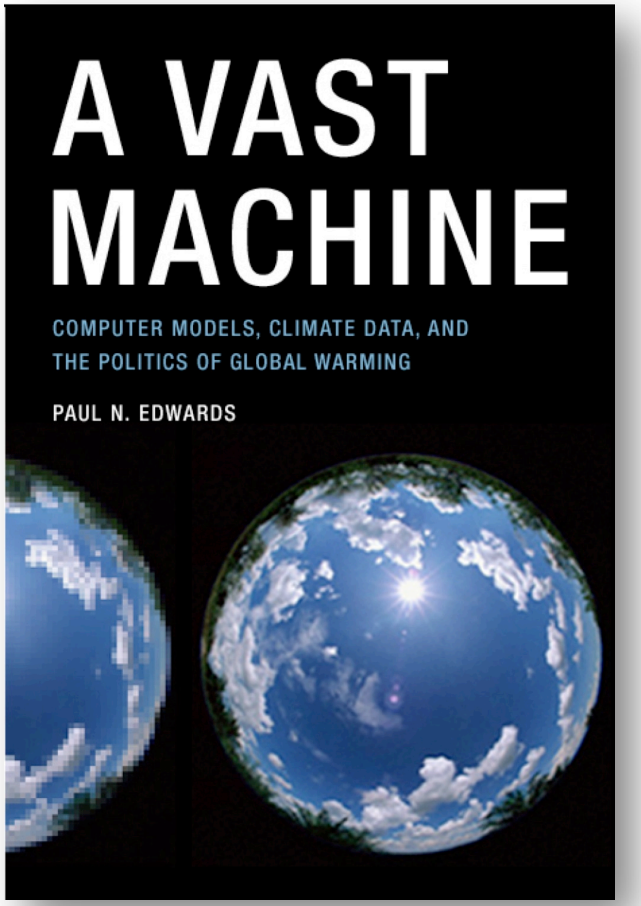
The history of infrastructures and the future of cyberinfrastructure in the Earth system sciences

Paul N. Edwards | SciencesPo, Paris (2012-13)
School of Information and Dept. of History, University of Michigan

"The Meteorological Society …wishes to be ***the central point, the moving power, of a vast machine, and it feels that unless it can be this, it must be powerless; if it cannot do all it can do nothing.***

It desires to have at its command, at stated periods, ***perfect systems of methodical and simultaneous observations****; it wishes its influence and its power to be omnipresent over the globe so that it may be able to know, at any given instant, the state of the atmosphere on every point on its surface…"*

— John Ruskin, 1839



Outline

- A historical model of infrastructure development
- Applying the model:
  - The world weather forecast network
  - Climate data
  - Climate modeling
- Some lessons from history for future infrastructure development

Infrastructure: a historical model

time

- System building (Hughes 1983)
  - Designed, coherent, centrally organized
- System variations proliferate
- Networks
  - Dedicated gateways link heterogeneous systems
- Internetworks
  - Generic gateways link heterogeneous networks
- Decentralization, fragmentation, service tiering


Understanding Infrastructure: Dynamics, Tensions, and Design

Report of a Workshop on "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures"

Paul N. Edwards
Steven J. Jackson
Geoffrey C. Bowker
Cory P. Knobel

January 2007

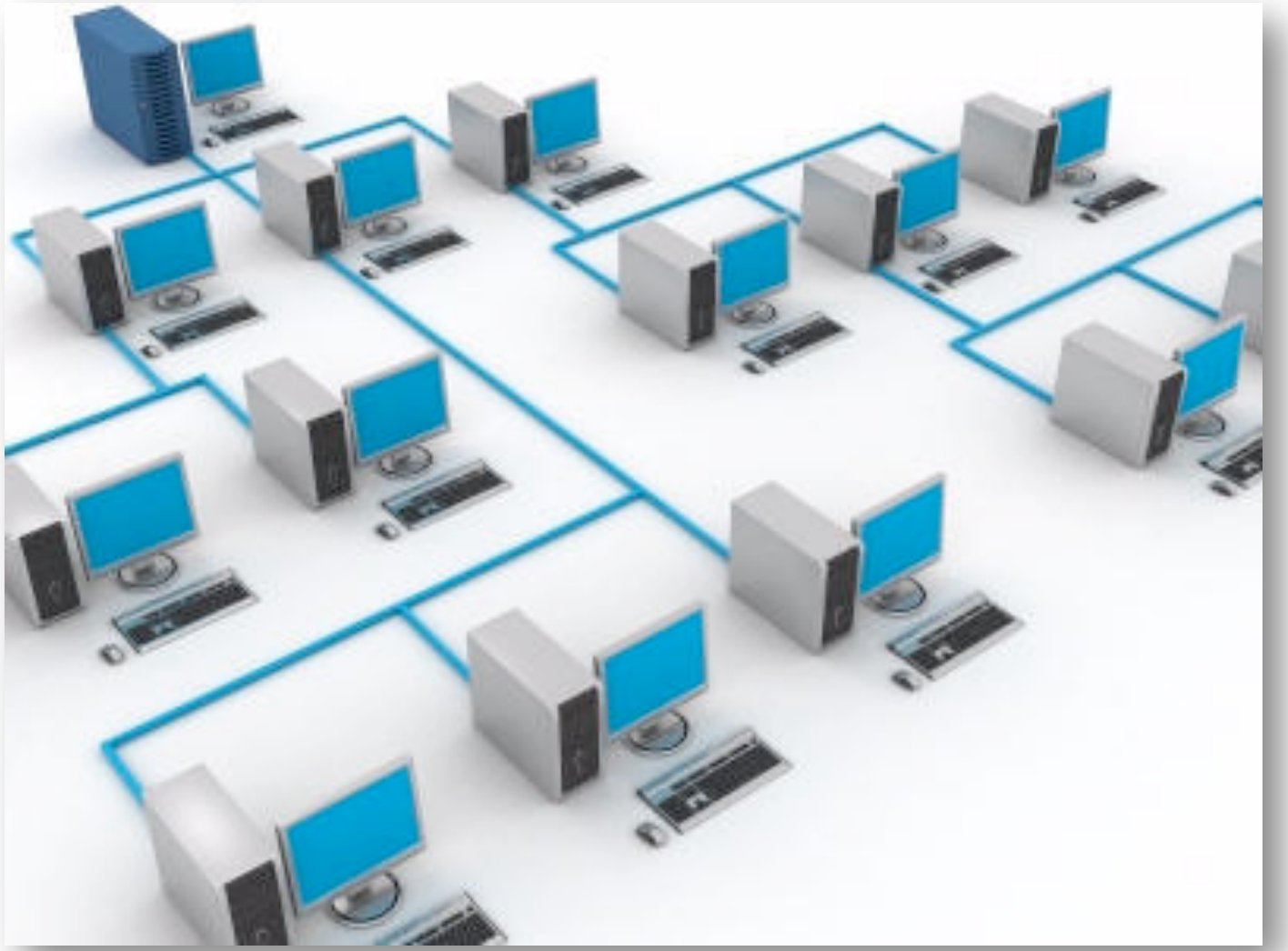
NSF Grant 0630263 • Human and Social Dynamics • Computer and Information Science and Engineering • Office of Cyberinfrastructure

Edwards et al. 2007

Dedicated or “improvised” gateways

- Connecting devices to different national power grids
  - Whose responsibility?
  - Who pays?
  - Who sets standards?

Computer networks link computers

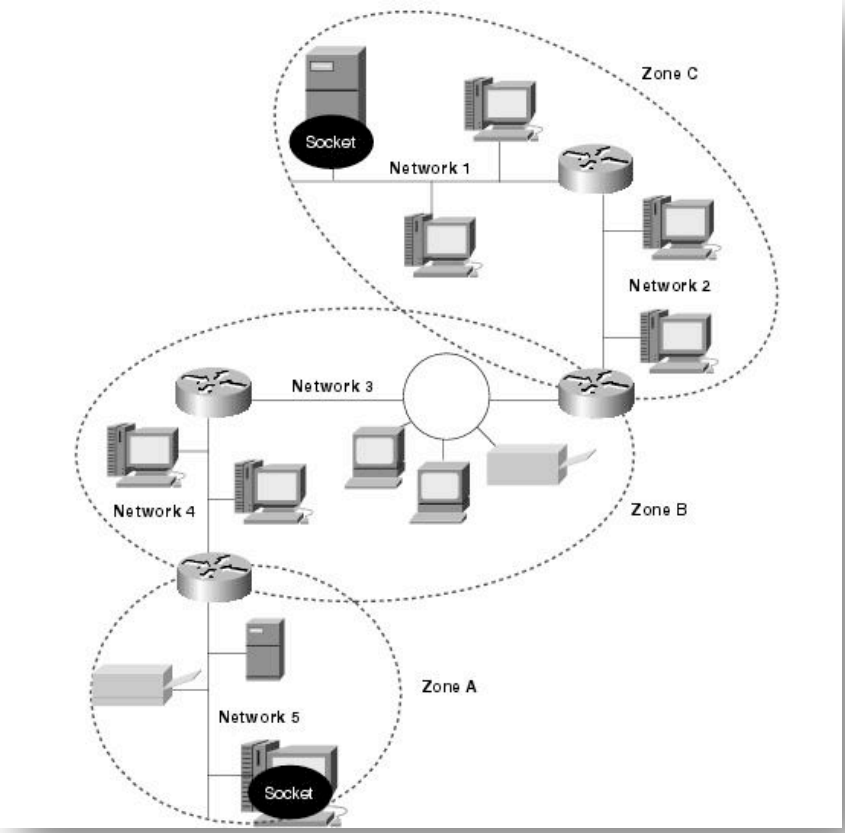


A generic gateway

- Global ISO standard
- Links road, rail, and shipping

Internetworks link networks

- Routers are gateways
  - connect computers to each other (network)
  - … and connect the local network to other networks
- "The" Internet connects millions of networks
- Protocols and standards permit this to work



Infrastructures are networks or internetworks (not systems) (Edwards et al. 2007)

	Systems	Infrastructures	
		Networks	***Internetworks or Webs***
Elements	Components, subsystems	Heterogeneous systems	Heterogeneous networks
Gateways and standards	Dedicated or improvised	Dedicated or generic	Generic
Boundaries	Closed, stable	Closed or open, reconfigurable	Open, reconfigurable
Examples	State electric company	National electric grid	International electric grid
	Mainframe computer	Network of similar personal computers	Internet, World Wide Web
	Telegraph-based synoptic weather mapping **Réseau Mondial**	**National weather services (some elements)**	**World Weather Watch** **GEOSS (Global Earth Observing System of Systems)**

Applying the historical model
Weather forecasting

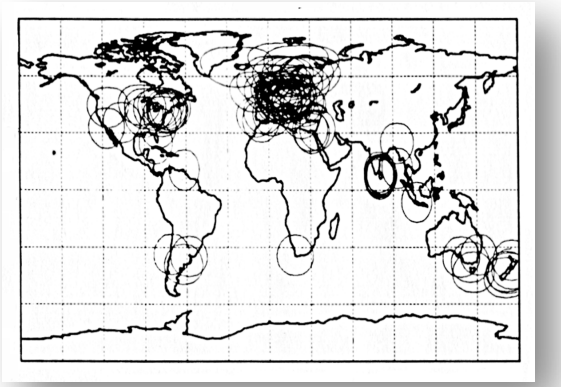
- Systems
  - National weather services
    - Telegraph-based synoptic weather mapping
  - Marine weather records
    - Semi-standardized reporting forms (1850s forward)
- Networks
  - Early regional telegraph networks (from 1850s)
- Internetworks
  - Both within and across national networks, integrating:
    - Surface stations
    - Air bases and airports
    - Marine data
    - Satellites (NASA, EUMETSAT, etc.)

1872 War Dept. synoptic map

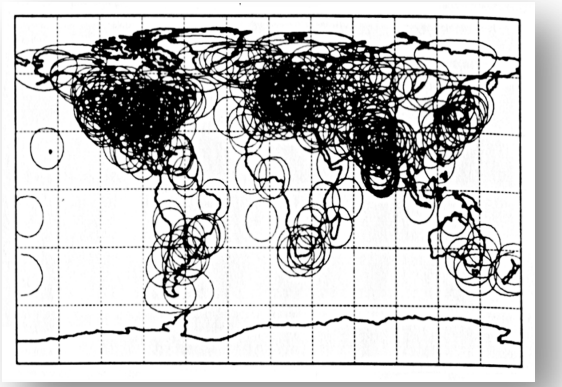
Paul N. Edwards

Surface station coverage: evolution

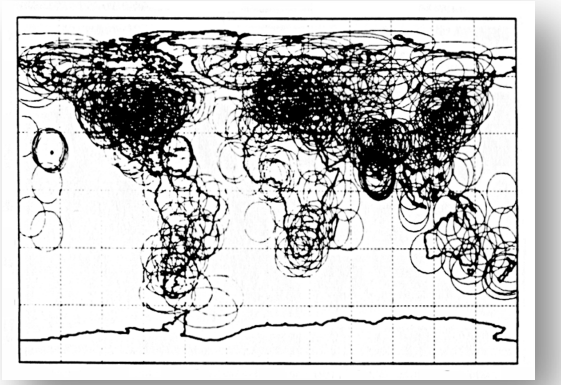
1870



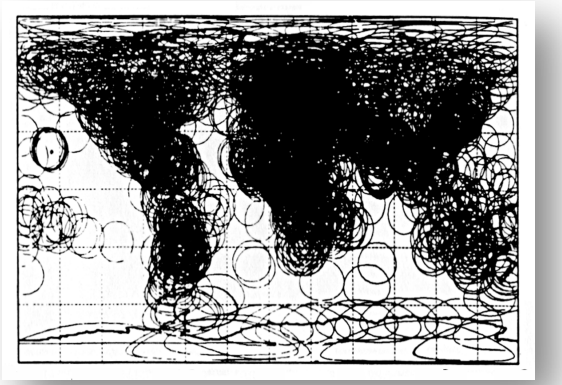
1900



1930



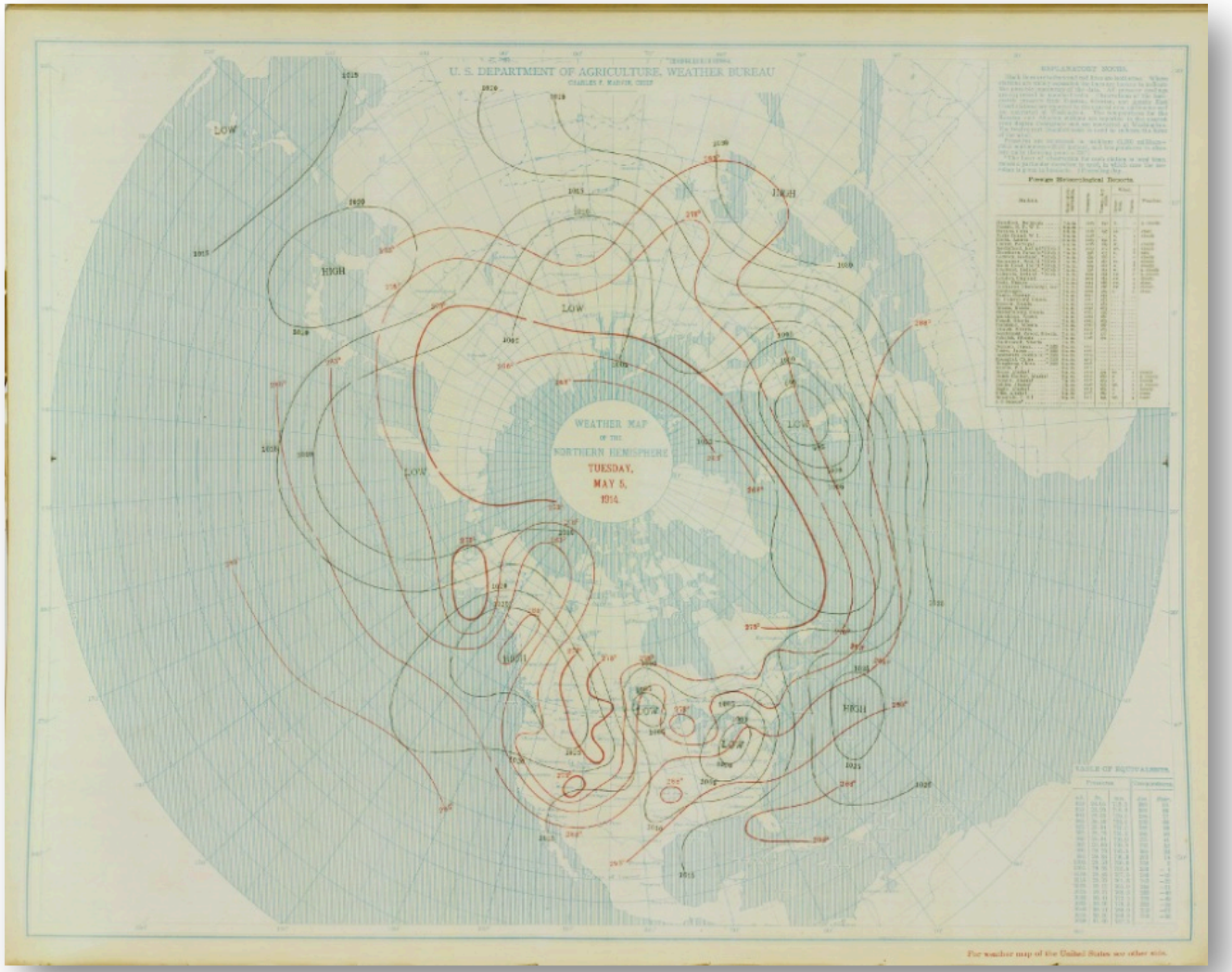
1960



Source: J. Hansen and S. Lebedeff, "Global Trends of Measured Surface Air Temperature," *Journal of Geophysical Research* 92, no. D11 (1987), 13,346-13,347. Circles drawn around each station is 1200 km,

International networks: gateway institutions

- International Meteorological Organization (1879-1947)
  - Non-governmental
  - Set common standards, but lacked power to enforce
- Established principle of free data sharing
- World Meteorological Organization (1947-present)
  - Intergovernmental
  - Encouraged network development
  - Regional meteorological centers



Synoptic map of the northern hemisphere, 5 may 1914

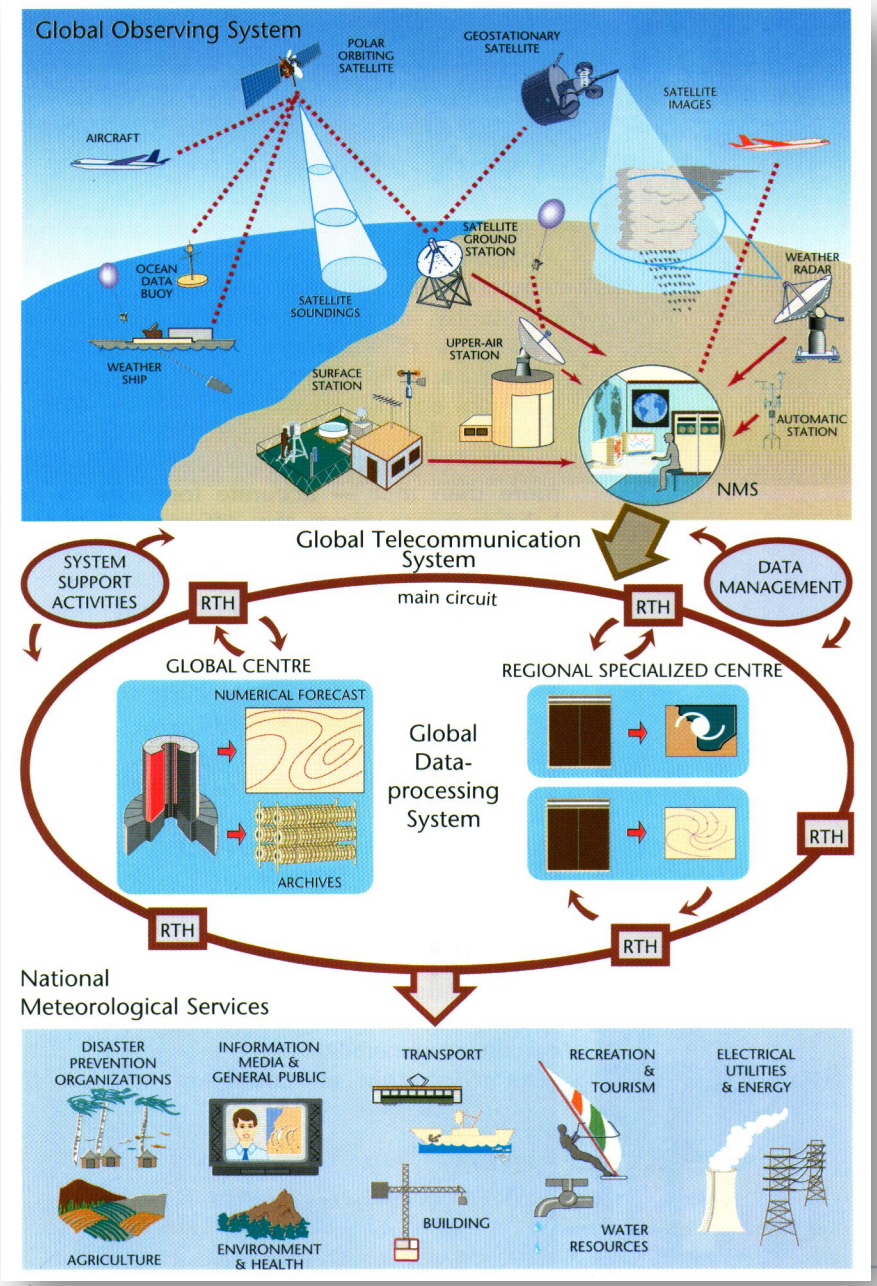
Paul N. Edwards

Gateways between weather data systems

- Many conversions
  - Temperature (°F vs. °C), pressure (Pa vs. bar), etc.
  - Reporting forms
  - Telegraph codes
  - Recording media
    - Telegraph, paper, punch cards, paper tape, digital tape, magnetic disk, RAM
  - Codes and communication networks
    - Telegraph, telex, fax, shortwave radio, telephone….
- Most required human intervention
- Today, computers have automated most of these conversions
  - …but new formats and conflicting standards still require human intervention

More gateways…

- Between multiple institutions within a nation…
  - NASA (satellites)
  - National Weather Service
  - Air Force observing systems worldwide
- And between mulitiple international data networks
  - WMO regional meteorological centers managed their own communication networks
  - European Centre for Medium Range Weather Forecasts
    - European national weather services
    - EUMETSAT



an internetwork

World Weather Watch

- Planned early 1960s
- Operational 1968
- Modern form of the world's oldest and best developed global digital data infrastructure

Paul N. Edwards

Applying the historical model
Global climate data: systems and networks

- The Réseau Mondial
  - "Worldwide network"
  - Proposed 1905
  - Organized by British Met Office (1911-1950s)
    - ~ 500 stations on a 10° lat-long grid, 80°N to 61°S
- World Weather Records
  - Smithsonian Institution, 1927-1947
  - 380 stations, including islands
  - Later, became *Monthly Climatic Data for the World*
- Both efforts were central collectors
  - Hampered by variations in standards

Different climate data systems, at different times…

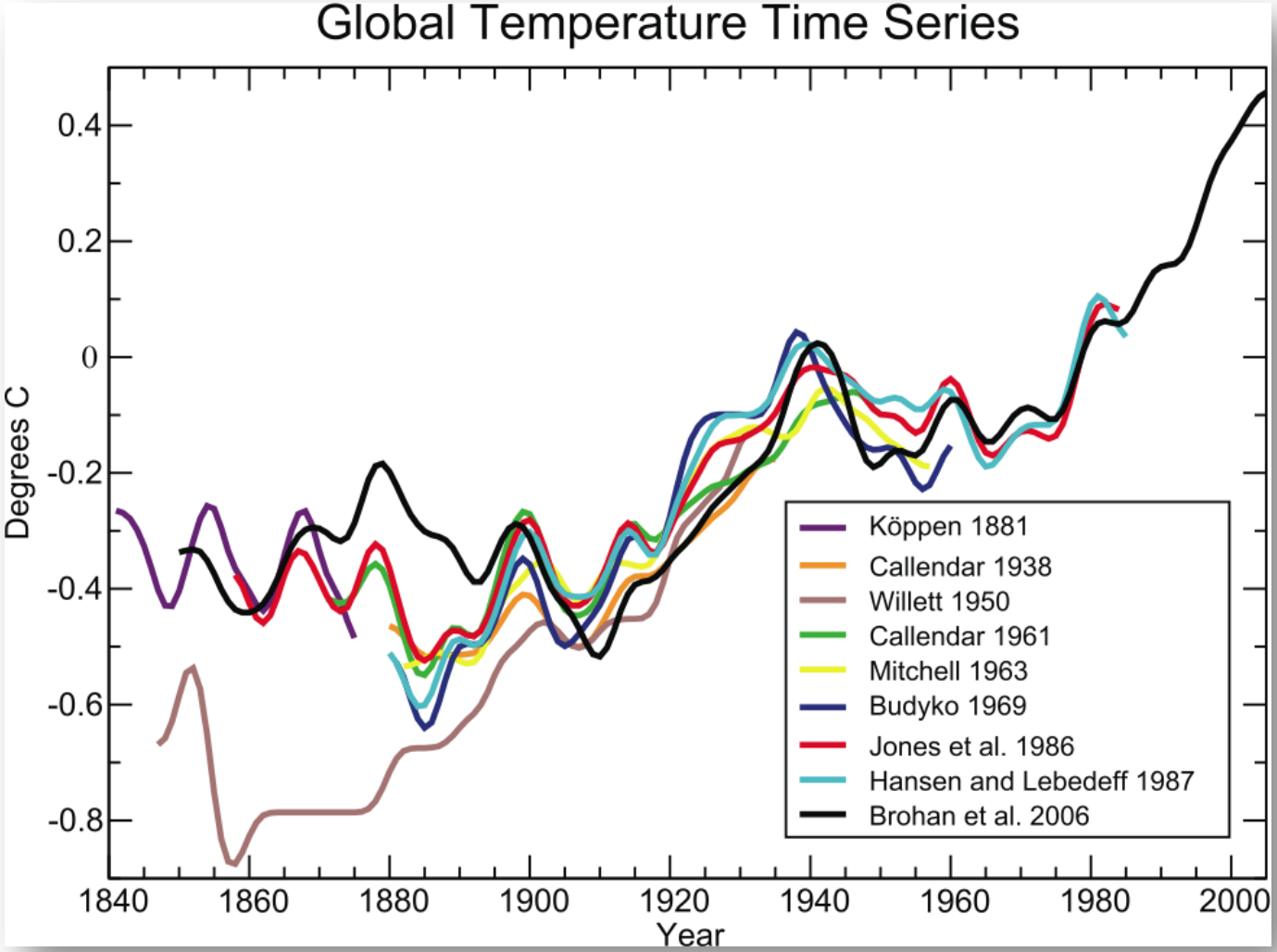
Country	*Methods used to calculate mean daily temperature*
Egypt	1/2(max + min); means of 3-hourly observations, 1/8(00 +03 + …21); 1/4(09 + 21 + max + min); 1/4(06*Z* + 12*Z* + 18*Z* + min); means of 24 hourly values (exact hours unknown)
France	1/2(max + min); 1/3(06 + 13 + 21); 1/3(06 + 14 + 22); means of 24 hours, 1/24(01*Z* + 02*Z* +… 24*Z*); means of 8 3-hourly observations
Ghana	1/8(03 + 06 + …24); 1/2(max + min)
Guyana	1/2(max + min); 1/12(00 + 02 + ...22); 1/3(07 + 13 + 18) local time; 1/2(12Z + 18Z).
Tunisia	Means of 24 hours (exact hours unknown); 1/2(max + min); 1/4(07 + 13 + 19 + (19 + min)/2)
U.S.S.R.	1/4(01 + 07 + 13 + 19); 1/3(07 + 13 + 21); 1/4(01 + 07 + 13 + 21); 1/4(07 + 14 + 21 + 21) 105°E meridian time; means of 2-4 daily observations in 53 different combinations

Source: Palutikof and Goddess, 1986

Gateways in climate data analysis

- "Data guys": historical knowledge used to clean and correct data
  - Station sites, instrument changes, etc.
- Data analysis models
  - Corrections for urban heating, instrument differences, etc.
  - Weighting data to correctly represent surface area
- Reanalysis models
  - Running historical weather data through a frozen weather analysis and forecast model
- Each of these combines heterogeneous national data sources to create homogeneous global data sets

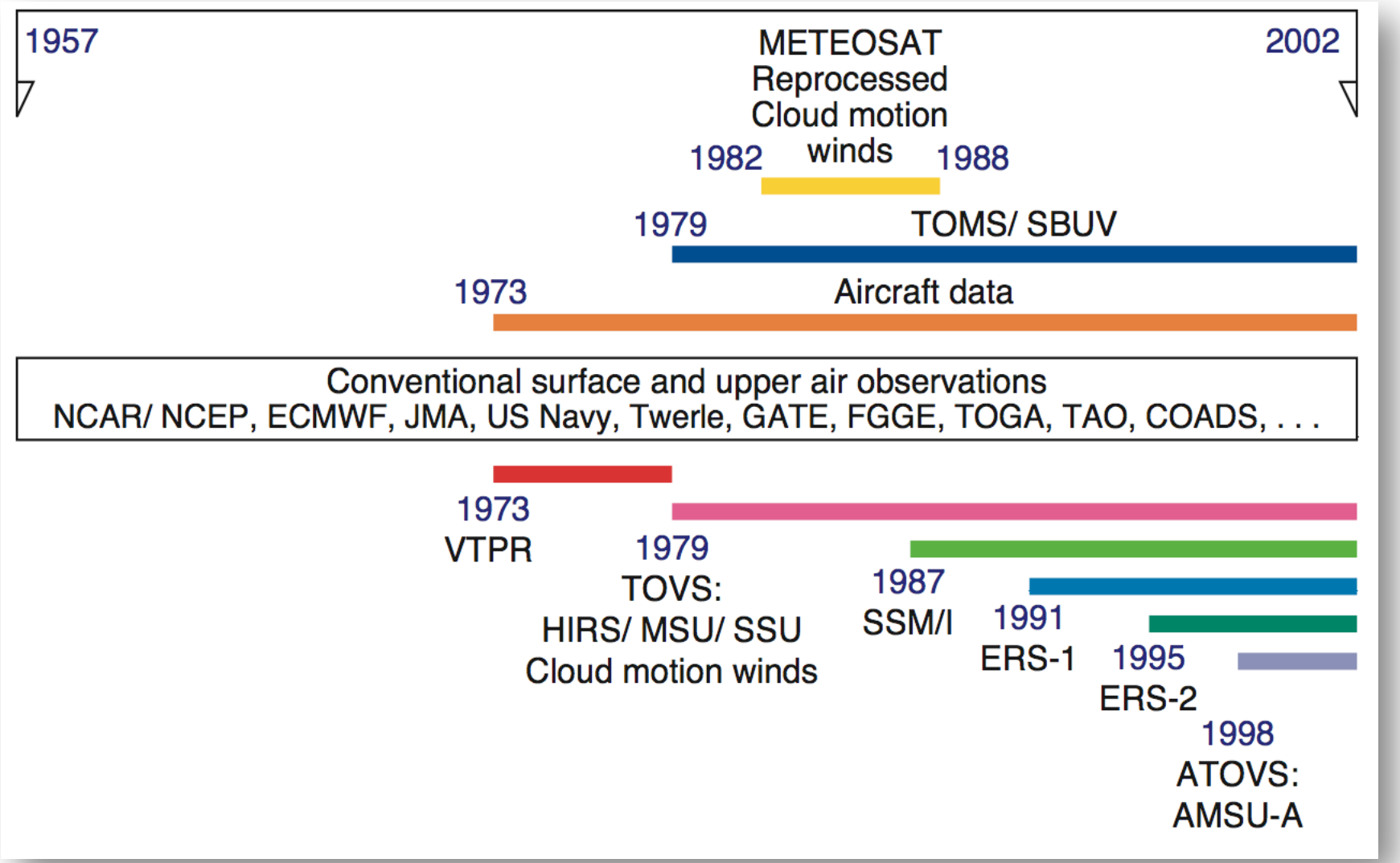
Global climate data are produced by an internetwork (linked, heterogeneous data networks)



IPCC 4th Assessment Report (2007), Fig. 1.3

Paul N. Edwards

A climate data internetwork: observing systems used in ERA-40 climate reanalysis



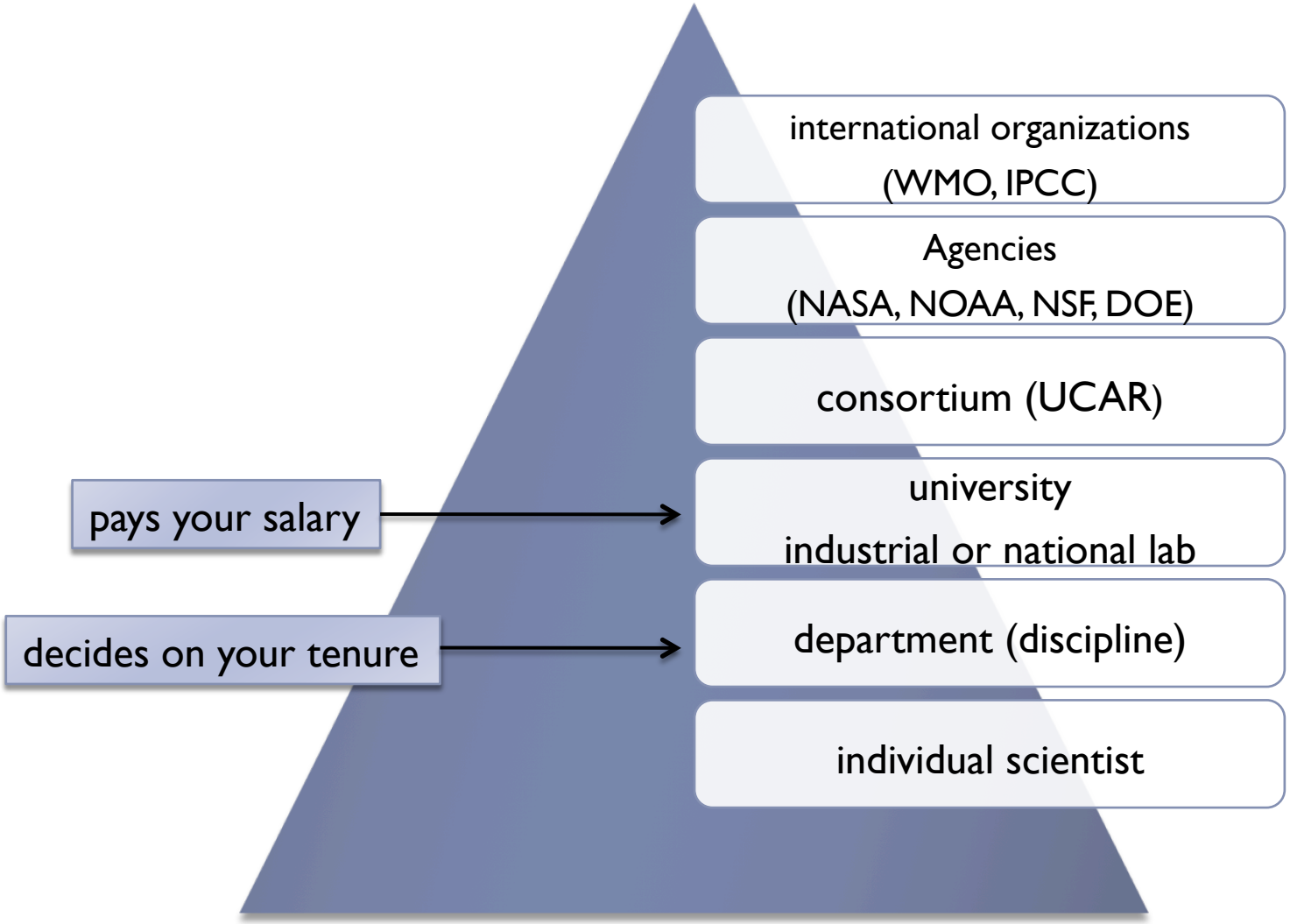
Other gateways in climate science

- Model couplers connect models
  - Model Coupling Toolkit (MCT)
  - Earth System Modeling Framework
- Model intercomparison projects
  - Standard exercises and benchmarks
  - AMIP (1988)
  - CMIP (Coupled Model Intercomparison Project) — ongoing
- Data portals
  - Earth System Grid
  - PCMDI (Program on Climate Model Diagnosis and Intercomparison)
  - GO-ESSP (Global Organization of Earth System Science Portals)
- IPCC
  - Connects multiple aspects of climate science
  - Creates/maintains global community
  - Connects science to policy

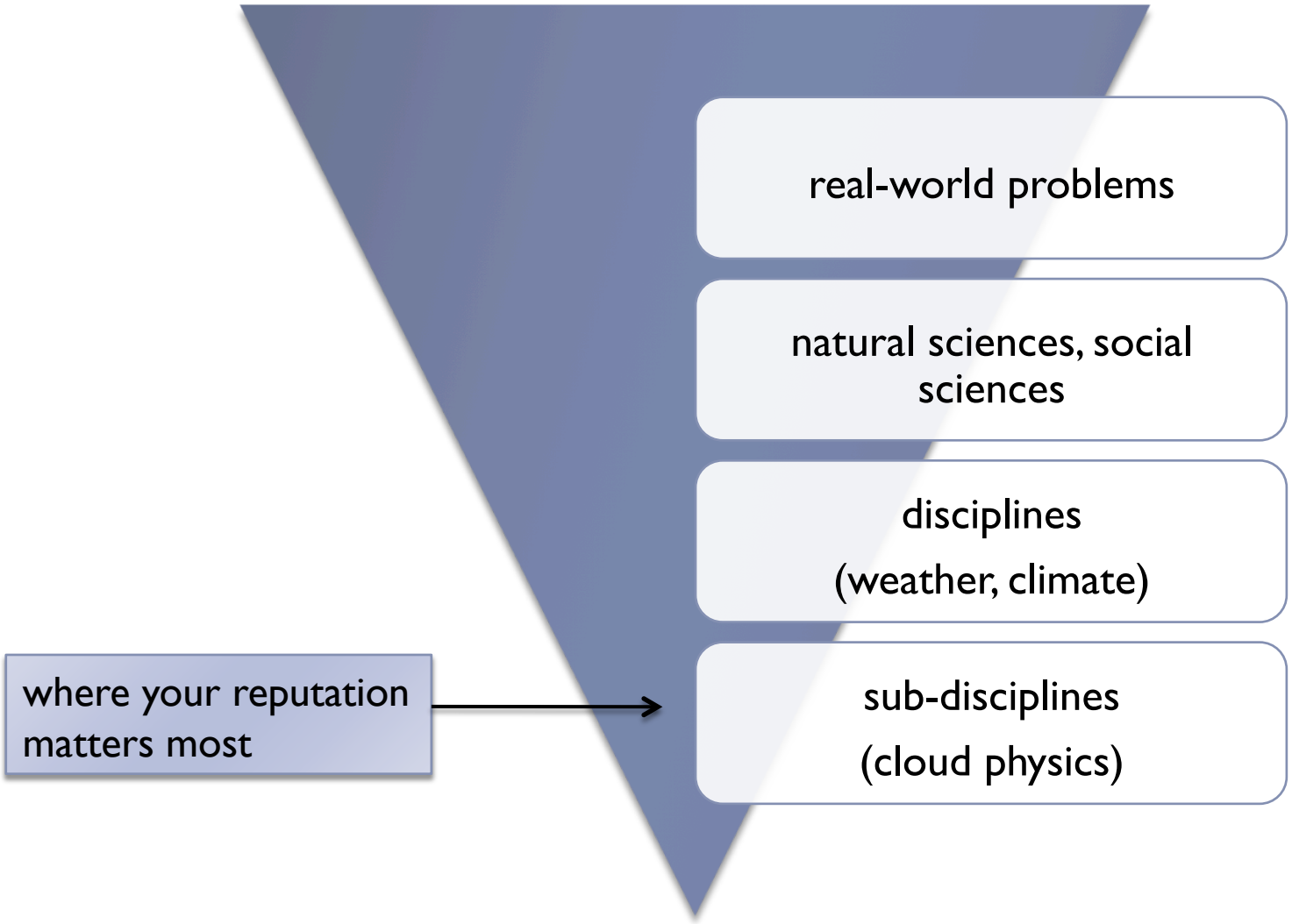
Some lessons from history

- Earth system sciences need global infrastructures
  - At least for data…
- Centralized design and control is not the primary path to working infrastructure
- Instead, build gateways (couplers)
  - Standards, technologies, institutions
  - Must be lightweight, readily understood, easily transferred across regions and cultures
- Interoperability requires ongoing work
  - Coupling with new tools, standards, institutions
  - Maintaining capability amidst changing software ecosystems

Hierarchies and incentives



Hierarchies and incentives



Cyberinfrastructure pitfalls

- Software makes it *seem* easy to build gateways between systems and networks…
  - "You just…"
- … but social, institutional, and security gateways are even more important
  - Multiple institutional cultures
  - Complex projects with many working groups
  - Multiple security and legal standards can block interchange
- New kinds of work and workers at the interface of science and software
  - Current career structures don't have a good place for them