Contemplating Internet History: Where Do We Go Next?

Jim Cowie

Fellow, Berkman Klein Center for Internet & Society at Harvard University

Fellow, Library Innovation Lab, Harvard Law Library

23 April 2025

How Should We Study Internet History?

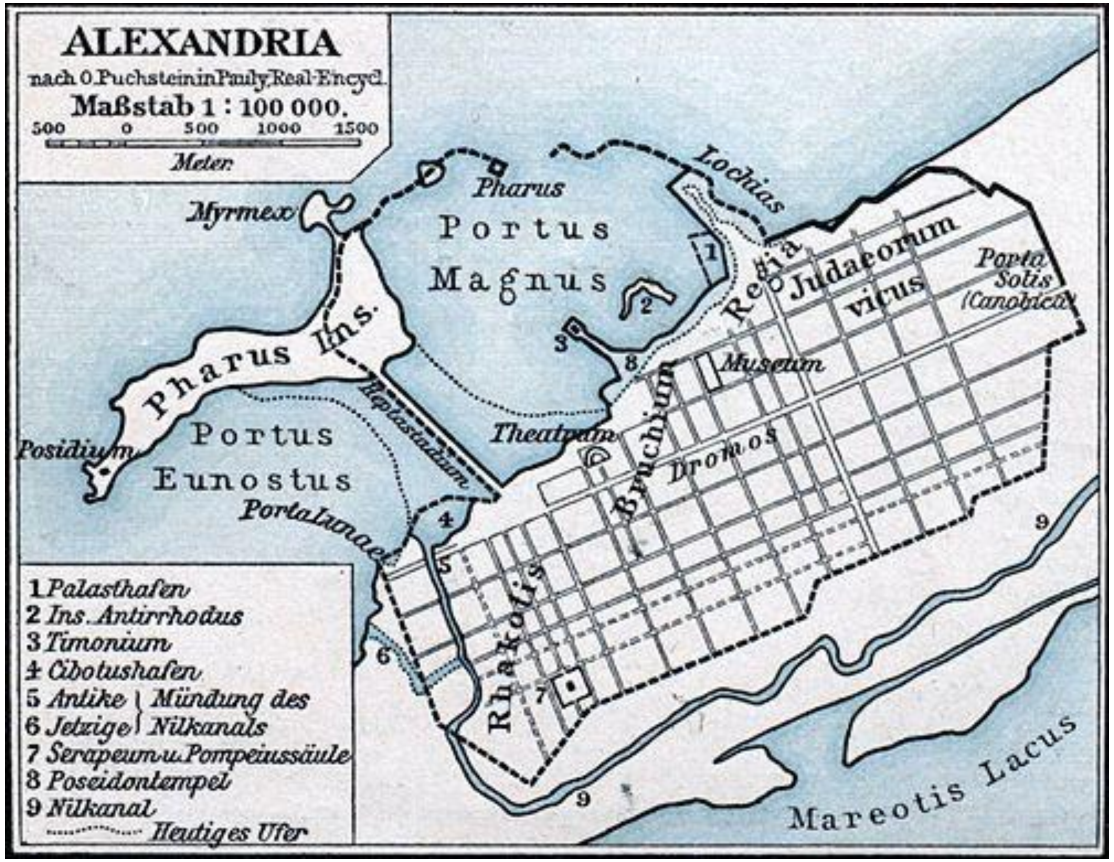
- Oral culture (various Internet History listservs)
- Network science (email archives, regional NOG attendance)
- Cultural anthropology (studying the IETF's "Loud Men Talking Loudly")
- Regional perspectives, colonialism, technological determinism
- Deep dives on specific parts of the ecosystem (example: IXP history)

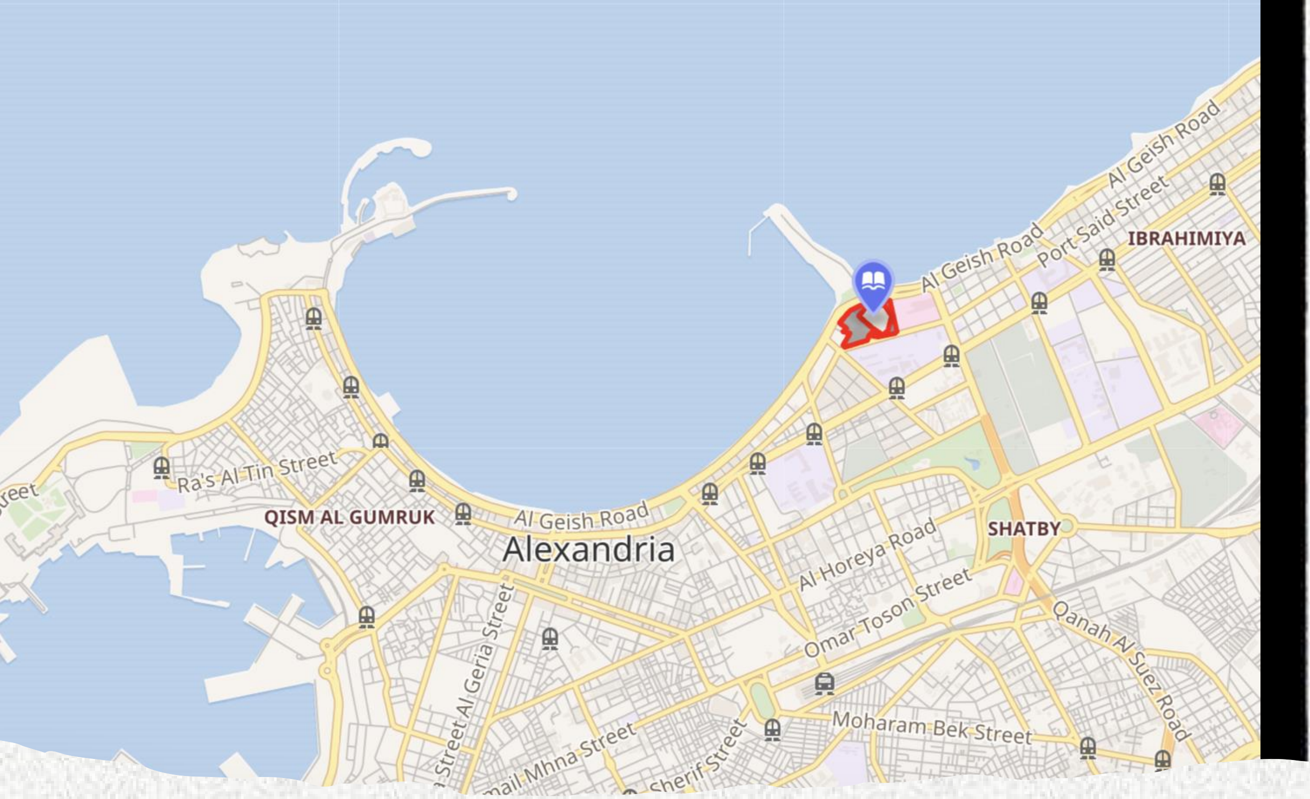
- Can study the prehistory (1960-1969), the competing standards era (1975-1992), the early expansion (1993-1999), and the modern era (1999-2024).

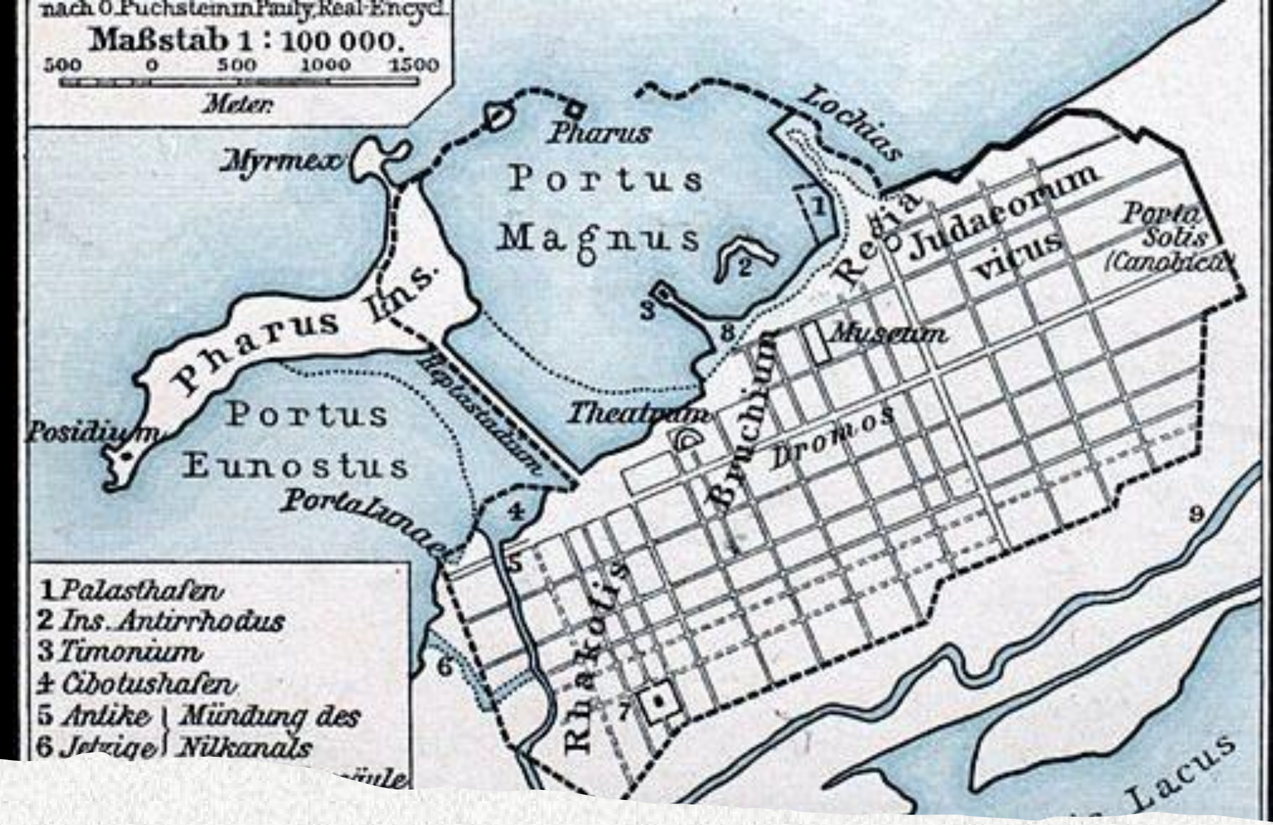
- My work this year tackles something a lot simpler and more straightforward: preservation of the Internet's recorded data history.

Consider the Great Library of Alexandria! 🔥

- Set in motion by Ptolemy I (or II) c.300 BCE
- Famously ‘burned’ by Caesar during his siege in 48 BCE 🔥
- In fact, the Library had already started to decline, and would continue to decline for centuries
- History requires **maintenance!**





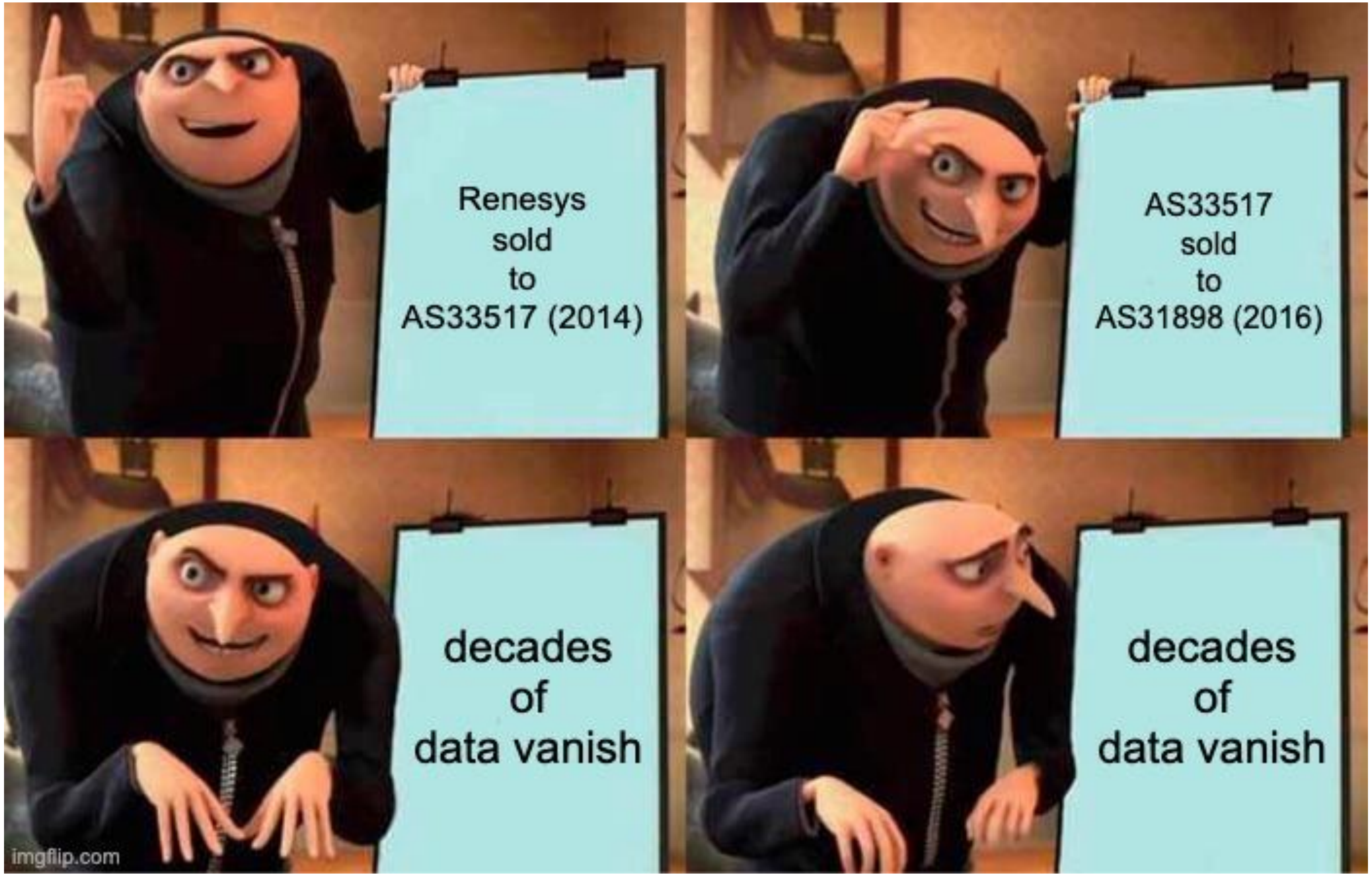


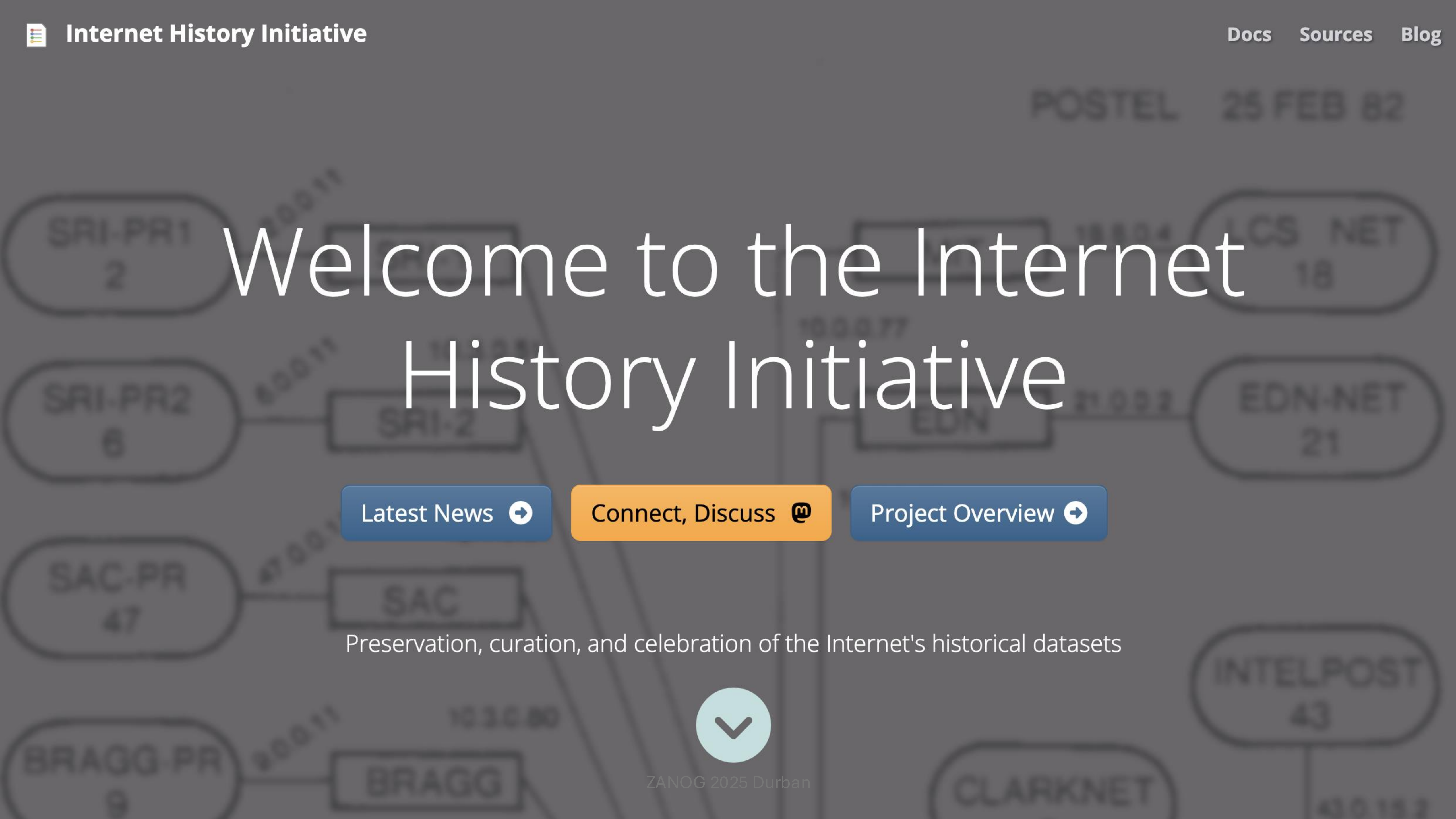
Biblioteca Alexandrina (2002-)

- Reborn as a new library, with massive digital collections available online
- Digital preservation for the centuries creates an entirely new set of challenges!

The Painful Origin Story of the History Initiative

or: an abrupt feeling of loss





Welcome to the Internet History Initiative

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

How can we help the historians of 2125?

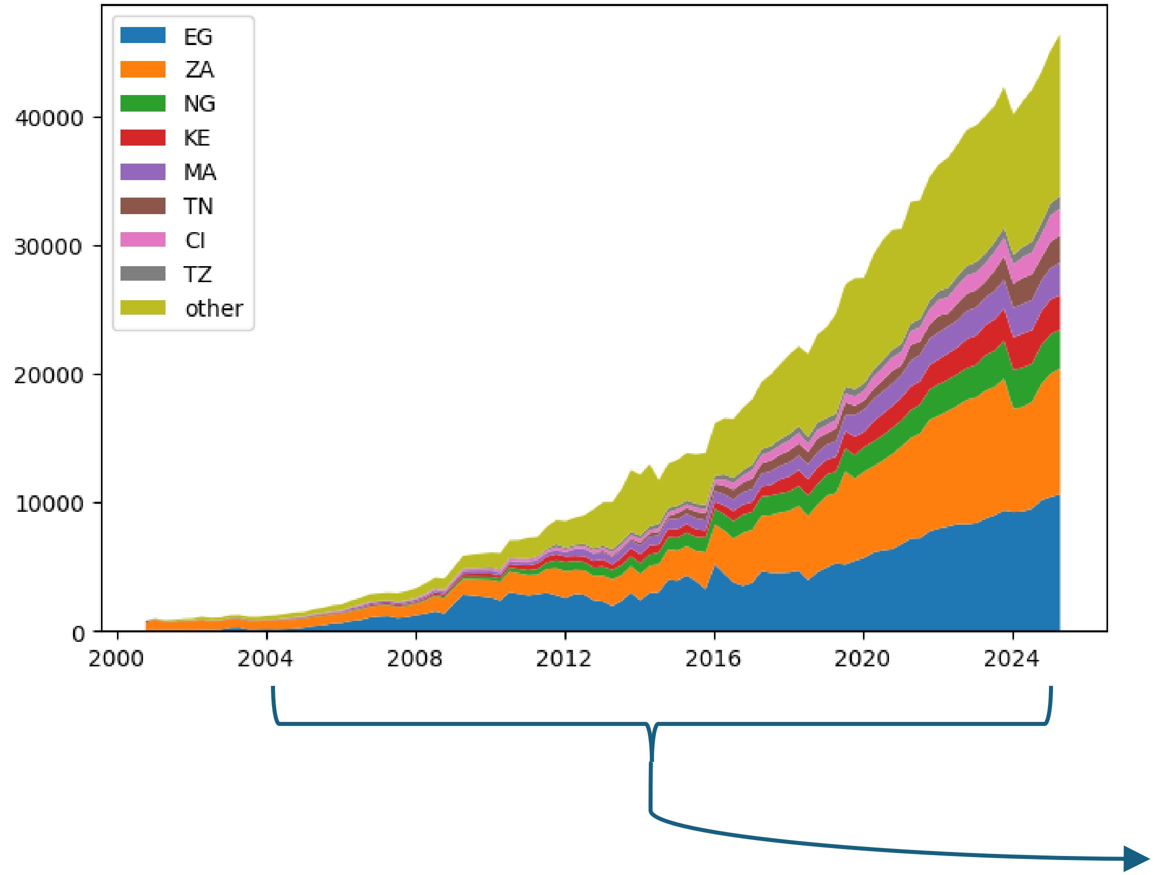
- Catalog our irreplaceable data sources
- Lots of Copies Keep Stuff Safe (LOCKSS principle)
- Document the “rapid expansion phase” of the Internet, from about 1999 through 2025, everywhere on Earth
- Derive time series that will support social science research (development economics, political science, conflict studies)
- Map out the sustaining framework ($) that will maintain these collections for future generations

Consider what regional NOGs built together as part of a global community…

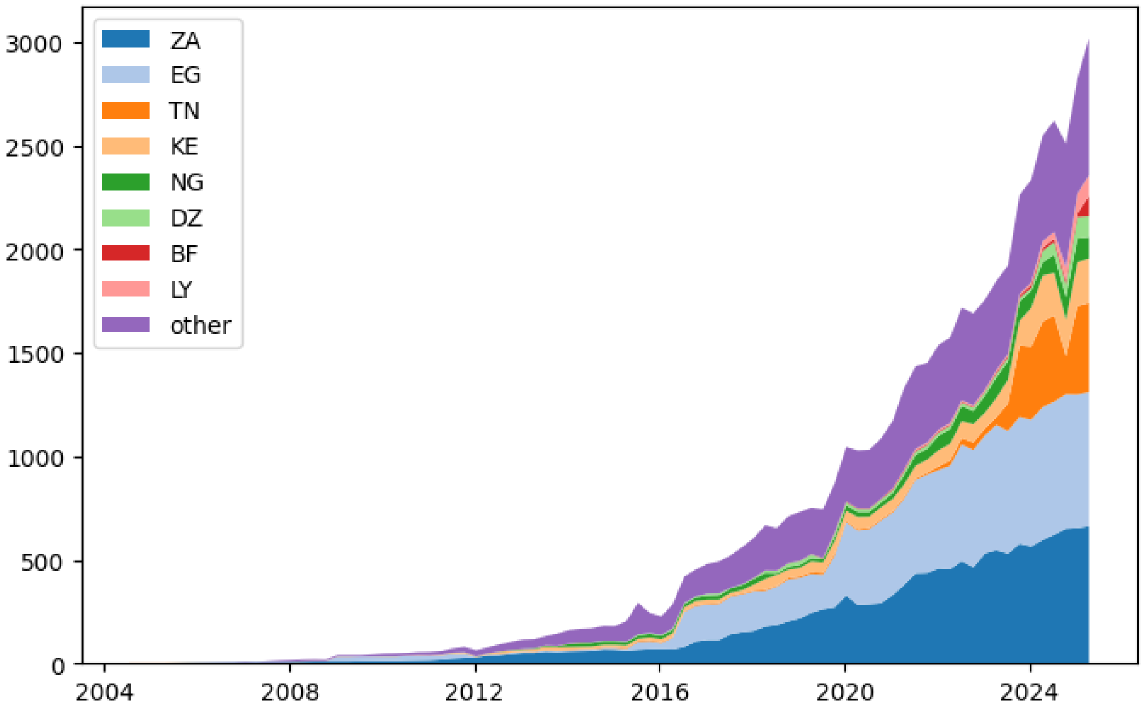
- Network operators have contributed decades of BGP sessions (locally at IXP collectors, multihop in remote collectors) to RIPE RIS and Oregon Routeviews
- These sessions record the evolving history of Internet interconnection from thousands of distinct operator perspectives, second by second, since ~1998!
- This is especially critical for telling the story of the Internet’s growth and diversification in regional markets like South Africa
- This data is complemented by decades of active measurements of latency, loss, and router paths traversed, like those performed since 1998 on the PingER platform, or since 2010 on the RIPE ATLAS platform

African BGP Routing Table Growth, 2000-2025

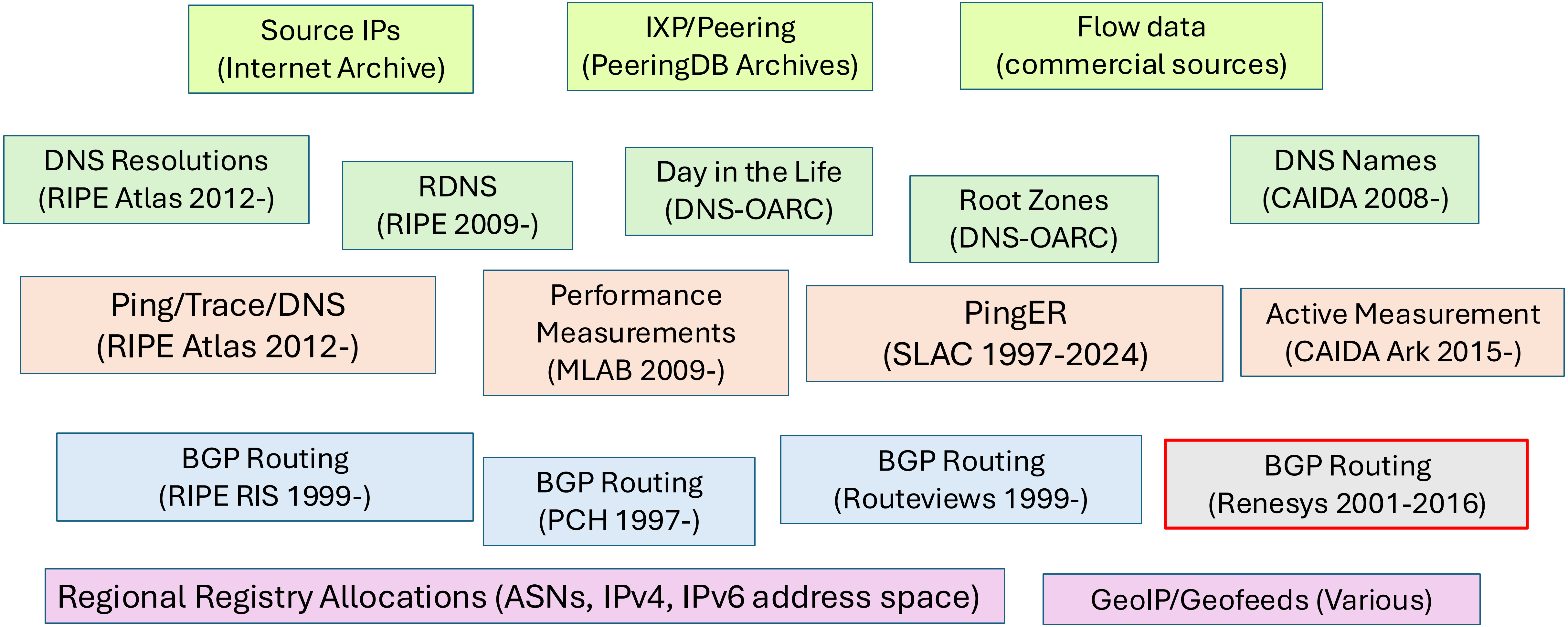
Total originated prefixes (RIS RRC00)



IPv6 originated prefixes (RIS RRC00)



The sum of our history is greater than its parts



Source IPs (Internet Archive)

IXP/Peering (PeeringDB Archives)

Flow data (commercial sources)

DNS Resolutions (RIPE Atlas 2012-)

RDNS (RIPE 2009-)

Day in the Life (DNS-OARC)

Root Zones (DNS-OARC)

DNS Names (CAIDA 2008-)

Ping/Trace/DNS (RIPE Atlas 2012-)

Performance Measurements (MLAB 2009-)

PingER (SLAC 1997-2024)

Active Measurement (CAIDA Ark 2015-)

BGP Routing (RIPE RIS 1999-)

BGP Routing (PCH 1997-)

BGP Routing (Routeviews 1999-)

BGP Routing (Renesys 2001-2016)

Regional Registry Allocations (ASNs, IPv4, IPv6 address space)

GeoIP/Geofeeds (Various)

Example: What We Stand To Lose

- The PingER project collected continuous data from 1997 to May 2024
- Estimates of latency, jitter, and bandwidth
- Specific attention to Global South universities and developing economies
- **Status: ~~Believed Extinct~~ Rescued**

SLAC
NATIONAL ACCELERATOR LABORATORY
ICFA
Fermilab
ICTP
FOTOSEARCH
UN Defined African Regions

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

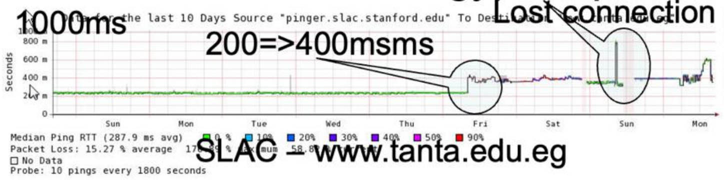
SEACOM is a 1.28Tbps, 15,000 kilometer-long undersea fiber-optic cable system that will provide high-speed communication between East Africa, South Asia and

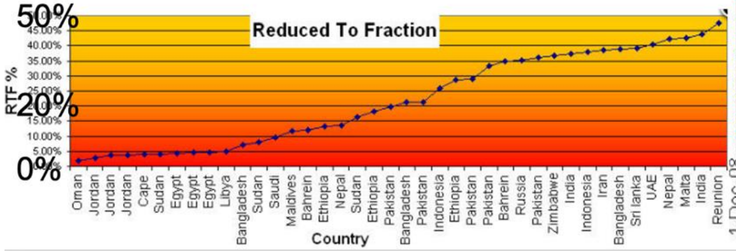
http://www.slac.stanford.edu/grp/scs/net/talk11/africa-sep11.pptx

Early, Important Work

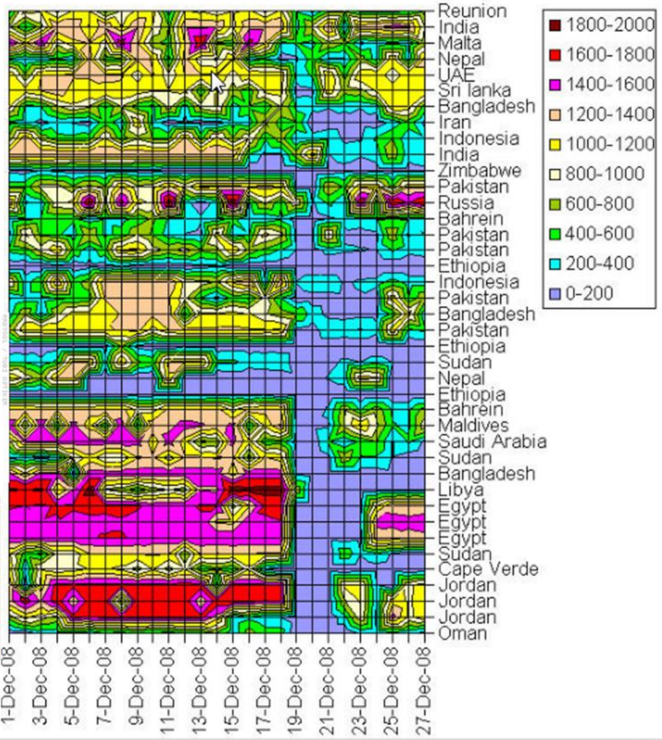
Example: Multiple routes important

- Not only for competition
- Need redundancy
- Mediterranean Fibre cuts
  - Jan 2008 and Dec 2008
  - Reduced bandwidth by over 50% to over 20 countries
- New cable France-Egypt Sep 1 '10





Derived TCP Throughput in kbits/sec from SLAC to Hosts in Countries Affected by the Mediterranean Fibre Cable Cuts December 2008



SLAC NATIONAL ACCELERATOR LABORATORY

PingER throughput & IDI

- Positive correlation between PingER throughput & IDI, especially for populous countries
- PingER measurements automatic

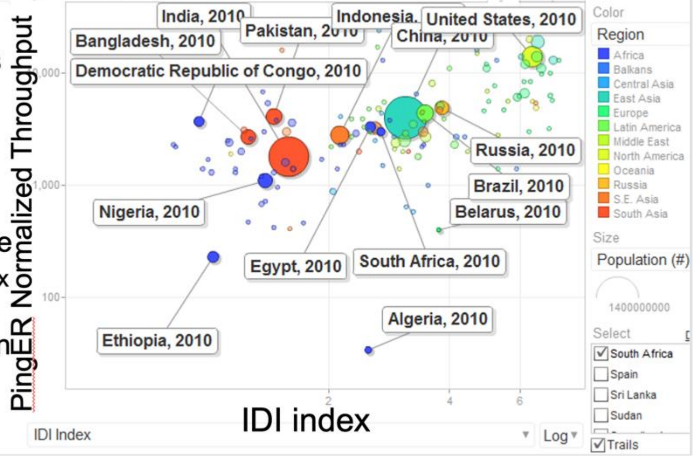
No army of data gatherers & statisticians

More up to date

- IDI 2009 index for 2007 data

Good validation

Anomalies interesting



A copy of this data was revived by SLAC IT in September 2024

It's being mirrored to IHI's S3 for intermediate preservation

Next steps: curation, transcription to modern data formats

The IHI challenge requires **Two** Collections

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

“Make sure we don’t lose irreplaceable datasets”

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

“Make sure the world understands why we preserved these datasets, why they matter”

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**: LLM research interfaces to make complex APIs available to nontechnical researchers

Sidebar: AI and the Internet

- **One of the reasons we study history is to learn from it**
- The Internet has evolved over 50 years to its current nonideal form
- AI infrastructure recapitulates Internet infrastructure
- We will have the same struggles over American control, regionalization, state control, centralization of power
- Except we now have them emerging on a timescale of months, not decades
- Prediction: we will see the Internet’s multistakeholder/multilateral governance wars fought again, this time over AI governance+safety

What comes next? What will we be able to do with the IHI collections?

- Throw the door open for collaborations with artists, storytellers
- Build an immersive AR walkthrough of the developing Internet in an urban space.
- Build a question-answering LLM that uses our analytic APIs to answer questions about how the Internet evolved in a particular region, with maps and timelines
- Find some truly new ways to make the Internet's geographic history and social benefits tangible, public, and participatory.

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

Thank you!

https://internethistoryinitiative.org

Mastodon: IHI@cooperate.social