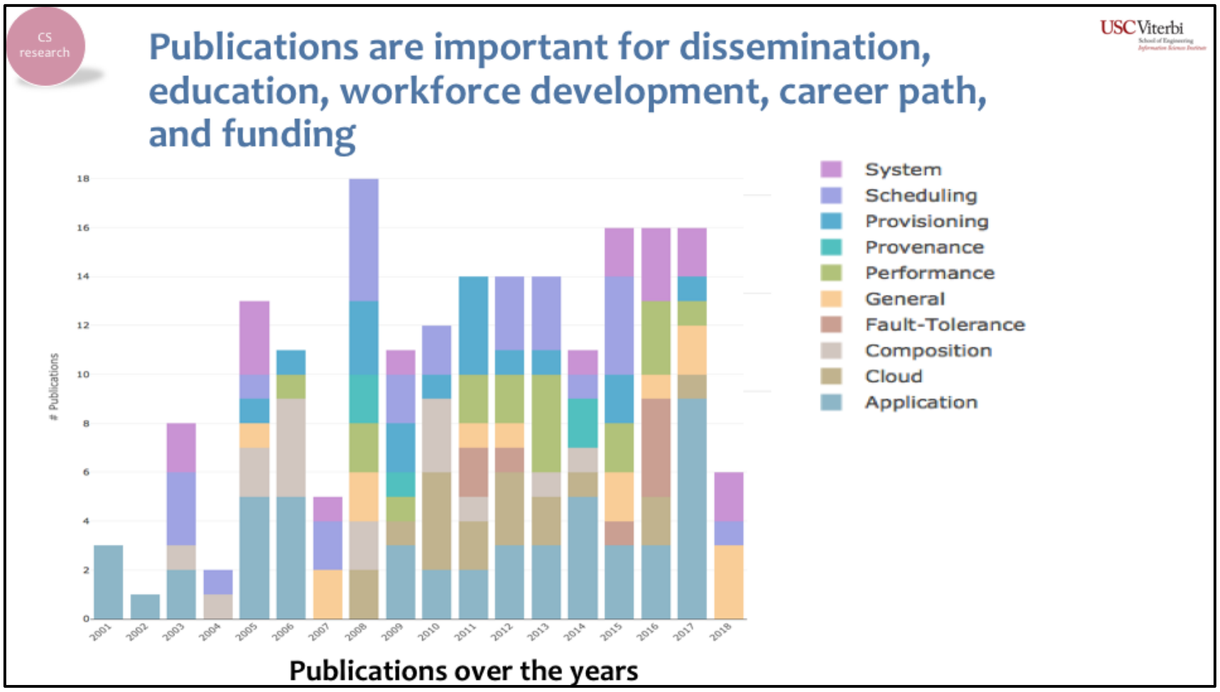


New Direction: In-memory coupling of simulation and analytics Collaboration with U Delaware, Cornell, UTEP

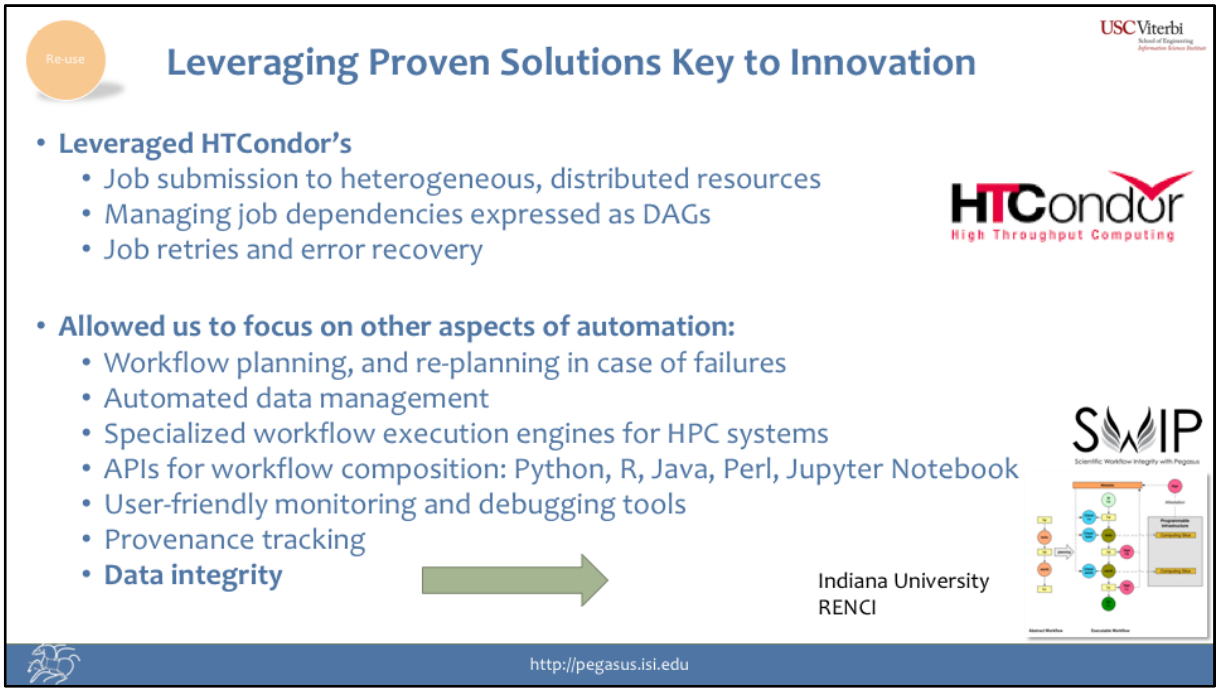
http://pegasus.isi.edu

Image credit: Michela Taufer, U. of Delaware

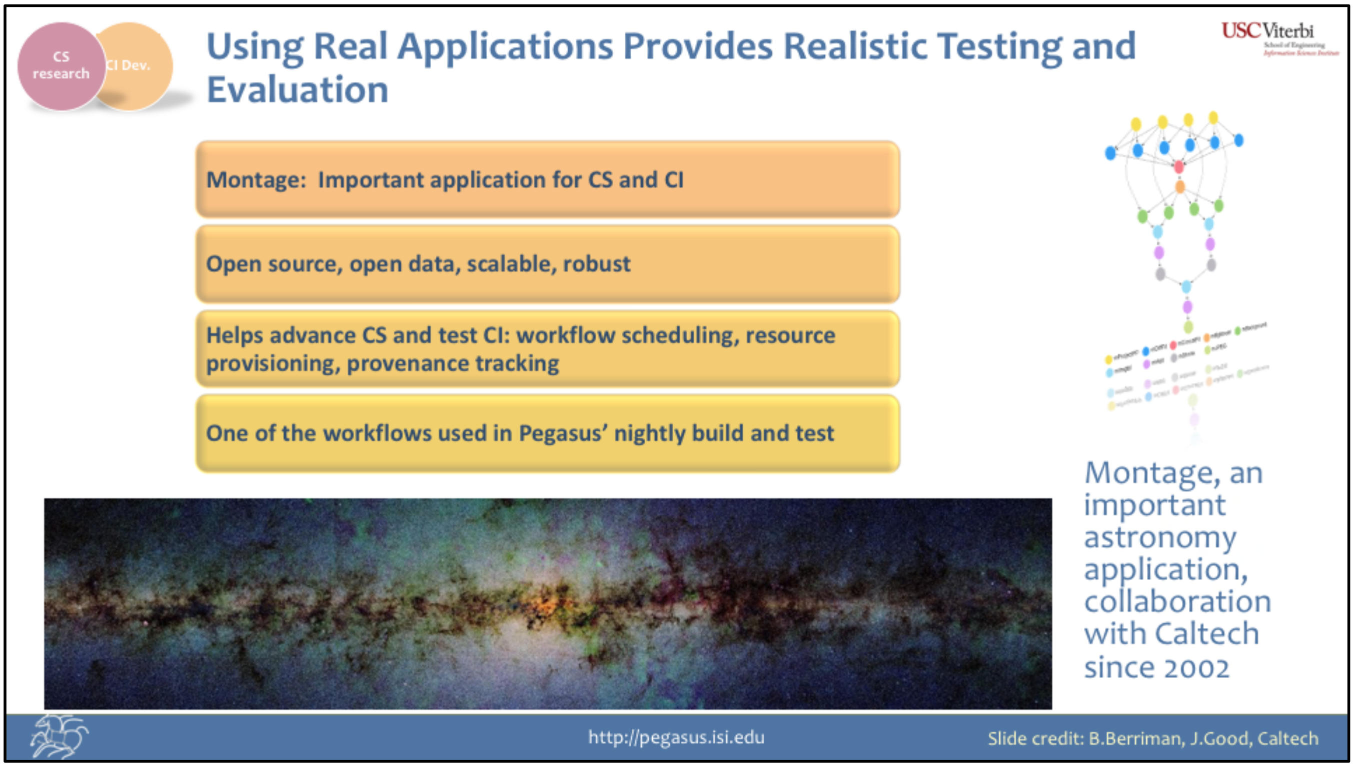
Slide 18: Describes CS principles in helping to develop cyberinfrastructure.



Slide 19: The graphs shows the publications by year starting in 2001 and ending in 2018. The publications span topics such as workflow system, job scheduling, resource provisioning, provenance, performance, general workflow management, fault

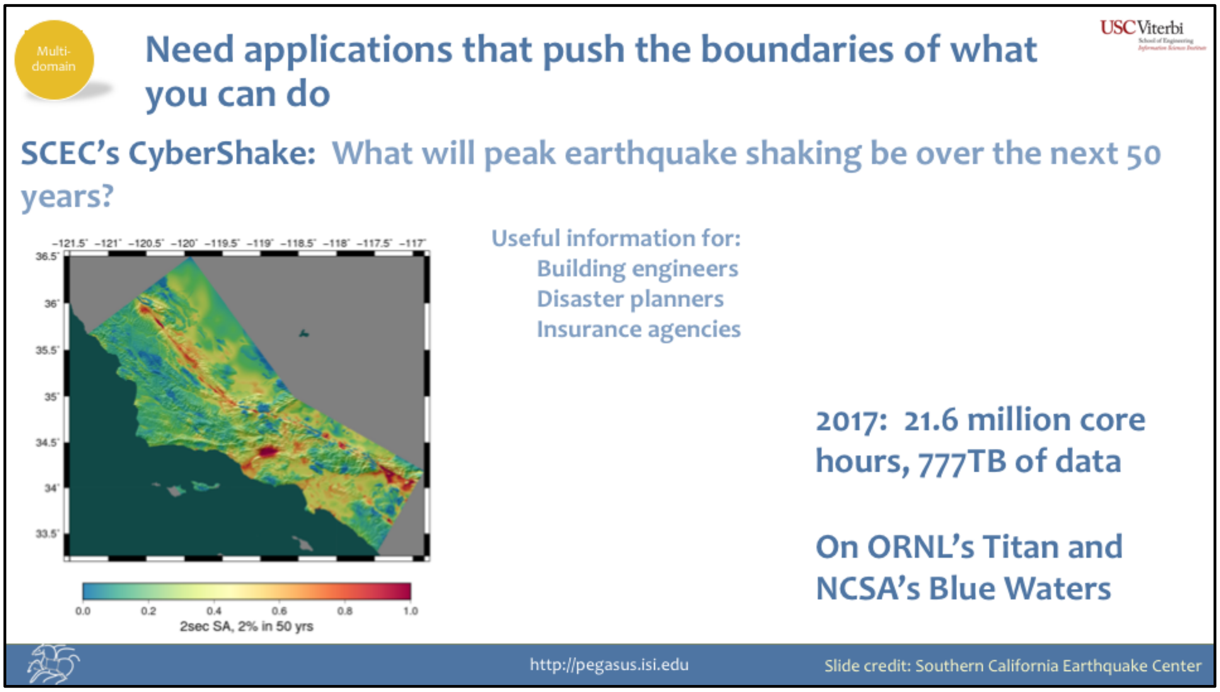


Slide 20: Mentions the SWIP data integrity project.

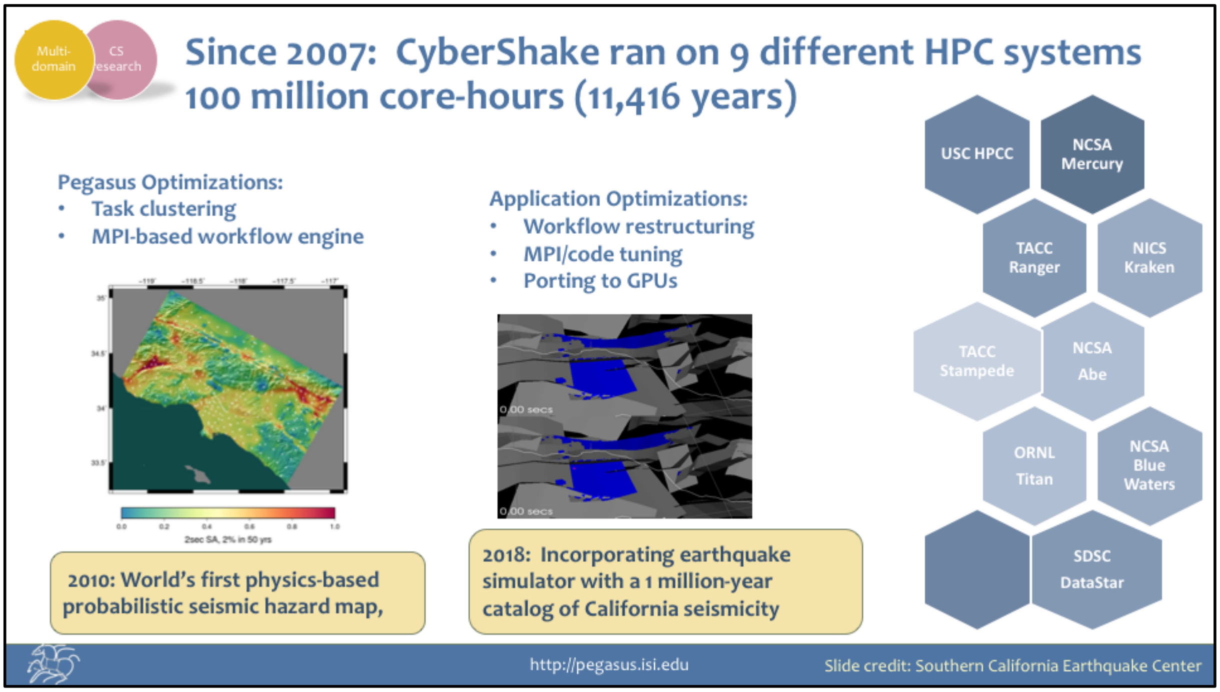


Slide 21: Describes the Montage astronomy application. The application is also important for computer science research and for the testing of the cyberinfastructure. Pegasus uses it as one of the applications for nightly builds and tests.

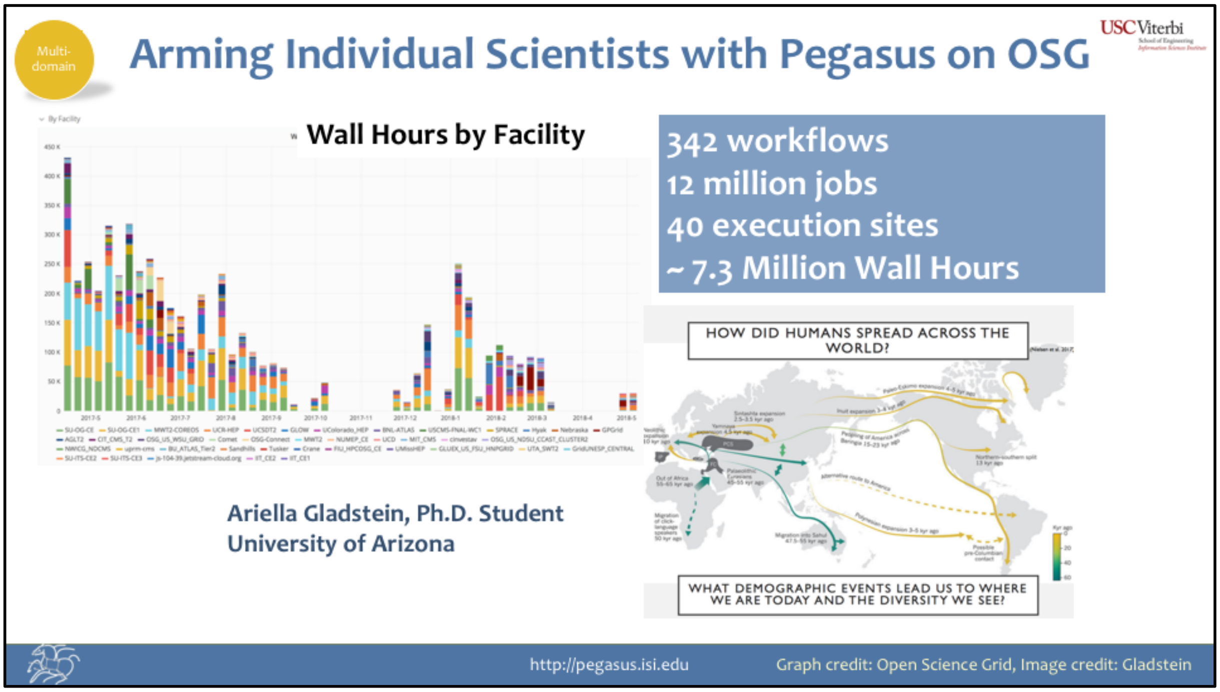
The picture shows the galactic plane generated by Montage.



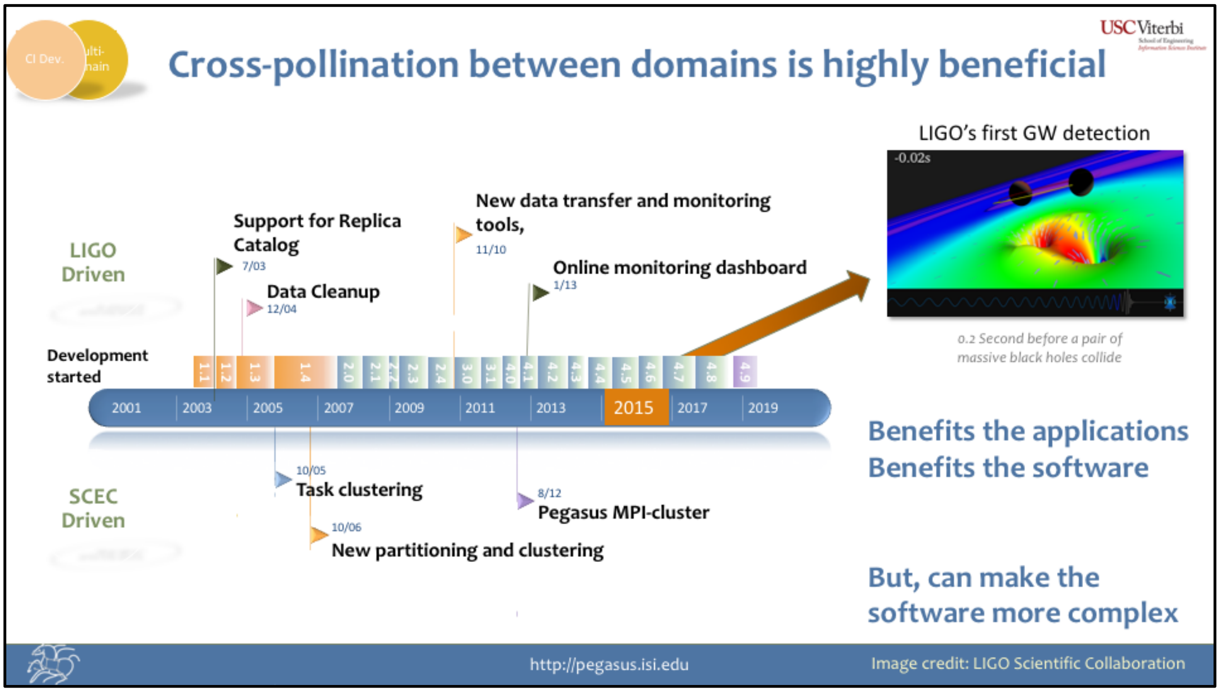
Slide 22: The picture shows a seismic hazard map for Southern California.

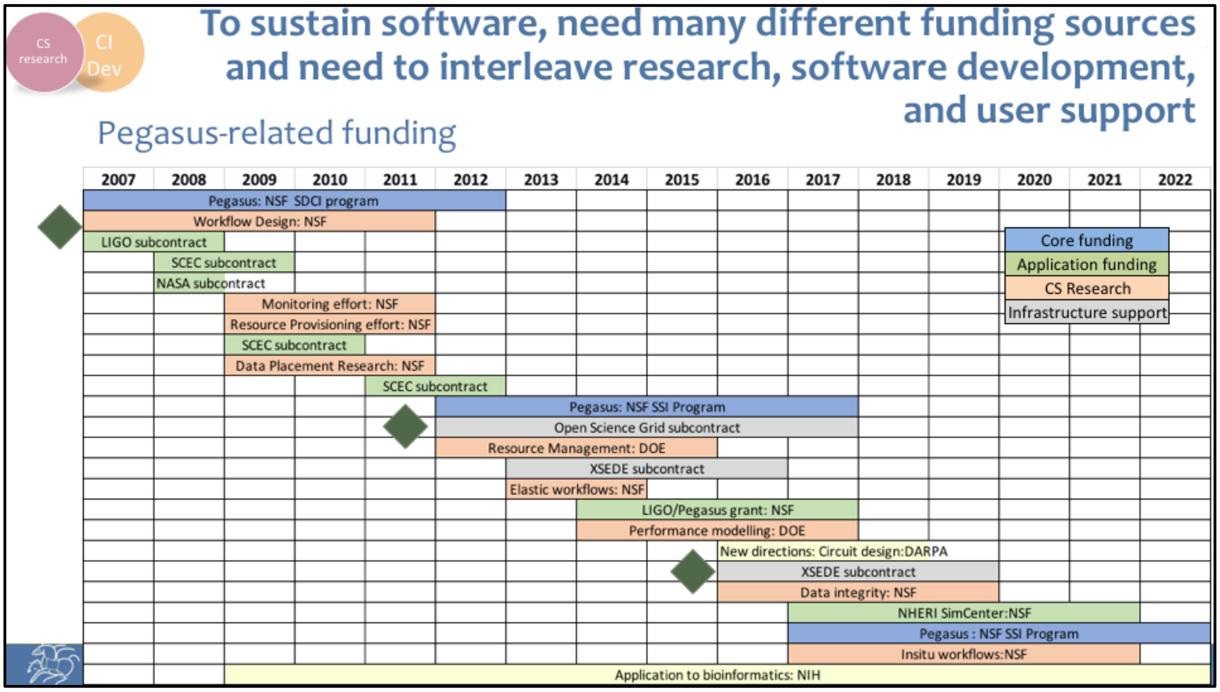


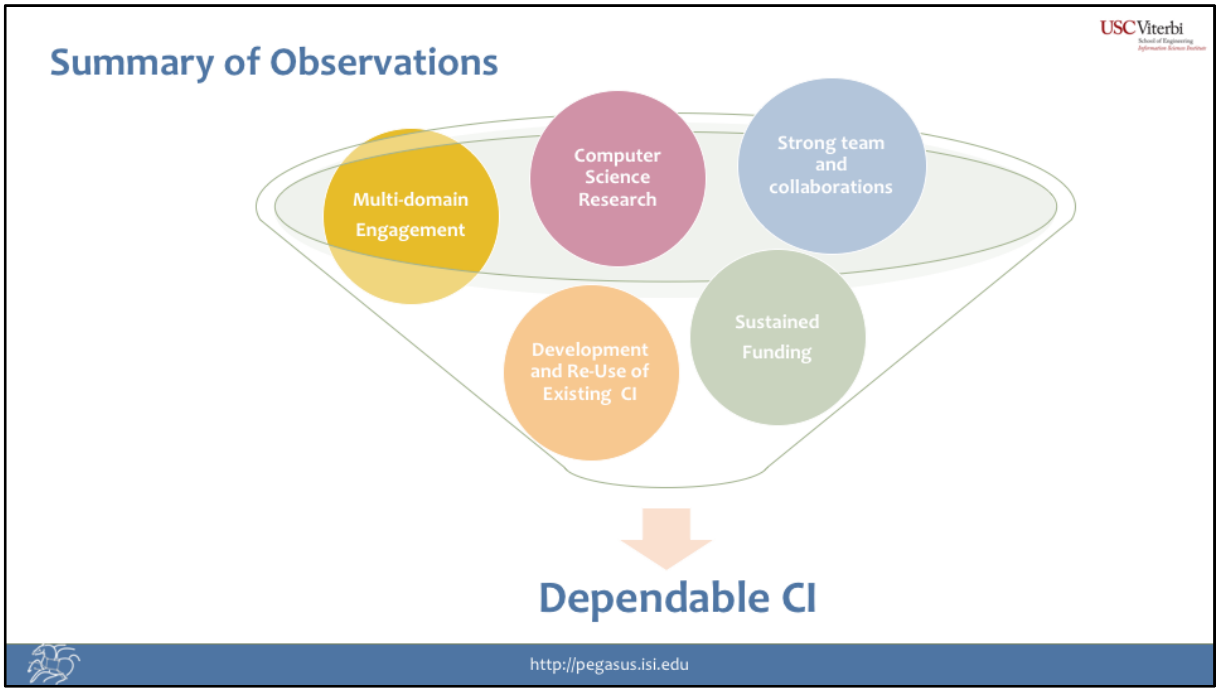
Slide 23: Shows the Cybershake map from 2010 and the results from the 2018 simulation of a million year of California seismicity. Also names the various HPC systems that CyberShake used over the last 16 years: USC, NCSA Mercury, TACC Ranger, NICS Kraken, TACC Stampede, NCSA Abe, ORNL Titan, NCSA Blue Waters, SDSC DataStar



Slide 24: Shows the amount of computing used by bioinformatics application supported by Pegasus on the Open Science Grid.


Cross-pollination between domains is highly beneficial
CI Dev.
USC Viterbi
LIGO's first GW detection
-0.02s
0.2 Second before a pair of massive black holes collide
LIGO Driven
SCEC Driven
Support for Replica Catalog
7/03
Data Cleanup
12/04
New data transfer and monitoring tools,
11/10
Online monitoring dashboard
1/13
Development started
1.1
1.2
1.3
1.4
2.0
2.1
2.2
2.3
2.4
3.0
3.1
4.0
4.1
4.2
4.3
4.4
4.5
4.6
4.7
4.8
4.9
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
10/05
Task clustering
8/12
Pegasus MPI-cluster
10/06
New partitioning and clustering
Benefits the applications
Benefits the software
But, can make the software more complex
http://pegasus.isi.edu
Image credit: LIGO Scientific Collaboration


CS research
CI Dev
To sustain software, need many different funding sources and need to interleave research, software development, and user support
Pegasus-related funding
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Pegasus: NSF SDCI program
Workflow Design: NSF
LIGO subcontract
SCEC subcontract
NASA subcontract
Monitoring effort: NSF
Resource Provisioning effort: NSF
SCEC subcontract
Data Placement Research: NSF
SCEC subcontract
Pegasus: NSF SSI Program
Open Science Grid subcontract
Resource Management: DOE
XSEDE subcontract
Elastic workflows: NSF
LIGO/Pegasus grant: NSF
Performance modelling: DOE
New directions: Circuit design:DARPA
XSEDE subcontract
Data integrity: NSF
NHERI SimCenter:NSF
Pegasus : NSF SSI Program
Insitu workflows:NSF
Application to bioinformatics: NIH
Core funding
Application funding
CS Research
Infrastructure support

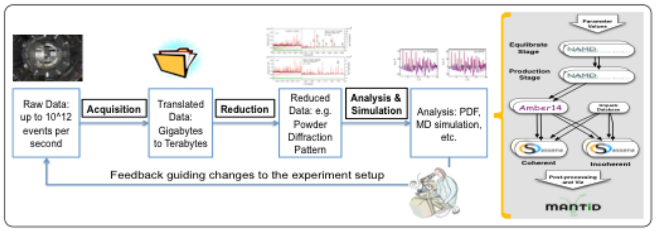


Slide 27: Slides shows the ingredients of dependable cyberinfrastructure: computer science research, multi-domain engagement, strong team and collaboration, development and re-use of existing cyberinfrastructure, sustained funding.

Looking ahead: Application Trends

USC Viterbi

Spallation Neutron Source



Planned CyberShake for Northern California:

- 869 geographic sites
- 16,000 workflow jobs
- 70 million core-hours on Blue Waters and Titan
- 800 TB of data

- More complex
- Faster time to solution: instrument steering
- More individual researchers in need of significant CI
  - Need intuitive workflow composition, better monitoring, error handling, assisted debugging

Outreach: How do you reach scientists that don't know you are out-there?
Many scientists are going through the same pain
Leverage/enhance existing engagement: NSF's Campus efforts? OSG/XSEDE outreach?
Education and outreach at instruments and experimental facilities?

http://pegasus.isi.edu

Image credit: Vickie Lynch, ORNL

Looking ahead: Growing Demand for Automation

HPC Systems
- Complex
- Heterogeneous
- Specialized data storage
- Increasingly faulty

Distributed Systems
- Software Defined capabilities
- Specialized data storage

Clouds
- New platform for science
- Very heterogenous
- Can be costly

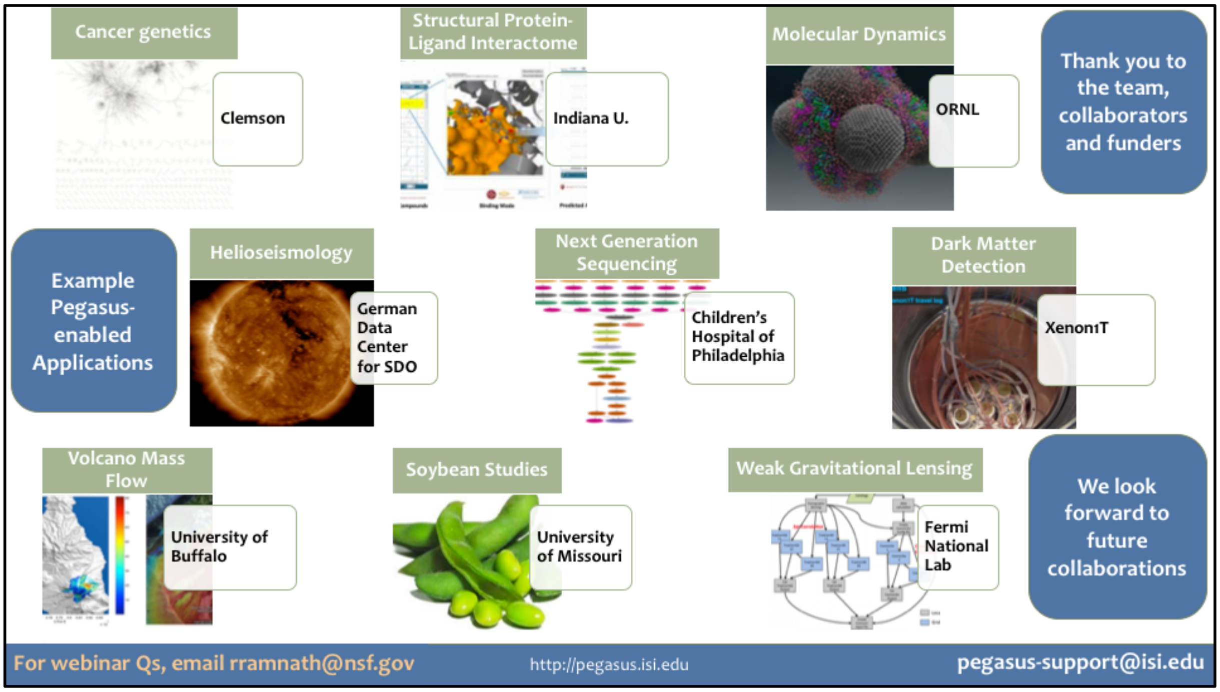
Resource Management is Key

Constraints: time to solution, budget

Faulty environment: failure detection and attribution

Heterogenous storage: memory, burst buffers, file systems, data xfer nodes

Need to keep track of big data technologies and machine learning solutions that are being developed at a rapid pace by industry



Cancer genetics
Clemson
Structural Protein-Ligand Interactome
Indiana U.
Molecular Dynamics
ORNL
Thank you to the team, collaborators and funders
Example Pegasus-enabled Applications
Helioseismology
German Data Center for SDO
Next Generation Sequencing
Children's Hospital of Philadelphia
Dark Matter Detection
Xenon1T
Volcano Mass Flow
University of Buffalo
Soybean Studies
University of Missouri
Weak Gravitational Lensing
Fermi National Lab
We look forward to future collaborations
For webinar Qs, email rramnath@nsf.gov
http://pegasus.isi.edu
pegasus-support@isi.edu