# IXPs and developments in networking: A 30-year personal review

Philip Smith

<philip@nsrc.org>

LINX 123

20th & 21st November 2024

UNIVERSITY OF OREGON

Last updated 17th November 2024

NSRC
Network Startup Resource Center

# Short biography

- Caught the Internet bug at University in 1986 while doing my PhD and PostDoc
  - HP 9000 Unix systems, early TCP/IP & 10base5/10base2 Ethernet
- 1993 to 1997 at PIPEX as Network Engineer etc
- 1998 to 2011 at Cisco as Consulting Engineer
  - BGP skills and best practices workshops, designing & building ISPs and national backbones, creating IXPs, co-founding NOGs, deploying IPv6…
- 2011 to 2013 at APNIC as Director of Learning and Development
- 2013 to date, working for myself
  - Major contract work as Senior Network Engineer & NOG Coordinator for Network Startup Resource Center
  - APRICOT Summit, numerous private clients
  - Pro bono work for NOGs and non-profits around the world
- 2021 – Elected to Internet Hall of Fame

# The journey…

- The first recognisable IXPs appeared in the early 1990s
- Emergence of global transit providers
  - Evolving from the US national backbone providers (Tier-1s)
- Emergence of global content providers & distribution networks
- Today:
  - Very rich IXP infrastructure around the globe
  - Global Content Services Providers: the “Hyperscalers”
- Future:
  - What comes next?

# The “Early” Internet

- What we now recognise as the modern Internet really started in the early 1990s
  - Moving from the command line to early browsers with hypertext
  - Computers for the general public had TCP/IP built-in → Windows 95
- Early customers at PIPEX:
  - Exploring/researching this new Internet thing
  - Moving connectivity from dedicated leased line to running over an IP network – significant cost reduction
  - Using Internet infrastructure between offices or to suppliers (domestic and international)
  - Overall: economies of scale using IP network

# The “Early” Internet: Challenges

- Bandwidth was small
  - 64K was the standard leased line in 1993; imagine that today!
  - “Internet connectivity” meant leased line to the US – 256K to E1 in 1993
- Applications were delay sensitive
  - 120ms or more between London and Falls Church, Virginia
  - LAN applications (running encapsulated inside IP) were not tolerant of delays or low bandwidth
- Early ISPs serving the same locality were not connected to each other
  - Often only via the US
  - Problematic for enterprises using more than one ISP for access

# Neutral Interconnect idea

- May 1994
  - ULCC circulated a proposal to form a “low key neutral Interconnect” in UK
  - PIPEX independently considering an IX concept, inspired by SE-GIX
- 5th August 1994
  - First meeting to discuss creating an IXP in UK
  - Participants – representatives from:
    - UKERNA (JANET) → UK academic (research & education) network
    - PIPEX
    - BTnet (newly announced)
    - Demon Internet
    - EUnet GB (UKnet joined EUnet alliance)

LINX98: https://www.bgp4all.com/pfs/_media/conferences/linx98-earlylinx-philipsmith.pdf

# Initial draft MoU

[Last updated 26-Aug-94, version 0.1]

Memorandum of Understanding
==========

This MoU is intended to help organisations who wish to join the LINX to understand what the LINX is, and what the requirements are for those who wish to join.

This MoU has accompanying documents which should be read by any potential members before they join. One details the legal basis for the LINX. The other is a prototype of a peering agreement to be used as a basis for a peering agreement between LINX Members.

Aims of the LINX
---
The LINX has two primary objectives.

1. To provide efficient interconnectivity for UK Internet Service Providers.
2. To further the cause of the UK Internet within Europe.

Cost Recovery
---
There are some costs associated with running the LINX. These costs will be recovered from the existing members on an equal share basis.

New members will pay on joining the agreed annual fee for existing members for that year. Members joining more than halfway through the year will pay half of annual cost. For the purposes of these calculations, the year starts on the 1st October.

The annual fee will be reviewed annually by the existing members, so as to cover costs during the year. The LINX will be run on a not-for-profit basis. The annual fee will provide for housing of one router at the LINX, not larger than a CISCO AGS+. If required, this can also include a megastream modem. If the member requires additional housing at the LINX, this will have to be paid for separately.

# Setting up LINX

- Work done by PIPEX network engineering staff:
  - Big boxes of cables, tools & parts in the back of my car, and off to Telehouse London
  - Bits included:
    - 8-port 10baseT ethernet hub, with 10base2 “uplink” port at the back, external DC PSU
    - “rack mount” frame for ethernet hub – could accommodate 3 side-by-side
  - Link from PIPEX PoP to the LINX hub connected as well
- Ethernet Hub came out of PIPEX Cambridge PoP
  - We had just replaced the hub with a Cisco Catalyst WS-C1202 switch
  - “Rebadged” Crescendo which Cisco had acquired late in 1993

# LINX

- LINX was proposed as the 4th site of a global D-GIX (distributed global Internet Exchange)
  - Joining CIX-West, MAE-East, and SE-GIX
  - But D-GIX never happened
- By May 1995, LINX had 6 members
- Early LINX benefits:
  - Each UK provider had a local (low latency & relatively high bandwidth) connection to the other via a common interconnect
  - No need to construct a full mesh of direct connections
  - Reduction in UK domestic traffic having to use expensive links to the US or mainland Europe
  - Attractiveness of locating early content in UK for easy access to consumers (Microsoft's European DataCentre for one – remember Windows 95)

# Early problems @ LINX

- Started life as an 8-port ethernet hub (10Mbps half duplex)
  - Dropping packets badly at around 2Mbps
- Replaced by Cisco Catalyst WS-C1202:


London Science Museum Picture

- Now each provider could run their 10Mbps link at near capacity
  - But LINX still needed more than 8 ethernet ports!

# Lots of lessons for the future!

- Need plenty of switch ports
  - Surely not a problem in 2024 😇
- Need to have redundant, reliable, fault tolerant infrastructure
  - IXP is critical to local Internet infrastructure, outages are “not tolerable”
- Member minimum requirement:
  - Independent address space and AS Number, plus fully BGP capable router
- IXP needs to focus on facilitating members to exchange traffic
  - It’s not an industry association, nor a lobby group (apart from advocating for IXPs)
- IXP must not compete with its members
  - Otherwise, why would members join
- Not every provider will join
  - Only if it makes technical ***and*** commercial sense

# Today ?

- We think we understand what IXP best practices are
  - A lot has been learned in the last 30 years, especially by the pioneers
  - There are 300+ IXPs listed in PeeringDB (https://peeringdb.com)
- But:
  - Many locations remain without an IXP
  - Many locations believe no IXP is needed
  - Many believe they can repeat the mistakes made over the last 30 years and get a better outcome

# Transit

- The early Internet was all transit
- The 1993 PIPEX had a transatlantic leased line that connected us to UUNET
  - We paid for the link ***and*** paid to connect to them
  - Very expensive – this is the transit model
- With more and more access providers, more opportunity to sell transit, leading to a significant growth in transit providers
- The US nationwide carriers (Tier-1s) transitioned into the global transit providers, joined by a few from Europe
  - By early 2000s, the global Tier-1s were reasonably well known

# Transit

- On 28th February 1999:

```
BGP routing table entries examined:                          56027
Total ASes present in the Internet Routing Table:             4634
Origin-only ASes present in the Internet Routing Table:       3834
Origin ASes announcing only one prefix:                       1268
Transit ASes present in the Internet Routing Table:            800
Transit-only ASes present in the Internet Routing Table:        40
```

- On 15th November 2024:

```
BGP routing table entries examined:                          968007   17.3x
Total ASes present in the Internet Routing Table:             76457   16.5x
Origin-only ASes present in the Internet Routing Table:       65497
Origin ASes announcing only one prefix:                       26856
Transit ASes present in the Internet Routing Table:           10960   13.7x
Transit-only ASes present in the Internet Routing Table:        532   13.3x
```

# Research & Education Networks

- Commercial Internet grew out of Research & Education infrastructure
  - Early to mid-90s PIPEX customers had to sign NSFnet's AUP to access the US NSFnet (no commercial traffic)
    - No agreement meant no access to NSFnet – many didn't realise the impact
    - NSFnet AUP resulted in the commercial transit providers in the US
- As commercial Internet transit grew, so did the global R&E infrastructure
  - National RENs (NRENs) and Regional RENs
    - Regional RENs were the transit providers for the NRENs and each other
  - RENs turned up at Internet Exchange Points and bought commodity access from commercial ISPs

# The rise of the Regional IXP

- What was the role of the IXP?
  - Some local Internet Exchange Points attracted network operators from other countries or regions
  - Leading to the birth of the regional IXP
- PIPEX went to MAE-East (Falls Church, VA) and to CIX-West (Palo Alto, CA)
  - Not just a LINX presence (which only had UK content)
- Getting to MAE-East give us access to ISPs in the US without relying on or paying for transit from UUNET
  - But we still bought transit from UUNET from our routers located nearby the MAE-East facility

# Motivation to connect to a Regional IXP

- Big growth in the early 2000s of new network operators:
  - Connecting to their local IXP, keeping local traffic local
  - Purchasing IPLC (International Private Leased Circuit) to an IXP in another country or continent
- Why go to foreign IXPs?
  - Chasing ***content***, value, and service quality for customers!
  - Place a router at that IXP
  - Settlement-free peering with members of that IXP (no traffic cost, only the IPLC cost)
  - Purchasing transit at that location (often significantly cheaper than locally)
  - Reduced dependency on quality & capacity of transit provider infrastructure
  - Opportunity to make its own routing policy decisions

# Regional IXP: side note

- A Regional IXP is one that has grown in significance and stature such that it appeals to members from outside its local area
  - It is still an IXP, but now with international members
  - In 2024 there are many Regional IXPs around the world – this is a good thing!

- A Regional IXP is ***NOT*** a group of IXPs interconnected by some mystical free limitless bandwidth
  - Some regulators and funding agencies seem to think this and are still trying to promote the concept – why??
  - Everyone likes free bandwidth, especially if someone else is paying for it 

# From Transit to Content

- But then something changed later in the first decade of this century:
  - Content moved from being a “single website” run by each content producer to a global distributed infrastructure
  - The emergence of content distribution networks as well as the content providers setting up infrastructure in strategic locations around the world
  - The use of BitTorrent (peer-to-peer data transfers) grew very quickly
- Why?
  - User experience – high latency / low bandwidth means a bad experience
  - Transit costs – sending the same data many times over was expensive (and wasteful)
  - Access providers ran web caches (remember Squid?)
  - Access providers tried to block BitTorrent
  - Content providers built their own distribution infrastructure
  - Emergence of on-ISP concepts such as Google’s Global Cache

# Content is King!

- Emergence of CDNs and Content Providers meant:
  - Less transit traffic to access the “single website” of your favourite content
  - Vastly improved performance because of the distributed infrastructure
    - Low latency, “unlimited” local bandwidth
    - Performance of the content delivery was now dependent on the access network, not the upstream transit to the other side of the planet
  - Pressure for improvements on the access networks, replacement of dialup and low speed copper links with higher speed access using fibre, point to point wireless, and 3G/4G mobile
    - Users in some countries had choices of access technologies

# Death of Transit?

- By 2015 we heard the predictions about the “death of transit”
  - Consolidation of the traditional Tier-1 transit providers was taking place:
    - Example: Global Crossing / Level-3 / Qwest / C&W / … becoming Lumen today
  - Pricing of transit in some locations was cheaper than being a member of the local IXP!
  - Some operators were de-peering at IXPs and simply relying on the super-cheap transit they were offered (despite the technical and policy downsides of having no choice)
  - 80% of all access network traffic was content available by peering

# Content at IXPs

- On the flip side of “death of transit”, content providers and CDNs were connecting to IXPs
  - Note: not every IXP (to the disappointment of some)
  - Not for altruistic support of the IXPs either
  - But because being at the IXP offered them the most reliable low latency path to their content consumers
  - Most implemented either *Open* or *Selective* peering policies
- As traffic volumes across the IXPs grew:
  - Content providers and CDNs encouraged direct cross connect with the IXP members – traffic volume too much to “risk” across the IXP fabric
- This is now the prevailing model today in many parts of the world

# Changing access to IXPs

- With the prevalence of content at IXPs, methods of getting there were changing
  - IPLC was not cheap
    - Some operators did without a protected circuit (saved money)
    - Cable breaks over the last 15 years seem to be more frequent and take much longer to repair!
  - Operators needed an IPLC for every IXP they wanted to get to
- Emergence of the “Network as a Service”
  - These providers were present at many IXPs around the globe and could connect their customer to any IXP as and when the customer required
  - The “point and click” to an IXP access via a convenient dashboard

# Network as a Service

- Regional IXPs partnering with the “Network as a Service” providers helped expand the IXP’s membership further
  - More enterprise networks interested in determining their own peering destiny, rather than relying on an outsourced infrastructure operator
    - Gave rise to a ”new breed” of IXP member and peer
    - Tools such as Peering Toolbox (https://www.peeringtoolbox.net/) to help with the basics
  - Meant that end user network operators (whether ISP or Enterprise) could choose where, when and how to peer, depending on business or end-user needs
    - No more fixed term IPLC contracts
    - No need to manage any IPLC, deal with outages or intermediates, or work with infrastructure providers on infrastructure problems
    - Significant growth again in “Remote Peering”

# Changes to the Transit model

- With the emergence of “Network as a Service”, the commercial traditional transit provider model was under more pressure
  - Consolidation already mentioned
  - Emergence of new multi-national transit providers with novel business models
  - Transit choice not so critical for most access providers – it was the other 10% of traffic to get to “the rest of the Internet”
    - Issues like service quality, reliability, capacity were no longer relevant
    - Price is now the determining factor – the cheapest way to the rest of the Internet

# Transit for R&E

- R&E networks have very different needs from today's commercial Internet
  - R&E eyeballs consume from the CDNs and Content Providers, which is why RENs turn up at IXPs (and are often key motivators in the creation of IXPs)
  - Transit is critical – R&E networks host infrastructure with clever people trying to do novel things, often requiring large network capacity
- Critical for R&E networks today:
  - Local peering: for access to commercial ISPs (staff and students work/study at home) → IXP !!
  - Transit: high bandwidth, minimal latency, short (ideally symmetric) paths
  - Routing: that works efficiently, optimally and reliably, a focus of the Global Network Advancement (GNA) Routing Working Group

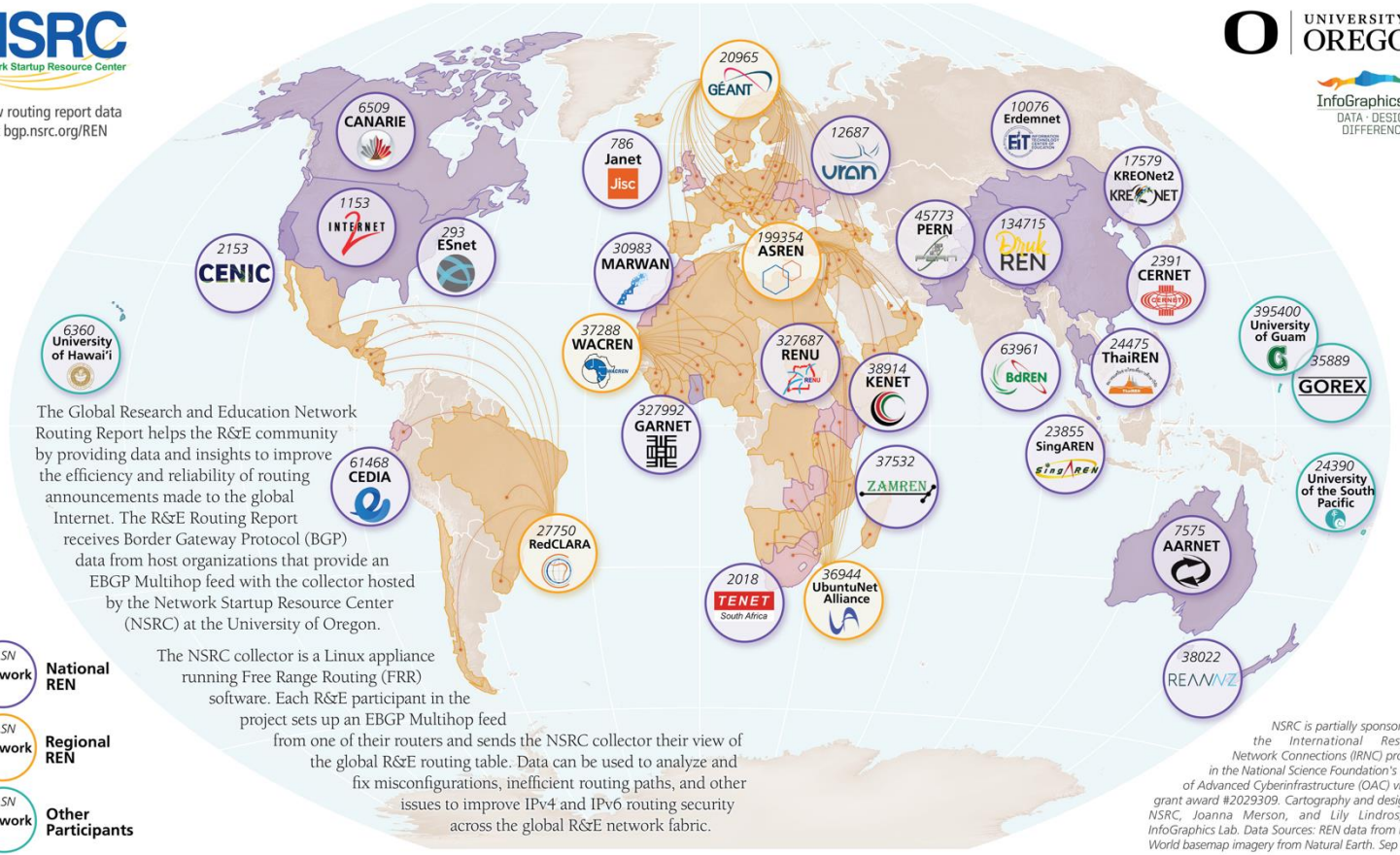
# GLOBAL RESEARCH AND EDUCATION ROUTING REPORT

NSRC
Network Startup Resource Center

View routing report data at bgp.nsrc.org/REN

UNIVERSITY OF OREGON

InfoGraphics Lab
DATA · DESIGN DIFFERENCE




The Global Research and Education Network Routing Report helps the R&E community by providing data and insights to improve the efficiency and reliability of routing announcements made to the global Internet. The R&E Routing Report receives Border Gateway Protocol (BGP) data from host organizations that provide an EBGP Multihop feed with the collector hosted by the Network Startup Resource Center (NSRC) at the University of Oregon.

The NSRC collector is a Linux appliance running Free Range Routing (FRR) software. Each R&E participant in the project sets up an EBGP Multihop feed from one of their routers and sends the NSRC collector their view of the global R&E routing table. Data can be used to analyze and fix misconfigurations, inefficient routing paths, and other issues to improve IPv4 and IPv6 routing security across the global R&E network fabric.

ASN Network **National REN**

ASN Network **Regional REN**

ASN Network **Other Participants**

*NSRC is partially sponsored by the International Research Network Connections (IRNC) program in the National Science Foundation's Office of Advanced Cyberinfrastructure (OAC) via NSF grant award #2029309. Cartography and design by: NSRC, Joanna Merson, and Lily Lindros, UO InfoGraphics Lab. Data Sources: REN data from NSRC. World basemap imagery from Natural Earth. Sep 2024.*

# Situation today

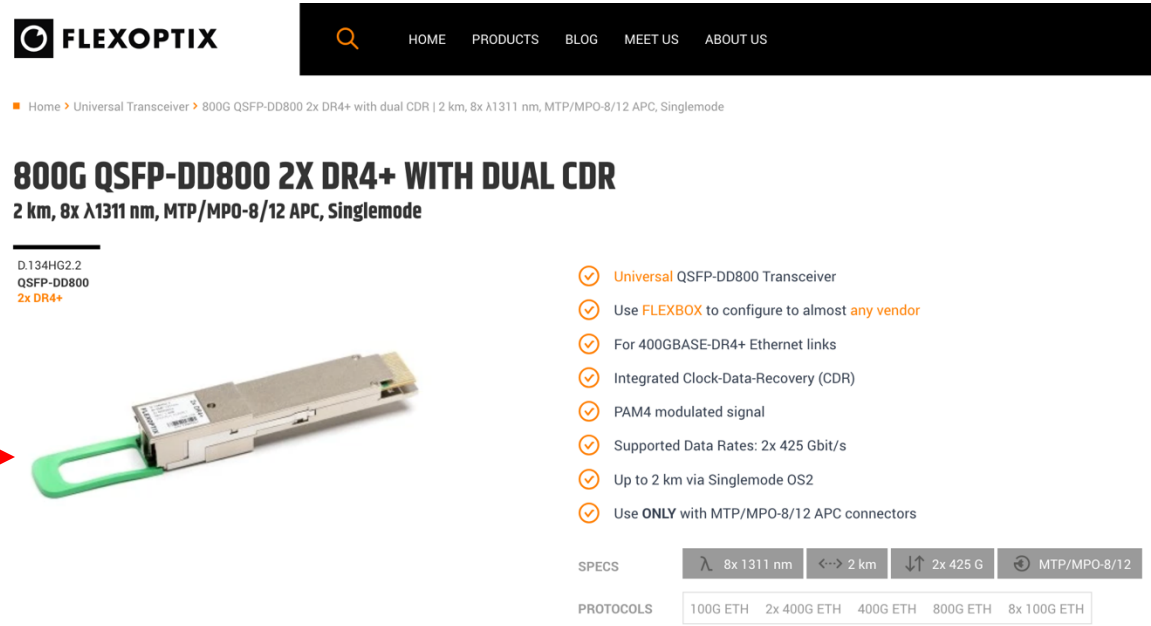
- It is hard to summarise for the entire globe, but:
  - Existing IXPs are still growing
  - New IXPs are still needed in some places
    - Some are being created
    - Others are still being debated (as if?!)
  - CDNs and Content Providers cluster around IXPs
  - CDNs and Content Providers build and run their own global infrastructure
  - “Network as a Service” is an alternative to IPLC for accessing IXPs and other operators
- What comes next?

# End of IXPs?

- Already see situations where biggest content services providers ("Hyperscalers") have too much traffic to be handled at IXPs
  - Now private cross-connect in the same datacentre
  - Backup private cross-connect in another datacentre, geographically separated
  - Desire not to have a third party involved between content and the eye-ball provider
  - Hyperscalers pushing for 400GbE and 800GbE connectivity
  - "Death of IXPs" has been predicted by some

# Faster and Faster and…

- In the past we thought 10GbE was physical limit for fibre infrastructure
- Then we got 100GbE, and those high-end switches are common today
- 400GbE and 800GbE QSFP available now
- IEEE 802.3dj is working on 1.6TbE
  - What will this ethernet switch look like? And cost?



# No end for IXPs!

- Even with faster and faster speeds, there won't be a "Death of IXPs"
  - IXPs will always be needed to keep local traffic local, as in early 1990s
  - They play incredibly important role in the global Internet ecosystem
- But the presence of the Hyperscalers at IXPs?
  - Is the IXP role to manage endless growth in traffic between a content services provider and eye-balls?
  - Or is the IXP role managing quality of interconnect for their members?
  - Depends where on the globe, small IXP vs large IXP, local market, etc

# Connecting with Content

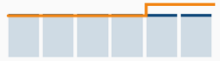
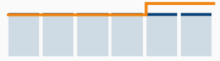
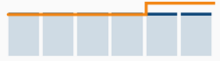
- As traffic levels grow, Hyperscalers will ask/tell their “peers” to connect directly to their infrastructure
  - Built out massive infrastructure globally: submarine fibre, datacentres,…
  - Connection to Hyperscaler is not at the IXP
  - Maybe not even the same datacentre
  - Who pays? What if access provider says “no” ?
- Even more pressure on the access provider
  - Must get best access to content provider for best experience for the end-user
  - In many countries, end-users have lots of choice, and churning (moving between access providers) is easy
  - Costs are back onto the access provider – fibre to the Hyperscaler as well as remaining connected to the IXP

NETFLIX ISP Speed Index https://ispspeedindex.netflix.net/country/australia

# Australia

## Leaderboard

Last Updated: October 2024

| Rank | Speed (Mbps) | ISP ⓘ | Type | Last 6 Months |
|---|---|---|---|---|
| 1 | 3.0 | Aussie Broadband | Fiber Cable DSL Wireless |  |
| | | Dodo/iPrimus | Fiber DSL Wireless |  |
| | | iiNet NBN | Fiber Cable DSL |  |
| | | iiNet Non-NBN | DSL |  |
| | | Optus NBN | Fiber Cable DSL Wireless |  |

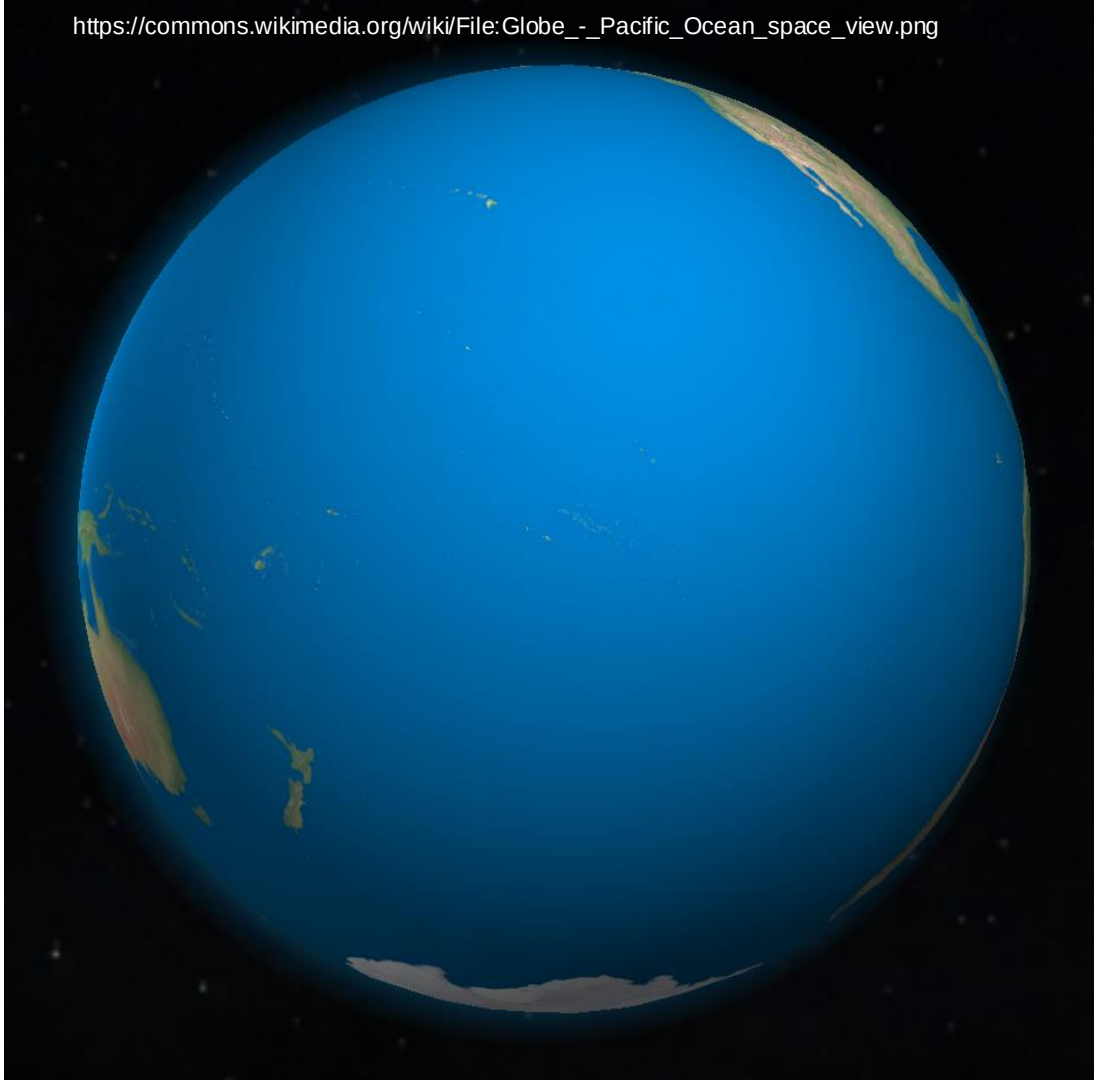
Consumers look at pages like this to choose their provider

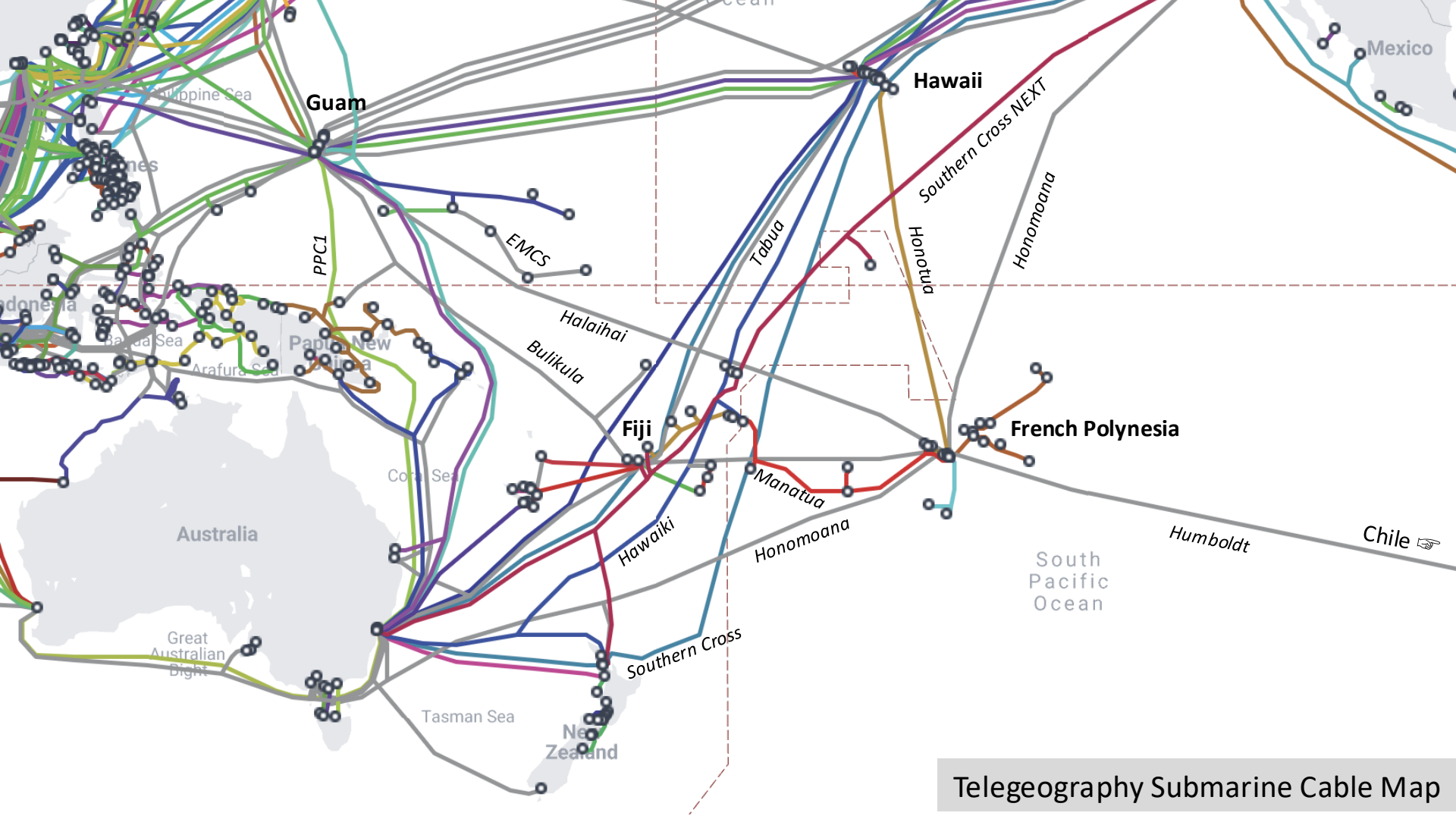
# Guessing the Future

- Impossible to predict the future – even though lots of people are happy to try
- But what **is** happening:
  - Hyperscalers are building massive global infrastructure – submarine fibre everywhere
    - Noticed how long it takes to get submarine fibre cuts fixed these days?
    - Lots of big announcements – cable ships are booked out for years!
  - We still need IXPs
    - Many countries have none, or have dysfunctional interconnects
  - We still need Transit
    - It may not be the dominant data transfer for commercial Internet, but we still need to access content & services not hosted by the Hyperscalers

# The Pacific Ocean

- Made up of:
  - Polynesia (690k people)
  - Micronesia (525k people)
  - Melanesia (13M people)
  - Source: https://www.worldometers.info
- Within next 2 years all Pacific nations will be on submarine fibre infrastructure
- 30 years ago there was only geostationary satellite!




Guam
Hawaii
Mexico
Philippine Sea
PPC1
EMCS
Halaihai
Bulikula
Tabua
Southern Cross NEXT
Honotua
Honomoana
Fiji
French Polynesia
Manatua
Hawaiki
Honomoana
Humboldt
Chile
Australia
Coral Sea
Arafura Sea
Great Australian Bight
Tasman Sea
New Zealand
Southern Cross
South Pacific Ocean
Telegeography Submarine Cable Map

# New interconnects: Pacific Example

- Hawaii & Sydney are well established as interconnects
- Guam is becoming a very significant interconnection location
  - MARIIX is first neutral IXP there – but submarine systems go straight through Guam without stopping (at L3)
  - Guam IX (by GuamExchange) is first real effort at attempting L3 interconnect for the multiple submarine fibre landings on Guam
  - New datacentres being established – intra-island fibre is still very expensive!
- Fiji is starting to look interesting as an interconnect
  - But the “IXP” there is closed (not open or neutral, run by the regulator), only for the incumbent telcos in Fiji + University of South Pacific – a tragedy for the South Pacific
- French Polynesia is starting to look interesting as an interconnect
  - Australia & New Zealand to Chile and the rest of South America, anyone?
  - But no IXP

# And then what?

- With the submarine fibre growth, every country on the planet will have highspeed Internet access
  - Potential anyway, local access network determines what really happens
  - Too often local access is still stuck with last century technology, despite all our highspeed infrastructure advances globally

- I've purposely not mentioned LEO satellite constellations
  - Apart from ruining the night sky and ground-based astronomy… 😔
  - Too susceptible to solar activity, exponential growth in orbital junk, and global politics to be relied upon any more than as a stop-gap pending the arrival of fibre infrastructure, or for backup of critical transmission infrastructure

# And finally

- Future for IXPs is bright
  - The long-established & experienced IXPs, like LINX, have a powerful future leadership role helping local communities establish new local interconnects
- Information delivery is being ever-more optimised
  - CDNs and content creators will always use the best way of getting content to their consumers
- The limiting factor in all this is the information consumption ability of humans! 🙄🤔

# PostScript: 30-year questions 😀

- What will the IXP fabric look like on LINX's Golden Anniversary?

- Will any IX want to handle zettabytes of traffic? And how?
- What will quantum computing and quantum networking mean for an IX?
- Dare I mention AI (aka machine learning) and its traffic plus operational impacts on an IX?