CICompass

Funded by the National Science Foundation, Grant #2127548

CI for Big Science as a Team Sport

by Frank Würthwein

NSF Ecosystem Lightning Talks

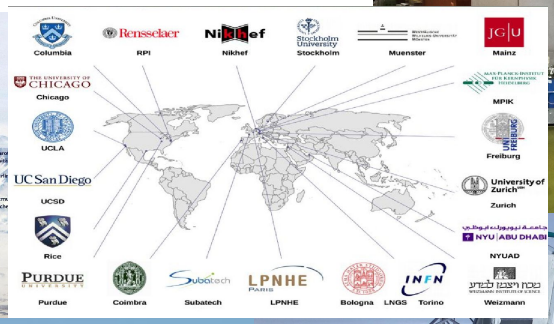
March 1, 2022
CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

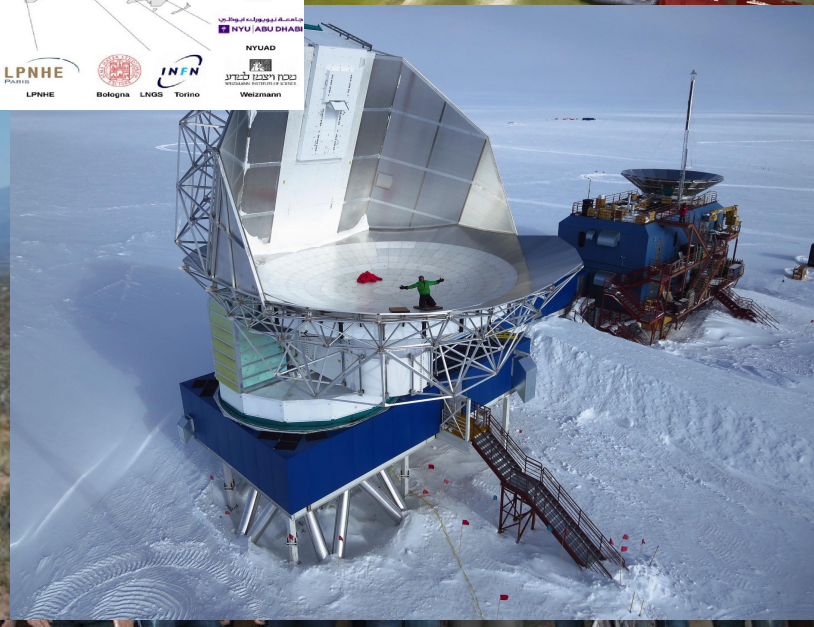
USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

CI for Big Science as a Team Sport

Frank Würthwein
Director
San Diego Supercomputer Center
March 1st , 2022

Big Science is a Team Sport





~100 members, 20 institutions

24 non-affiliated members

+35 associate members

Smithsonian Astrophysical Observatory

Adler Planetarium

Argonne National Lab

Barnard College / Columbia University

Bartol Research Institute / University of Delaware

Georgia Institute of Technology

Iowa State University

Purdue University

University of California, Los Angeles

University of California, Santa Cruz

University of Chicago

University of Iowa

University of Minnesota

University of Utah

Washington University in St. Loui

McGill University, Montreal

University College Dublin

Cork Institute of Technology

Galway-Mayo Institute of Techno

National University of Ireland, Galway

Argonne NATIONAL LABORATORY

BERKELEY LAB

NIST

Fermilab

KICP

Berkeley University of California

Case CASE WESTERN RESERVE UNIVERSITY

McGill Colorado University of Colorado at Boulder

UCDAVIS UNIVERSITY OF CALIFORNI

Kavli Institute for Cosmological Physics AT THE UNIVERSITY OF CHICAGO

THE UNIVERSITY OF CHICAGO

LMU

Harvard-Smithsonian Center for Astrophysics

M

I

SLAC NATIONAL ACCELERATOR LABORATORY

NSF Major Facilities benefit from an open CI that enables their communities to contribute in kind resources.

Open Science

Open Data

Open Source

Open Infrastructure

- Open Compute
- Open Storage & CDN
- Open devices/instruments/IoT, …?

Openness for an Open Society

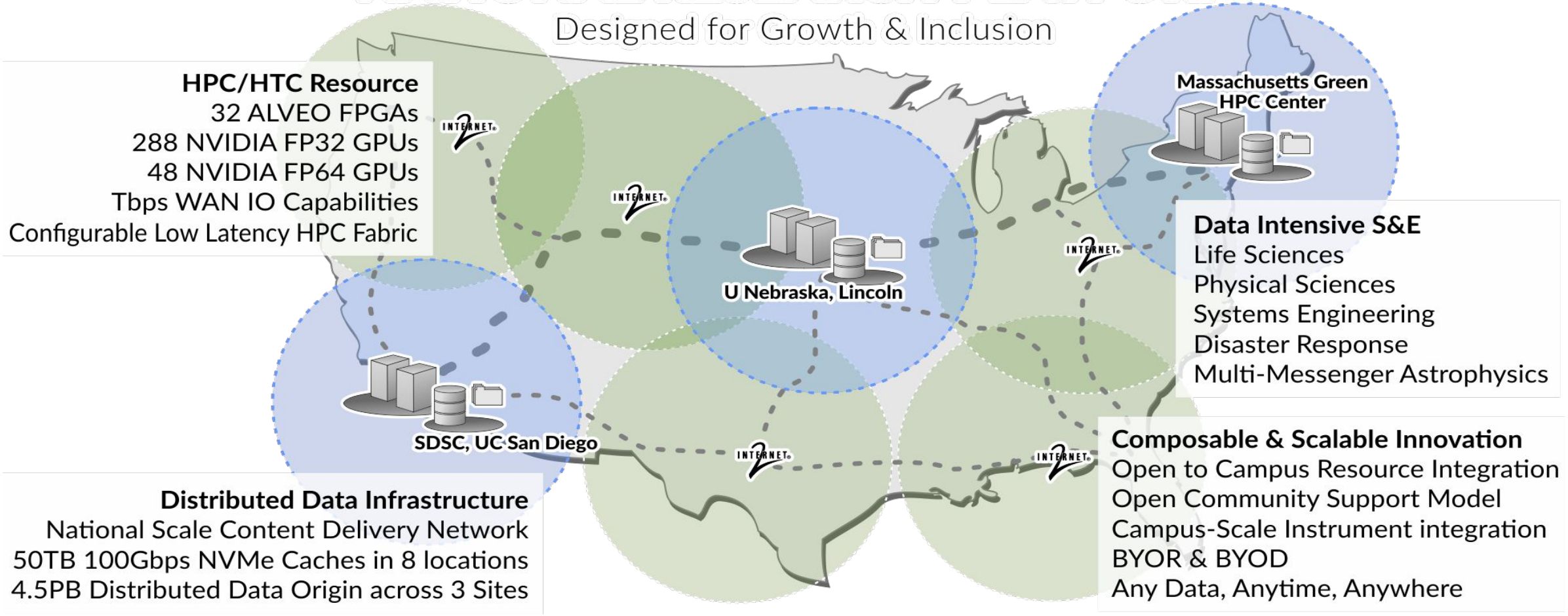
3 Ways to build Open CI Federations

- **At the container orchestration layer, e.g. K8S as implemented in National Research Platform.**
- Federating independent container orchestration frameworks, e.g. via SLATE or via Admiralty on K8S.
- **Federating storage and compute clusters at the storage and batch system layer., e.g. via OSG and PATh.**

Institutions choose the integration layer
based on the Cybersecurity controls they want to have.
Lower layer of integration => less effort to join & operate.

NATIONAL RESEARCH PLATFORM

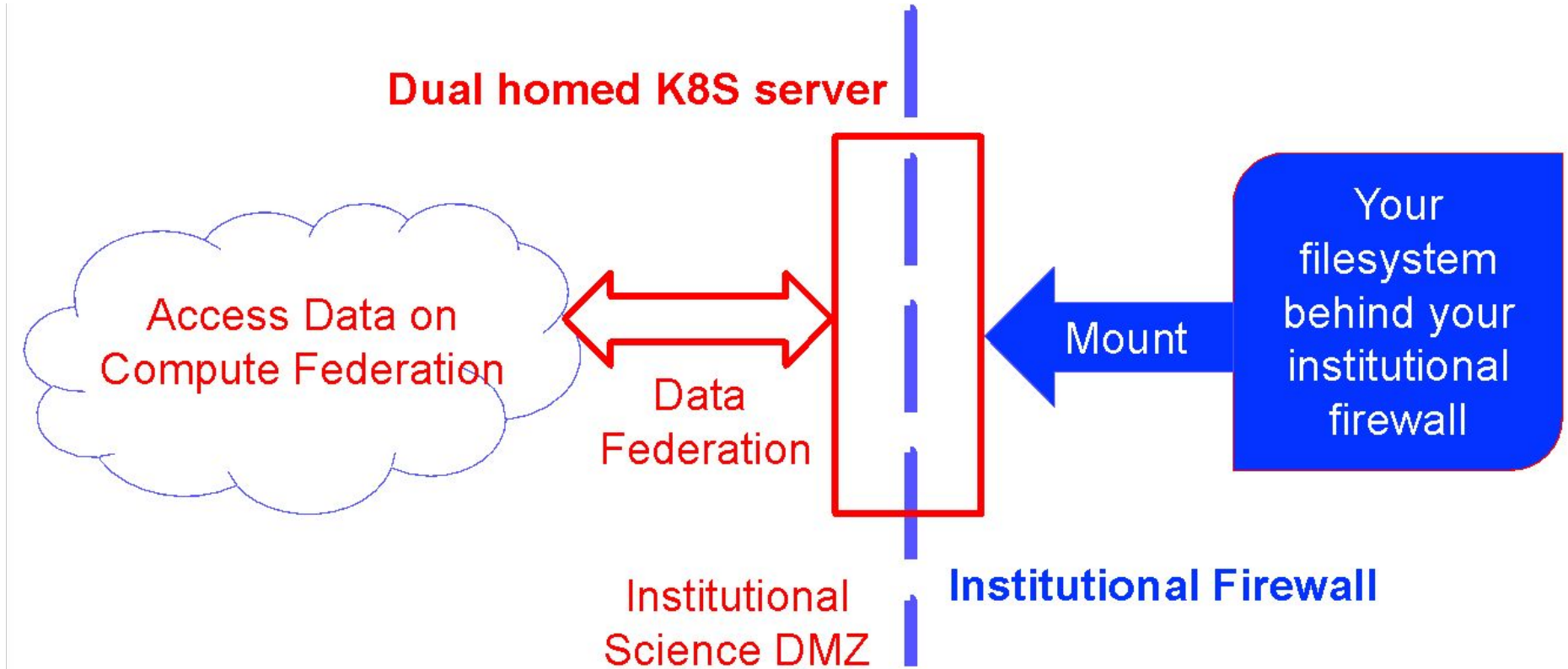
Designed for Growth & Inclusion



NSF 2112167

- Regional Ceph storage systems.
  - Researchers control replication of their data across regions.
- Content Delivery Network with Caches to cover the entire USA at less than 500 miles

Integration of an Institutional Data Origin



Your fileserver with your data is behind your institutional firewall.
A dual homed K8S server mounts only those filesystems you want to export.
We operate the Data Federation API by deploying container into K8S.

Federation Model … Goals for 2022

- Support **regional Ceph storage systems** across the USA.
  - Campuses can join individual storage hosts to the Ceph system in their region.
  - All regional storage systems are Origins in Data Federation
  - Deploy replication system such that researchers can decide what part of their namespace should be in which regional storage.
- Deploy **caches in Internet2 backbone** such that campuses in USA are within 500 miles of a cache.
- We have enough hardware funded via multiple different NSF projects to deploy nationwide in 2022.
- We have the software stack in production to allow any campus in the USA to join the open infrastructure.
- **NRP is fully interoperable with OSG & PATh**
  - Any compute resource in NRP is accessible from OSG submission points.
  - Any data resource in OSG and NRP is accessible from any compute resource on either OSG or NRP

Thoughts on Access

- "Free" as in the Open Science Pool operated by PATh
  - Any student/researcher working with any US based faculty/researcher/project can get an account and "hunting license" to use the fair-share access to federated resources.
- In-kind resources committed by your collaborators
  - Resources dedicated to your major facility based on a handshake or MoU
- Credit received as part of an NSF Award or supplement.
  - The NSF program director plays a significant role here.
- Allocation received as part of a national allocation process
  - requires writing an allocation proposal that is reviewed a few times per year.

Open Infrastructure provides plumbing to exercise any of the above.

Acknowledgements

- We gratefully acknowledge the NSF support via NSF 2112167, 2030508, 1841530, 2120019, 2100237, 1541349, 1540112.
- Apologies for any awards I might have missed ….