

---

# Internet History Initiative:

*Preserving Our Collective Data Legacy*

Jim Cowie

Fellow, Berkman Klein Center + Library Innovation Lab, Harvard University

# Gilbert Simondon (1924-1989)

"Transforming all the conditions of human life, augmenting the exchange of causality between what man produces and what he is, true technical progress might be considered as implying human progress if it has a **network structure, whose mesh is human reality**"

-- "The Limits of Human Progress: A Critical Study"  
(1959)



[https://en.wikipedia.org/wiki/File:Lemonde-Gilbert\\_Simondon.jpg](https://en.wikipedia.org/wiki/File:Lemonde-Gilbert_Simondon.jpg)

---

# Why preserve Internet history?

*No, seriously, why do we bother*

01

# These are perilous times.

---

- It feels as if the next decade may be one that is as transformative as the 1990s were for the Internet.
- Like it or not, AI is changing the ways information ecosystems function, and that seems likely to drive significant changes in the ways people around the world depend on our network of networks
- At this moment in history, the world has **a lot** to learn from the history of the Internet .. and there are entire academic research communities who we can connect with our community's recorded history
- ...If we work quickly to preserve our datasets and help the institutions that retain them.

# What This Community Can Do To Help

---



Nominate endangered datasets / services (even internal ones)



Introduce us to the humans behind them (who can authorize preservation)



Volunteer a cold-copy host institution (university/NREN/RIR/IXP/library)

In order to make sense of historical events, you have to wait for some time to pass.

## **The Historian's Dilemma**

As time passes, the original recordings of historical events start to get lost, and fade.

In order to make sense of historical events, you have to wait for some time to pass.

As time passes, the original recordings of historical events start to get lost, and fade.

We need to faithfully preserve recordings of what happened, so that future historians will have what they need.

# Welcome to the Internet History Initiative

[Latest News](#) 

[Connect, Discuss](#) 

[Project Overview](#) 

Preservation, curation, and celebration of the Internet's historical datasets



---

# A lightning tour

*In which we show how tricky Internet history interpretation can be, and why we should probably leave it to the experts*

02

# Internet Prehistory (1969-1980)



**1969** – ARPANET first nodes (UCLA, SRI, UCSB, Utah), invention of RFC system

**1973** – Cerf and Kahn begin TCP/IP design work;  
First international ARPANET nodes via satellite

**1976** – X.25 approved by CCITT (competition!)

**1978** – First international X.25 network (IPSS)

**1980** – TCP (RFC 761) and IP (RFC 760)

*All over the world,  
researchers and vendors  
are building testbeds,  
placing bets on  
competing standards*

*None has attained  
critical mass ... yet*

# The 1980s ex-USA: Independent growth



- 1980** – ACSNet (Australia)
- 1982** – SDN (Korea, early TCP/IP)
- 1984** – JUNet (Japan)
- 1985** – PCCS Conference (Korea)
- 1987** – WIDE project (Japan)
- 1988** – ERNET (India)
- 1989** – AARNet (Australia)
- 1989** – Postel assigns 43.0.0.0/8 to Japan

*Addresses and ccTLDs are being assigned manually throughout the region*

*Interconnections to the ARPANet (and later NSFNet) are expensive or forbidden*

*Most early international content is store-and-forward content (UUCP, Usenet)*

# The 1980s: Protocol Groundwork for Scaling



**1982** – Exterior Gateway Protocol (RFC827)

**1983** – "Flag Day": ARPANET mandates TCP/IP

**1984** – DNS TLDs (RFC920, .com .mil .edu .org)

**1986** – NSFNet launches; TCP-based backbone

**1989** – first BGP (RFC1105)

**1990** – NSFNet *also* supports OSI's CLNP

*ARPANet and NSFNet provide significant scaling, but exclude non-NATO partners and commercial interests*

*Competing standards like those from OSI are very much alive; this isn't over yet!*

# The 1990s: Rapid Global Self-Assembly



**First wave:** PACCOM (Hawaiian 64 Kbps satellite, connected Australia, New Zealand, Japan, and Korea to the NSFNet over TCP/IP in 1989)

**Second wave:** Singapore (1990), Hong Kong, Taiwan (1991), India, Malaysia, Thailand (1992)

**IPv4 Exhaustion:** RFC1519 (1992) estimates that Class B space will deplete "within 15 months."

*The Internet's advocates have a terrifying realization: they did not design for the kind of demand scaling that appears by the early 1990s...*

*... and it was becoming apparent that this would need to be a truly global Internet*

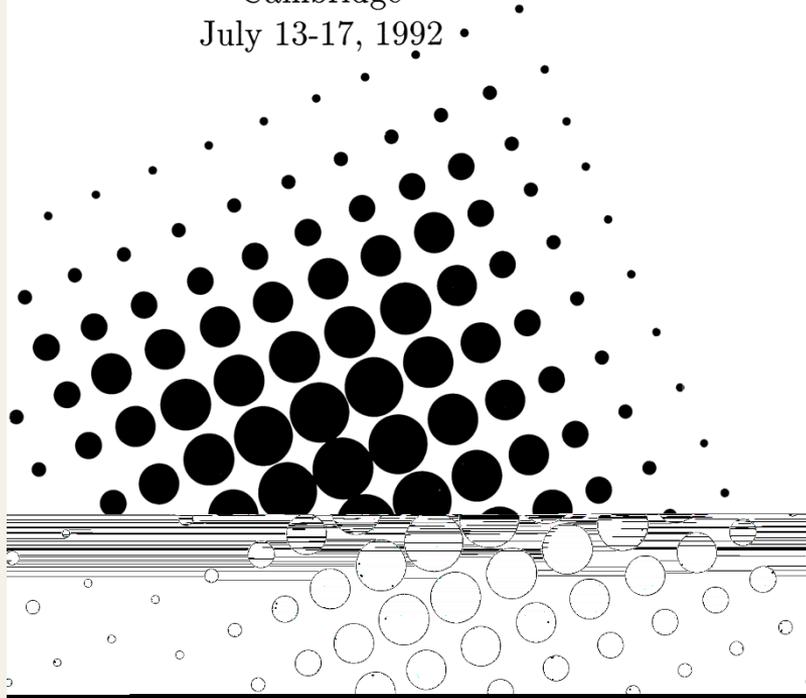
Proceedings of the Twenty-Fourth  
Internet Engineering Task Force

Massachusetts Institute of Technology

NEARnet

Cambridge

July 13-17, 1992



“As an initial allocation, North America and its service providers would be allocated the Class C networks under <220.0.0.0 – 252.0.0.0>, Europe and its service providers would be allocated the Class C networks under <216.0.0.0 – 252.0.0.0>, and **the remaining continents** would be allocated about 10,00 Class C networks each.” [sic]

-- Minutes of the IP Addressing Plan BOF

<https://www.ietf.org/proceedings/24.pdf>

# IPv4 exhaustion was almost avoided in 1992

## Abstract

Internet growth has created serious problems of address space consumption and routing information explosion. A solution to these problems requires a new version of the Internet Protocol, which we call IP version 7 ('IPv7'). This memo presents architectural guidelines that any IPv7 should meet. It then discusses how an IPv7 based upon the OSI CLNP protocol would meet these requirements, and presents the reasons for the IAB's preference for this solution. Finally, it makes a three-part recommendation: (1) proceed at full speed on CIDR; (2) do the design work on IPv7 based on CLNP; and (3) continue to pursue research in advanced routing and other future extensions of the Internet architecture.

-- draft-iab-ipversion7-00 (Kobe, 1992)

# ...But the protocol wars were too bitter

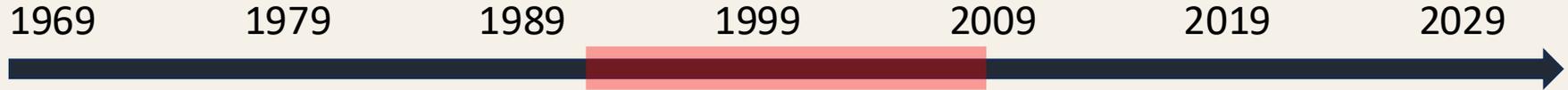
RFC 1752

Recommendation for IPng

January 1995

In the late spring of 1992 the IAB issued "IP version 7" [[IAB92](#)], concurring in the ROAD group's endorsement of CIDR and also recommending "an immediate IETF effort to prepare a detailed and organizational plan for using CLNP as the basis for IPv7." After spirited discussion, the IETF decided to reject the IAB's recommendation and issue the call for proposals recommended by the ROAD group. This call was issued in July 1992 at the Boston IETF meeting and a number of working groups were formed in response

# So from 1992, "our Internet" unfolds in data



- Nepal (1993), China (1994), Sri Lanka (1995), Mongolia (1996), Cambodia (1997), ...
- **1992** – RIPE NCC
- **1993** – APNIC (pilot)
- **1993** – Mosaic browser
- **1994** – APNIC gets 202/8,203/8
- **1995** – end of NSFNet
- **1995** – Netscape IPO
- **1995** – APNIC (formalized)
- **1997** – ARIN (from InterNIC)
- **1998** – ICANN
- **2002** – LACNIC (from ARIN)
- **2003** – NRO (from the RIRs)
- **2005** – AFRINIC (from ARIN+RIPE)

# From 1992, "our Internet" unfolds in data

1969                      1979                      1989                      1999                      2009                      2019                      2029



From: Randy Bush <randy@psg.com>

To: goldstein@nsf.gov

Date: Tue, 7 Jun 1994 12:18:18 -0700

Indonesia is up on the new 64kb link,  
traceroute caddsys.iptek.net

...

202.46.1.2 (202.46.1.2) 1830 ms 755 ms 759 ms

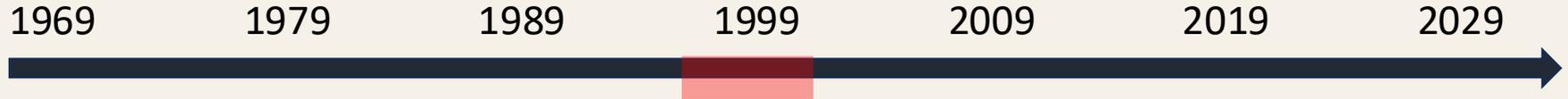
---

# The age of measurement arrives

*In which we describe the earliest global Internet measurement repositories, and wonder where the rest got to over the years*

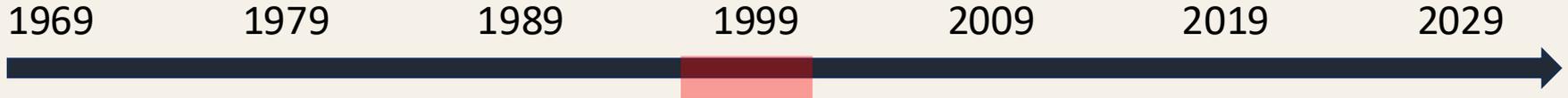
03

# Institutional measurement begins (~1997)



History becomes legible once institutions begin compiling the Internet’s “operational exhaust”

# Institutional measurement begins (~1997)



## BGP Route Collection

- Oregon Routeviews (1997)
- Packet Clearinghouse (1997)
- RIPE RIS (1999)

## Ping/Traceroute

- SLAC PingER (1997) ←
- CAIDA Skitter (1998)

# SLAC PingER: Endangered Data

- **Continuous** collection from 1997-2024
- Estimates of latency, jitter, and bandwidth
- Specific attention to Global South universities and developing economies



**Quantifying the Worldwide Digital Divide: The Emergence of Africa**

*Prepared by: R. Les Cottrell  
SLAC, Stanford, USA*

*ISPA/iWeek, Pretoria, S. Africa, Sept 21, 2011*

*[cottrell@slac.stanford.edu](mailto:cottrell@slac.stanford.edu)*



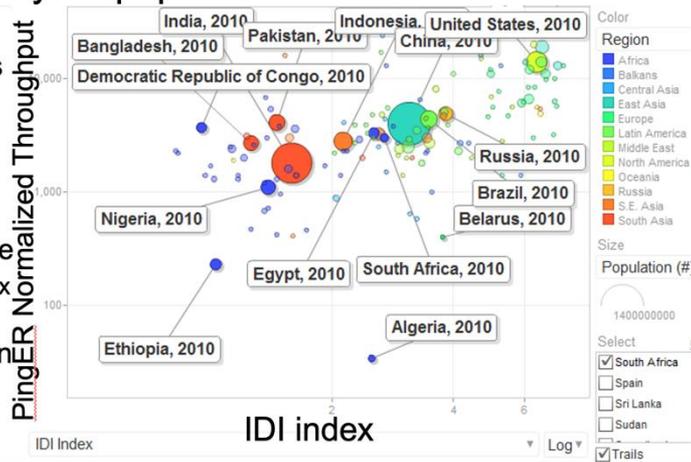
SEACOM is a 1.28Tbps, 15,000 kilometer-long undersea fiber-optic cable system that will provide high-speed communication between East Africa, South Asia and

# All Projects Come to an End Eventually...



## PingER throughput & IDI

- Positive correlation between PingER throughput & IDI, especially for populous countries
- PingER measurements automatic
- No army of data gatherers & statisticians
- More up to date
  - IDI 2009 index for 2007 data
- Good validation
- Anomalies interesting

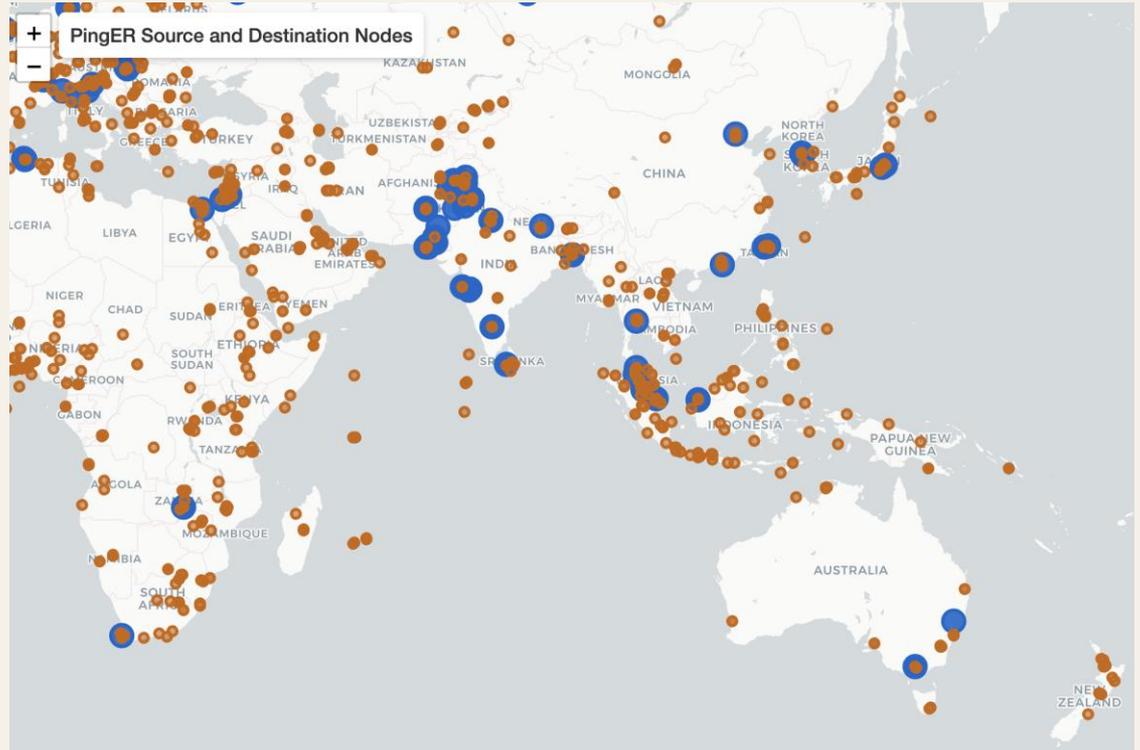


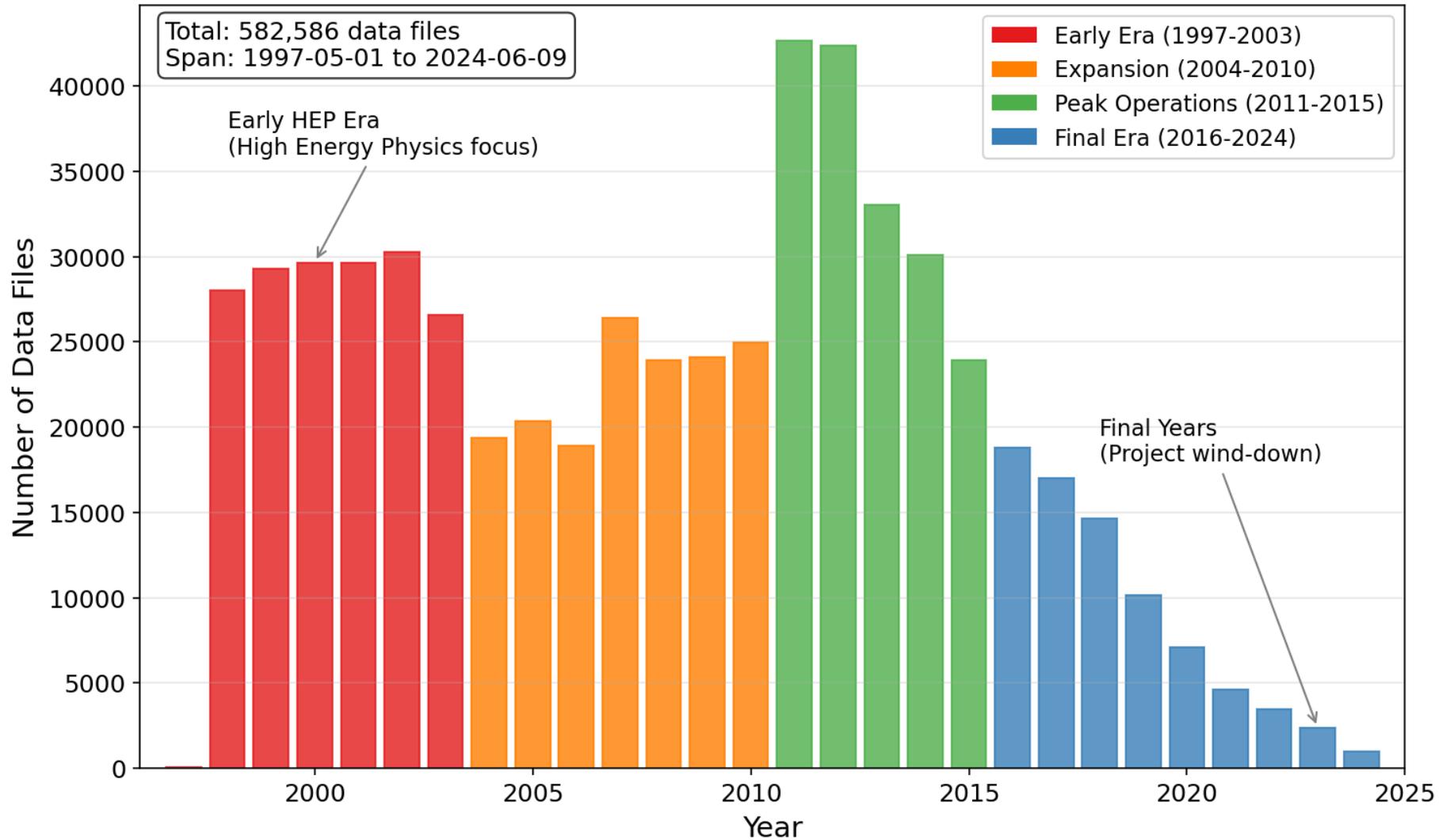
- Les Cottrell was among the first researchers to begin to draw the connections between Internet connectivity and economic development
- **Upon his retirement, this system went down and the data was believed lost.**

# IHI was able to rescue the Pinger data!

In late 2024, SLAC restored access to the servers just long enough for us to make a copy of the project's directories

*Sources (blue)*  
*Destinations (red)*





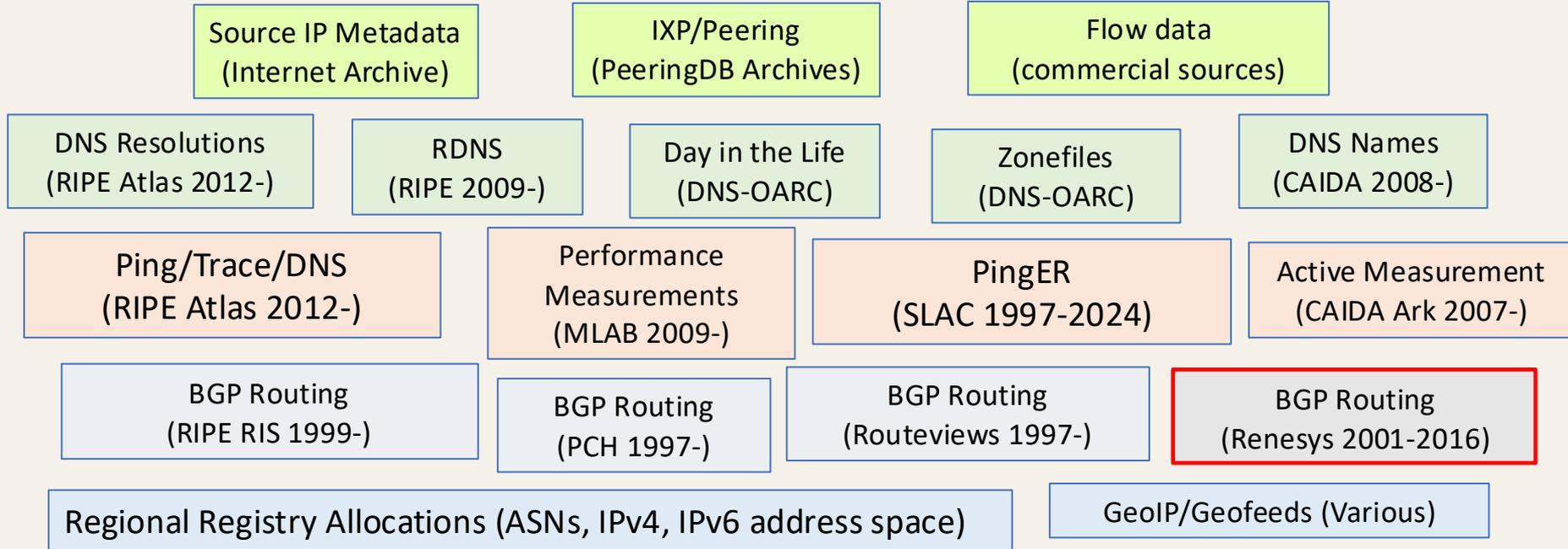
---

# Convening a community

*Celebrating the preservers of Internet history*

04

# Our history is greater than the sum of its parts



# The IHI challenge will require **Two** Collections

---

- The first collection is purely for **preservation**

“our offline backups”

- The second is a working collection, to support ongoing **research**

“our rebuildable data lake,  
open to the public”

# The first collection is purely for **preservation**



Archival copies of primary datasets



Minimal curation, but clear chain of evidence from original sources



Apply checksums, create metadata, break into volumes that can be replicated and shared to as many institutions as would like to host 'cold copies'



Plan to recopy this to new media every decade to meet century-scale retention

## Century-Scale Storage

Maxwell Neely-Cohen

 HARVARD LAW SCHOOL  
Library Innovation Lab

<https://lil.law.harvard.edu/century-scale-storage/>

# The second is a working collection, for **research**

- Derived from the cold preservation collection, in modern formats to support **integrated exploration**
- We can tear this data lake down and **rebuild it over time** to suit the challenges of the day
- Derive **open-source tools** that expand the horizon of ways researchers can approach the collection and find meaning in it
- Provide UIs for **coders and noncoders**: make complex APIs available to nontechnical researchers 

# What Might We Do With Such a Collection?

- Throw the door open for collaborations with artists, storytellers
- Build an immersive AR walkthrough of the developing Internet in an urban space.
- Integrate LLM UIs to let the collection speak to a broader audience
- Build ‘science museum’ exhibits that answer questions about how the Internet evolved across particular regions, with maps and timelines
- Find some truly new ways to make the Internet's geographic history and social benefits **tangible, public, and participatory.**

---

**Would you like to  
help?**

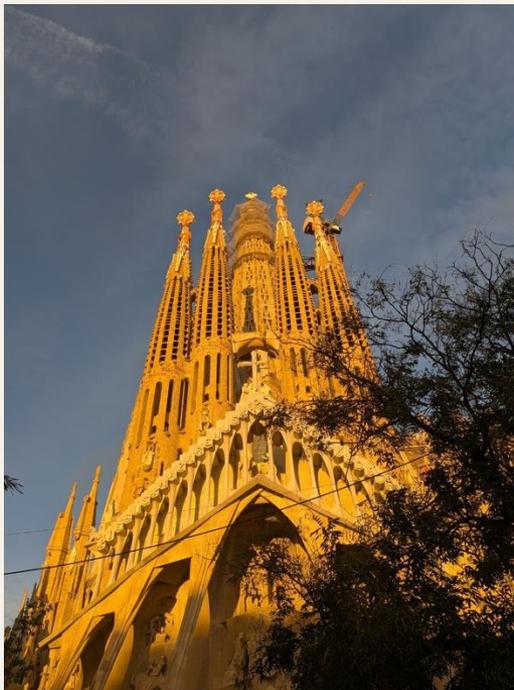
05

# Internet Data Trust Workshop, Barcelona, Dec 2025



- Over 60 participants as part of the Internet Society's Pulse Measurement Week - **thank you!**
- Conversations around prioritization, principles, preservation
- Valued guests from the library science and archivist community, as well as public science platforms

# Internet Data Trust Workshop: Key Findings



- Preservation is expensive and can be hard to justify in terms of resourcing
- Potential for mutual aid to avoid disasters
- Bringing datasets into the light together creates opportunities for new research
- Archivists have seen this all before; Lots of Copies Keep Stuff Safe (LOCKSS)

# Pay Attention; Help Save Endangered Data

- Most people would guess that a preservation initiative needs storage .. machines .. rackspace .. and this is true.
- But the primary resource that runs out is **time**
- By the time we learn about an endangered dataset, it may be too late, and it's lost or destroyed

# What This Community Can Do To Help



Nominate endangered datasets / services (even internal ones)



Introduce us to the humans behind them (who can authorize preservation)



Volunteer a cold-copy host institution (university/NREN/RIR/IXP/library)

---

# Thank You

*Lots of Copies Keep Stuff Safe*

[internethistoryinitiative.org](https://internethistoryinitiative.org)

Mastodon: [@IHI@cooperate.social](https://cooperate.social/@IHI)

[cowie@internethistoryinitiative.org](mailto:cowie@internethistoryinitiative.org)

