

History & Theory of Infrastructures

Robust cyberinfrastructure will develop only when social, organizational, and cultural issues are resolved in tandem with the creation of technology-based services.



Report of Workshop on
"History & Theory of Infrastructure: Lessons for the New Scientific Cyberinfrastructures"
(P. N. Edwards et al., January 2007)

Dynamics of Infrastructure Development

From P. N. Edwards et al., January 2007

1. System Building

- Deliberate and successful design of 'local' systems that deliver technology-based services.
- Combination of technical, organizational, and social capabilities is crucial - "System building typically begins as a social act."

2. Growth

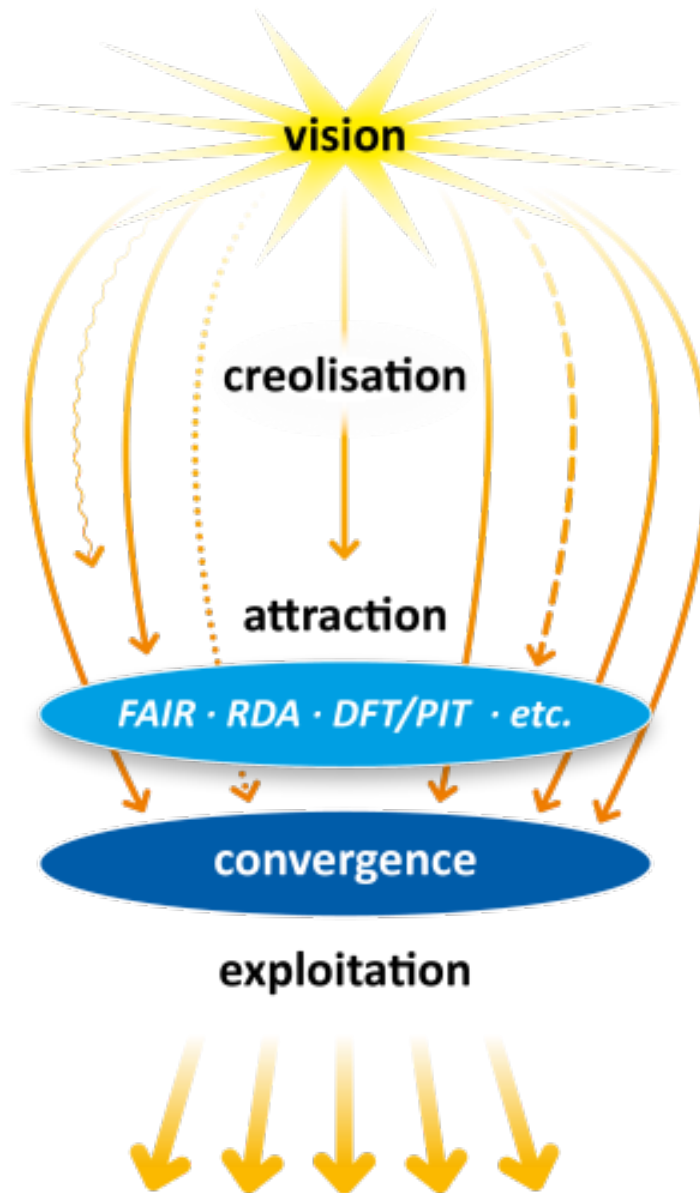
- Technology transfer across domains and locations
- Results in variations on the original design, as well as emergence of competing systems

3. Consolidation

- Infrastructures formation via gateways that allow dissimilar systems to be linked into networks
- Standardization and inter-organizational communication techniques are critical.

Gateways = technologies and standards applied across multiple communities of practice

From: "People – the Ultimate Challenge for Geoinformatics" Presentation at the NSF/EAR Geoinformatics Workshop in Denver, CO, March 2007



Take Aways

- Foundation: disciplinary data repositories
 - Provide the “I & R” through community engagement
 - Keep the landscape manageable, don’t keep re-inventing the wheel & fragmenting
 - Dedicated support for coordination and collaboration, rethink funding structures
- Co-locating data and computing resources
- Governance
 - requires effort and needs to be supported
 - Current funding structures
- Learn from other efforts (including ‘Epic Failures’!!)

Breakout session 2:

Vision for the Future CDF

- How could the CDF become a sustainable organization?
- What are potential organizational models for the CDF of the future? (e.g. consortium of data facilities)
- Can you identify any transition opportunities for the CDF? (e.g. NSF solicitation for Synthesis Center)

Sharing Infrastructure among Domain Repositories:

A Vision and Way Forward of the Council of Data Facilities



Kerstin Lehnert*, *Columbia University (IEDA)*
Jerry Carter*, *Incorp. Res. Inst. for Seismology*
Robert Downs*, *Columbia University (CIESIN)*
Corinna Gries*, *University of Wisconsin-Madison (EDI)*
Daniel S. Katz, *University of Illinois (NCSA)*
Steve Diggs*, *UC San Diego/Scripps (CCHDO)*
Mohan Ramamurthy, *UCAR (Unidata)*

* Members of the CDF Executive Committee

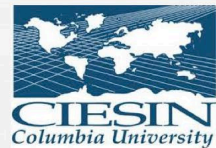
CDF advances coordination among Earth, Space, & Environmental data facilities to enable integrative science.

CDF promotes the unique value of domain-specific data services to achieve truly FAIR data.

CDF addresses common challenges of data facilities to provide scalable, trustworthy, & sustainable services.

<https://www.earthcube.org/council-of-data-facilities>

CDF was founded in 2014 through NSF's EarthCube program, which ends in 2022.



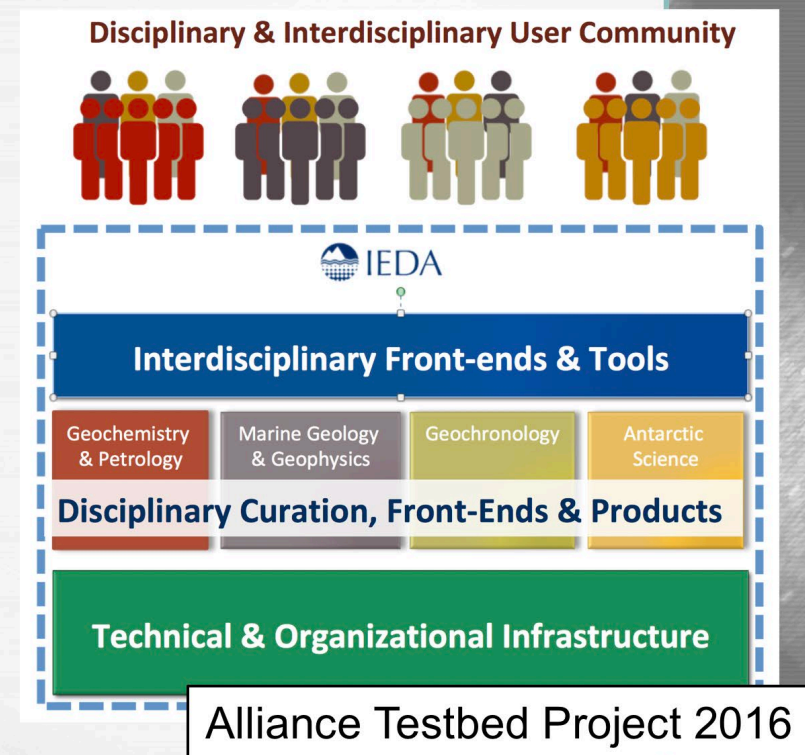
... and many more



CDF Shared Infrastructure: Concept



- Operate a **collective infrastructure** that co-locates computational and storage resources with “Cloud” and High-Performance Computing (HPC) environment.
 - Different levels of service (IaaS, PaaS, SaaS)
- Focus data facility resources on domain-specific data curation following **common best practices**.
- Design, develop, and adopt a more mature **culture of coordination** and shared resources among data facilities.



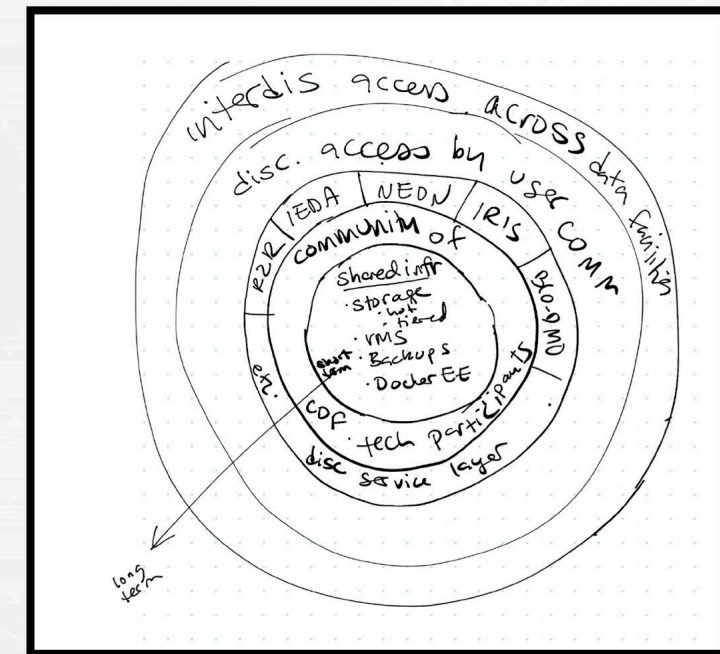
“allow repositories to retain autonomy while benefiting from collective infrastructure”

From CDF Shared Infrastructure Initiative White Paper 2019
<https://www.earthcube.org/council-of-data-facilities>

CDF Shared Infrastructure Value Proposition



- Will help research scientists **do interdisciplinary & innovative science** by easily accessing data and services across repositories and using big data and HPC approaches.
- Will help data facilities **become more efficient, resilient, trustworthy, sustainable,** and convergence on **consistent best practices.**
- Will help funding agencies **accelerate scientific discovery** by supporting data resources to be located in close proximity but curated and managed by sustainable domain repositories.



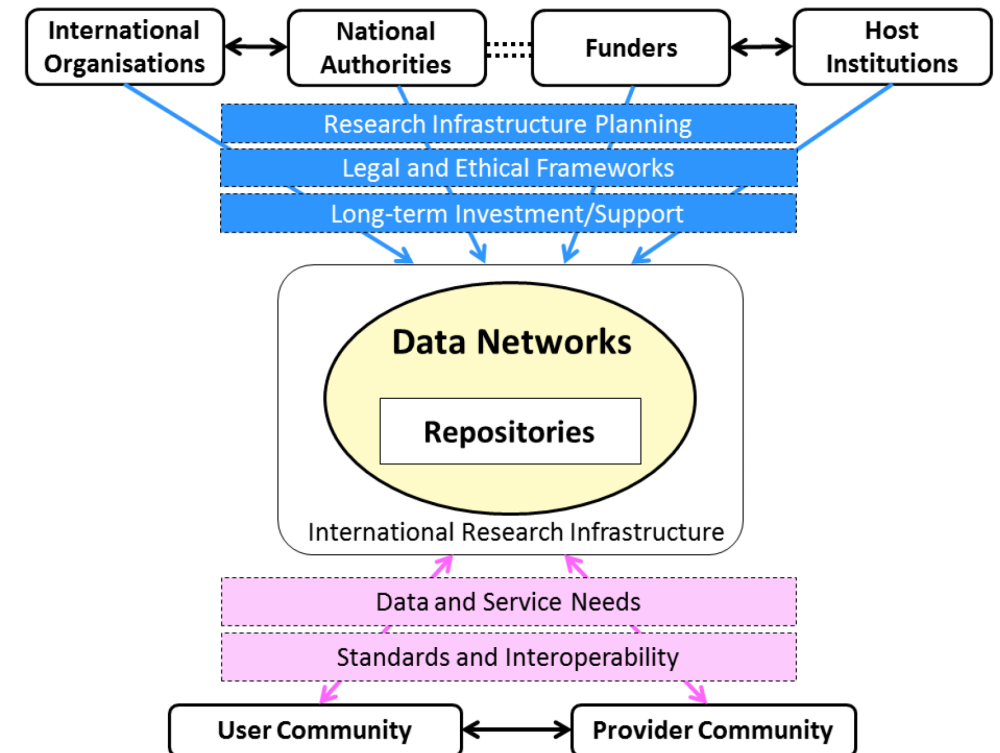
Napkin drawing from
CDF Workshop in July 2019

ESS Data Are Global

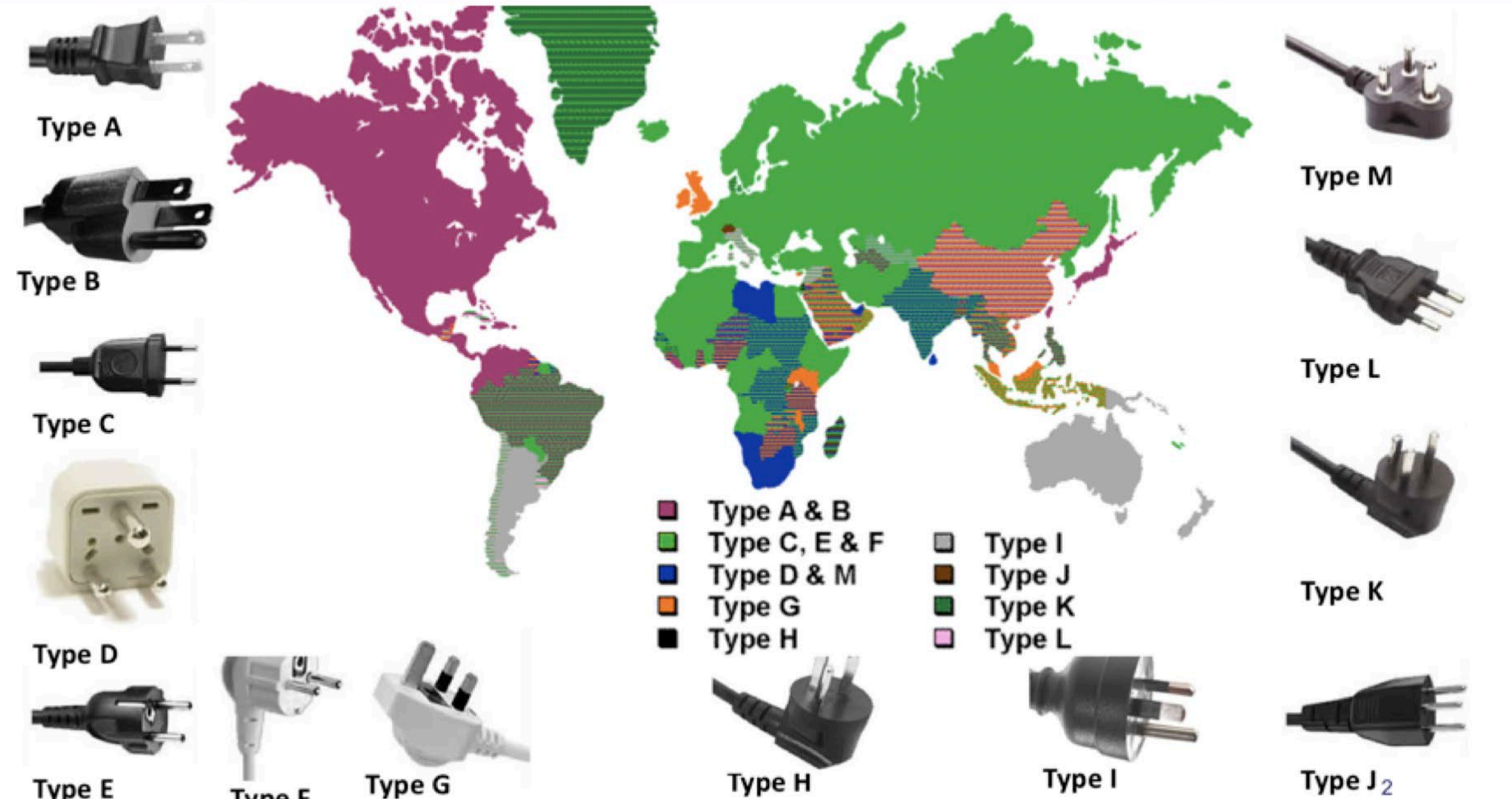
- Joint development, adoption, and promotion of standards that ensure meaningful aggregation, analysis, and interpretation.
- **National engagement and governance** to ensure responsiveness to the needs of specific communities.
- Implementation of **national policies** to ensure compliance.
- Long-term commitment for sustainability.

CO-ORDINATION AND SUPPORT OF INTERNATIONAL RESEARCH DATA NETWORKS

OECD SCIENCE, TECHNOLOGY
AND INNOVATION
POLICY PAPERS
December 2017 No. 51



Where are the community agreed standards for Interfacing?



From: “First steps towards internationally integrating data and services in the solid Earth sciences and beyond”

by L. Wyborn, B. Evans, K. Lehnert, T. Rawling, J. Klump, K. Elger, S. Cox, H. Glaves, M. Ramamurthy, E. Robinson, S. Stall; EGU General Assembly 2018

COUNCIL OF DATA FACILITIES



The Council of Data Facilities (CDF) is a federation of existing and emerging **geoscience data facilities** that serve as a foundation for EarthCube and related aspects of the cyberinfrastructure for earth system science.

- Founded 2014 as an EarthCube Governance entity
- Currently 42 members
 - **28 in Category A** (*NSF-funded not-for-profit or academic data facilities*)
 - **9 in Category B** (*Federally Funded Research and Development Centers (FFRDCs) and other federal, state, and local data facilities*)
 - **2 in Category C** (*International, private, and other not-for-profit or academic data facilities*)
 - **3 in Category D** (*Associate Members*)