

# Office of Cyberinfrastructure Update

NSF Advisory Cyberinfrastructure Committee

December 12, 2012

## 3 “new” topics

- Research Data Alliance (RDA)
- International Research Network Backbone concept
- National Data Infrastructure

# Biggest Threats to Establishing a Global Data Research Infrastructure

- Not understanding the critical importance and the need to share data for next century science and education
- Relying on additional workshops, conferences, committees and so forth to study and provide more recommendations
- Waiting for standards to be approved that will enable data sharing, interoperability, and support the entire data life cycle

Just do it

# RDA Purpose

Accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be achieved through the development and adoption of infrastructure, policy, practice, standards, and other deliverables.

# Research Data Alliance (RDA)

- IETF model focused on data sharing and exchange
- “Rough consensus and early deployment”
- Create and adopt focused “chunks” of code, policy, infrastructure, standards, or best practices that enable data to be shared and exchanged
- Efforts that have substantive applicability to some segments of the research community (but may not apply to everyone at the outset)
- Involvement of working scientists and data researchers who can start sharing and exchanging data today, while other long-term or far-reaching solutions are being appropriately discussed and explored in other venues

# Initial PIs funded to support RDA

## US:

Fran Berman – RPI

Bill Michener – DataOne

Beth Plale – Indiana University

Sayed Choudhury – Johns Hopkins University

## Australia:

Ross Wilkinson – Australia National Data Service

Andrew Treolar - Australia National Data Service

## Europe:

Leif Laaksonen (iCORDI/CSC)

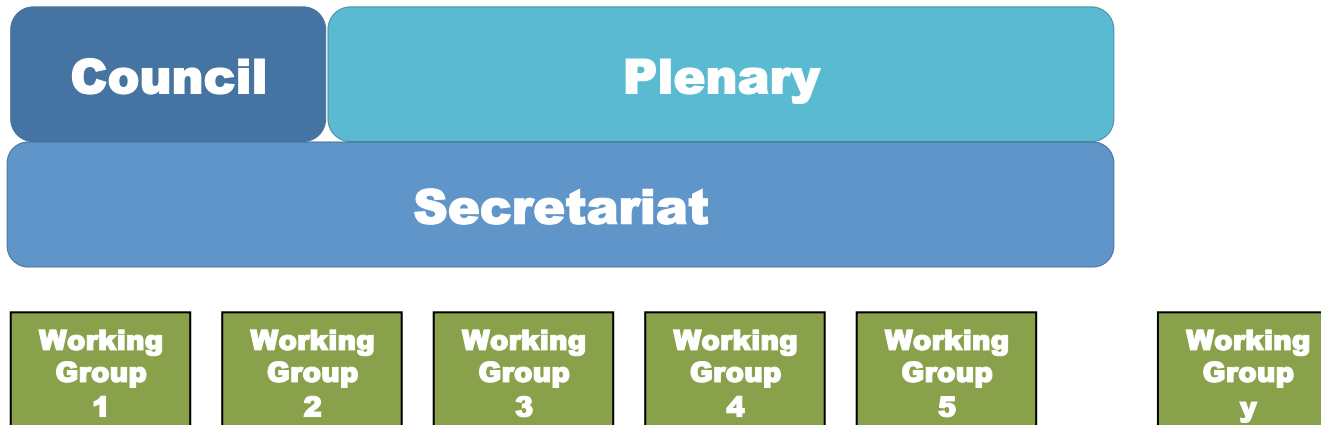
Peter Wittenburg (iCORDI/Max Planck Institute)

Juan Bicarregui (iCORDI/STFC)

# RDA Working Principles

- Openness – Membership is open to all interested individuals and organizations, all meetings are public, RDA processes are transparent, and all RDA work products are freely available to the public;
- Consensus – The RDA moves forward by achieving consensus and resolving disagreements through appropriate mechanisms;
- Balance – The RDA is organized on the principle of balanced representation for individual organizations and stakeholder communities;
- Harmonization – The RDA works to achieve harmonization across standards, policies, technologies, tools, and other data infrastructure elements;
- Voluntary – The RDA is a public body responsive to its members. It is not a government organization or regulatory body and, instead,; and
- Non-profit - RDA is not a commercial organization and will not design, promote, endorse, or sell commercial products, technologies, or services.

# Research Data Alliance





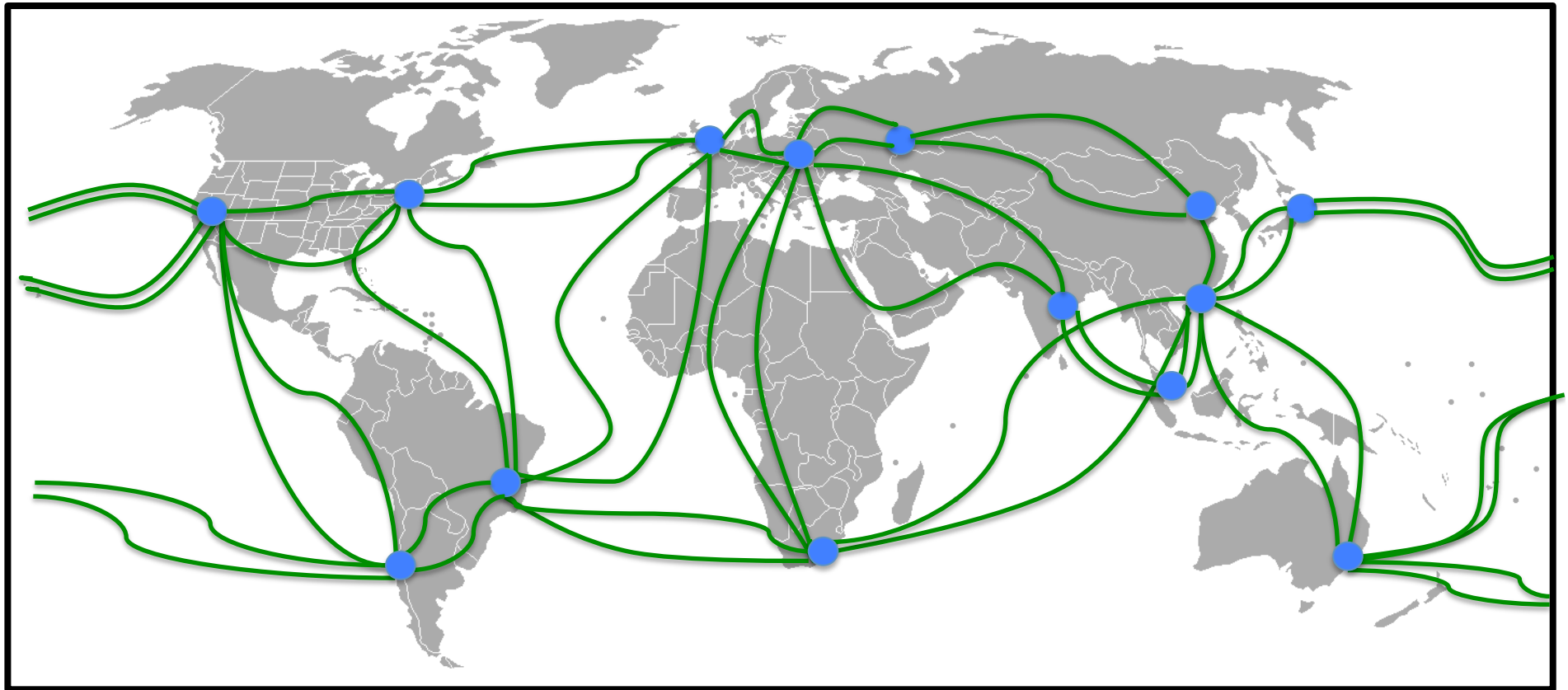
# RDA Timeline

- White paper- Apr 2012
- RDA projects start - Sep 2012
- Multiple collaboration teams start to meet weekly- Oct 2012
- RDA Steering group established- Oct 2012
- Convening workshops
  - Oct 1-3 (Washington DC)
  - Oct 23-24 (Barcelona, Spain)
- 1<sup>st</sup> Plenary – Mar 18-20, 2013 (Gothenberg, Sweden)
- 2<sup>nd</sup> Plenary – Oct 2013 (Washington DC)

# International Research Network Backbone Concept

## Notational only

Architecture, aggregation nodes, locations, bandwidth, connectivity to be determined



- Aggregation nodes: Primary connection to International Backbone; every country and economy has the opportunity to be an aggregation node or connect to one
- 🟢 Shared High Bandwidth network (multi 10/100 Gig) interconnecting Aggregation nodes with redundant links

# Global Network Guiding Principles

- Global coordination, planning, rationalization, governance and coordinated investments (e.g. simultaneous calls, joint calls)
- Open exchange points: “policy-free” operation of open exchange points allowing for bi-lateral peering at all layers, open transit
- Matching Investments: investments will be matched among parties such that relationship between operators will not be hampered by asymmetry.
- End-to-end interoperability to ensure end-to-end visibility and interoperability across appropriated links and paths
- Development of a coordinated strategy by key scientific funding agencies to meet the international networking needs of science, engineering and education for the next 10 years.

# Coordinating Cyberinfrastructure in multi-user facilities

- Goal: coordination of cyberinfrastructure development between two or more NSF-supported scientific facilities as a supplement to their existing award
- 7 supplement awards in August 2012 (BIO, ENG, GEO, MPS, OPP)
- Examples
  - Software tools for researchers working in synthetic biology
  - Common data pipeline between multiple high energy physics facilities to analyze data gathered about neutrinos
  - Joint informatics education software and curriculum for BIO-supported centers.

# Building a National Data Infrastructure

- The data infrastructure will be complex and involve a range of modalities
  - Multiple data centers, clouds, distributed systems, replication
  - Partnerships between campuses, government, business
- Leveraging and building on the myriad of data efforts and projects underway and being planned
- New focus on curation, interoperability, sharing of data, common approaches and data policies
- New sustainability models for data stewardship will emerge, driven by the needs of individual research communities

**Questions?**