Fiber route maps in the conference record: extraction and georeferencing, with first results

Internet History Initiative*

July 2026

Summary

Operators have been putting maps of their fiber builds on conference slides for thirty years. Most of those maps were shown once and filed away. This note describes an attempt to find them in a large archive of network-operations conference material, to judge which ones carry real geographic information, and to trace that information back out into standard geographic form (KML) aligned against present-day OpenStreetMap data. It also reports, candidly, how far automated methods got and where they stopped.

The short version: from roughly 1.2 million archived slide and document pages we identified 1,708 route-map slides, of which 54 passed a strict screen for georeferencability. A pipeline of image processing and model assistance traced the drawn routes and, for a subset, produced KML aligned to within roughly 50–150 meters. The rest of the candidates have their routes extracted in pixel form but need a person to place control points. Every automated output here should be treated as a draft for human review, not as an authority.

1 Source material

The maps come from the Internet History Initiative's reading-room collection: presentation decks and proceedings from operator and standards meetings (NANOG, RIPE, IETF, APNIC/APRICOT, AfPIF, SANOG, UKNOF, ESNOG, and many regional groups), 1985–2026. All pages in the collection carry a short model-written description of their visual content, produced during accessioning.

Selection ran in three passes, each narrower and more careful than the last:

1. A text search over the page descriptions for map-like content produced 3,806 candidate pages.
2. A vision-model triage over the page images kept 1,708 that actually depict fiber or cable routes (the rest were logical topology diagrams, coverage cartograms, and ordinary illustrative maps).

*Prepared within the IHI reading-room program; methods questions welcome via `history.cooperate.social`.

3. A second, stricter scoring pass asked a narrower question: could a cartographer produce a useful KML by tracing this image against real map data? The scoring rubric was calibrated against a human sample — the curator's estimate, from paging through the set, was that only one slide in fifty or so is genuinely actionable. The final screen (a score of 90 or above, with the model's "actionable" flag set) kept 54 slides.

Two honest caveats about this selection. First, the model's place identifications are advisory only: it labeled a downtown Chicago street map as Milwaukee, plausibly because the talk came from a Wisconsin university. Named streets and landmarks on the map are more reliable than the model's summary of where the map is. Second, the screen is certainly not exhaustive — maps drawn in unusual styles, or in the yellow/orange range that basemap roads occupy, are under-selected.

2 Tracing the drawn route

The route extraction is ordinary image processing, and it works well on this material precisely because the material is ordinary: a presenter draws their fiber in one saturated color over a lighter basemap.

The steps: find the map region on the slide (the largest non-white rectangle); mask pixels near the route's hue with sufficient saturation (the hue is chosen automatically by testing several candidates and scoring how "stroke-like" the resulting ink is — a route is long and thin, water fill and tinted regions are not); discard components that are neither large nor elongated (legend swatches, markers); thin the mask to a one-pixel skeleton; walk the skeleton into ordered polylines, preferring to continue straight through junctions; and simplify. Closed small loops are dropped — on slides these are almost always boxes and markers, not fiber.

Known weaknesses, observed in practice: dashed or faint strokes break into fragments; routes drawn in yellow or orange are excluded from automatic hue selection because basemap arterials use those colors; text or symbols drawn over the route in the same color merge into it. The per-map overlay images in this bundle (`*.qc.png`) show exactly what was traced, and are the fastest way to judge each trace.

3 Georeferencing

This is the interesting part, and the part that resisted automation longest. The goal is a transform from slide pixels to coordinates. For maps of the screenshot era the true model is known: web-map screenshots are Web Mercator, north up, so a similarity transform (scale and translation, no rotation) is the right family, and we fit exactly that. Assuming north-up is not a small convenience — it eliminates whole families of wrong answers.

Control points

A vision model proposes control points: street intersections and named point landmarks it can see on the map. We learned quickly not to trust these proposals naively.

- Asked once over the whole slide, the model drifts into plausible invention — it proposed a tidy series of numbered-street crossings, most of which are not printed on the map at all, and placed them on a line. (Collinear control points cannot constrain a map transform, so this failure is doubly damaging.) Asking separately about each quarter of the map, with instructions to use only labels it can actually read, spreads the points and reduces invention.

- Every proposed point is then independently verified: we crop a small window around the claimed position and ask again whether the feature is there. Invented points tend to fail this check and are dropped. Pixel positions from the small crop are also simply more accurate than positions claimed against the full slide.
- Because proposals vary from run to run, we pool several proposal rounds and take the median pixel position per feature.

Control points are turned into coordinates using OpenStreetMap services, politely and with attribution: named landmarks through Nominatim, street intersections by fetching the named ways from Overpass in a single query and computing the crossing of their centerlines geometrically. Computing crossings ourselves matters for two reasons: Nominatim does not resolve intersections, and grade-separated crossings (an expressway over an avenue) share no OSM node yet are perfectly good control points on a drawn map. Route numbers ("I-275", "60") are matched against OSM `ref` tags rather than names. A name-match whose crossing points spread over more than 250 m is discarded as ambiguous rather than averaged.

One archival lesson worth recording: landmark names rot. A Tampa arena that appears on a 2010s slide no longer exists under that name in OpenStreetMap, having been renamed in 2025. Street intersections are the durable currency for this work.

Fitting, and a trap

With control points in hand we fit the north-up similarity robustly (RANSAC over point pairs, rejecting any hypothesis rotated more than a few degrees or at an absurd scale), then refine against the map's own geometry: we build a mask of the basemap's drawn arterial ink, take its distance transform, and adjust the transform so that real OSM road geometry, projected onto the slide, lands on drawn roads (chamfer alignment, scoring only the best-matching portion since no slide draws every arterial).

The trap: street grids are self-similar. Pure geometry matching confidently locked onto an alignment 25% off in scale — every road landed on *a* road, just not the right one, and every control point disagreed by kilometers. The remedy is to let the two sources of evidence constrain each other: control points anchor the neighborhood of the answer, road geometry sharpens it. When the control-point fit and the chamfer refinement agree on scale independently, as they did on the pilot map, that agreement is itself the best available evidence the alignment is right.

On the pilot (a Tampa metro fiber build drawn street-by-street over a Google Earth screenshot), the result aligns to real geometry within roughly 50–150 m along the route. Residuals per control point are recorded in each map's `georef.json`.

4 Results over the candidate set

Of the 54 candidates, 3 aligned automatically to KML, 45 were traced but not aligned, and 6 produced no usable trace.

event	year	talk	place (advisory)	status
arin-on-the-road-madison-2014	2014	Barford Atlas	Chicago, Illinois (Loop and South Side	aligned to KML
nanog-68	2016	Wednesday.General.Barford.Atlas.23	Chicago, Illinois (Loop and South Side	aligned to KML
nanog-94	2025	Backhoe Fade Fault Location	Tampa, Florida	aligned to KML
afpif-2024	2024	Introducing The Open Fibre Standard	Johannesburg, South Africa	traced only
afpif-2024	2024	Introducing The Open Fibre Standard	Kyiv, Ukraine	traced only
afpif-2024	2024	Introducing The Open Fibre Standard	Maputo, Mozambique	traced only
aims-5	2013	Internet Atlas: A Geographic Database of the	Milwaukee, Wisconsin	traced only
apnic-54	2022	A Realignment of Submarine Cables in Asia and	Tarawa, Kiribati	traced only
apricot-2024	2024	Intelligent IoT for Sustainable Development G	la Bisbal d'Empordà, C-66 road, Catalo	traced only
arin-40	2017	Hagens Zayo	San Jose, California	traced only
ausnog-2025	2025	Ausnog 2025 01 10 Craigaskings	Melbourne Docklands, Australia	traced only
esnog-33	2025	isoc steve song open fibre data standard	Johannesburg, South Africa	traced only
esnog-33	2025	isoc steve song open fibre data standard	Lagos, Nigeria	traced only
esnog-33	2025	isoc steve song open fibre data standard	Gaborone, Botswana	traced only
hknog-4.0	2017	BevanSlattery HKNOG 4.0 TKO Express Undersea	Sydney Olympic Park, Sydney, Australia	traced only
ietf-104	2019	Fibre Network Deployment Paradigms - Xavi Poz	Barcelona, Spain (specifically the San	traced only

event	year	talk	place (advisory)	status
ietf-104	2019	Best Current Practice on Alternative Operator	Barcelona, Spain (specifically the San	traced only
ietf-109	2020	mdi-qkd-quantum-internet	Delft and The Hague, Netherlands	traced only
ietf-120	2024	Tribal Broadband Bootcamps - Strengthening So	RantanenTown Ranch, El Rancho area	traced only
ietf-124	2025	The Open Fibre Data Standard: Shedding Light	Johannesburg, South Africa	traced only
ietf-125	2026	Connecting the Unconnected: It Takes A Villag	Pipidan, Brgy. Pipidan, Philippines	traced only
ietf-126	2026	01 Open Fibre Data Standard (OFDS)	Johannesburg, South Africa	traced only
ietf-126	2026	01 Open Fibre Data Standard (OFDS)	Kyiv, Ukraine	traced only
ietf-126	2026	Chair Slides	Johannesburg, South Africa	traced only
ietf-99	2017	Zenzeleni and other community networks in Afr	Rwenzori Mountains, Uganda	traced only
ix-br-2012	2012	Marcos	Triumph Drive, Independence Drive, Atl	traced only
ix-forum-2020	2020	IX Forum 2020 Panorama Cabos Submarinos no Br	Santarém, Pará, Brazil	traced only
linx-117	2022	Linx117 Zayolightning Temahassan	Manchester, UK	traced only
linx-121	2024	Linx121 Commsworld-cwc Charlieboisseau	Glasgow, Scotland	traced only
linx-2014	2014	Day 1 P2 Telehouse	London Docklands (Canary Wharf area)	traced only
nanog-68	2016	Monday General Barford Atlas 10	Chicago, Illinois (Downtown/Loop area)	traced only
nanog-92	2024	To Bng Or	Ann Arbor, Michigan (Washtenaw County)	traced only
nanog-93	2025	Backhoe Fade Fault Location	Tampa, Florida	traced only
ripe-64	2012	RIPE Neo 2012	Paris, France	traced only

event	year	talk	place (advisory)	status
ripe-88	2024	Open Fibre Data Standard RIPE NCC SEE	Johannesburg, South Africa	traced only
ripe-88	2024	Open Fibre Data Standard RIPE NCC SEE	Luanda, Angola	traced only
ripe-88	2024	Open Fibre Data Standard RIPE NCC SEE	Gaborone, Botswana	traced only
sanog-13	2013	Sanog Conference Jugaad Dileep	San Francisco, California	traced only
sanog-27	2013	SANOG Conference Jugaad Dileep	San Francisco, California	traced only
sanog-27	2016	SANOG Conference Jugaad Dileep	Unknown (likely a specific town or dis	traced only
sanog-32	2016	SANOG32 Conference Overview Of FTTx Deploymen	Dhaka, Bangladesh	traced only
sanog-32	2016	Sanog32 Conference Overview Of Fttx Deploymen	Dhaka, Bangladesh	traced only
sanog-32	2016	SANOG32 Conference Overview Of FTTx Deploymen	Dhaka, Bangladesh	traced only
uknof-17	2010	Digital Scotland	Scotland	traced only
uknof-2014	2014	Hare Gigaclear	Fyfield, Essex, UK	traced only
uknof-2020	2020	UKNOF 2020.v2	Balquhidder, Scotland	traced only
uknof-2020	2020	UKNOF 2020.v2	Monachyle, Tuarach, Bealach (Loch Awe,	traced only
uknof-2020	2020	UKNOF 2020.v2	Mhor 84, Isle of Mull, Scotland	traced only
afpif-2024	2024	Introducing The Open Fibre Standard	Gaborone, Botswana	no usable trace
apnic-60	2021	Setting Up a Telecom-Grade Temporary Conferen	Taipei, Taiwan (near Taipei Main Stati	no usable trace
ietf-124	2025	The Open Fibre Data Standard: Shedding Light	Kyiv, Ukraine	no usable trace
ietf-124	2025	The Open Fibre Data Standard: Shedding Light	Kyiv, Ukraine	no usable trace
ix-forum-2020	2020	IX Forum 2020 Panorama Cabos Submarinos no Br	Amazon River Delta, Pará and Amapá, Br	no usable trace

event	year	talk	place (advisory)	status
menog-6	2010	E An Building A Terabit Internet Exchange 0	Frankfurt am Main, Germany	no usable trace

Where alignment failed, the commonest causes were: the place hint from the triage stage being wrong (the transform is then sought in the wrong city); basemaps without distinct arterial ink (satellite imagery, minimalist cartography) giving the chamfer stage nothing to grip; too few verifiable labeled intersections; and maps whose geography is real but whose projection is not a web-map screenshot (scanned print cartography, stylized national maps). None of these are dead ends — they are exactly the cases where a person placing three or four control points by hand completes the job, with all the tracing and plumbing already done.

5 A second pass: librarian-placed control points

After the automated batch, we made one more attempt on the most promising of the traced-but-unaligned maps — this time with the librarian reading each map directly and placing control points by hand (well, by eye), while the geocoding, robust fitting, and verification stayed automated. The slides were re-rendered from their source PDFs at twice the working resolution for this pass; several placements that failed at 170 dpi succeeded once the street labels could actually be read.

This rescued five more maps: the Chicago Loop pair (whose "Milwaukee" place hint had sent the automated pass to the wrong city), a second Tampa slide that turned out to be byte-identical to the already-aligned one (the same deck shown a year apart — the transform transfers exactly), a block-level Zayo map of downtown San Jose (the tightest fit of the whole exercise: five of five control points within 18–37 m at 0.4 m/pixel), and a Dark Fibre Africa coverage view of northern Johannesburg, anchored almost entirely on country clubs and an airport, where the control-point fit and the road-geometry refinement agreed on scale independently.

The pass also surfaced two methodological points worth recording. First, not every screenshot is faithful: the Chicago slides had been squashed about 19% horizontally by their author, which no rotation-free similarity transform can fit — the model family had to be widened to a full affine (with sanity bounds), and the overlay format switched to quadrilateral corners so stretched fits render exactly. Second, several automated defenses tuned against model error turned out to fight the librarian: a minimum-scale sanity bound rejected a perfectly good 0.4 m/pixel block-level map, and a duplicate-detection radius collapsed legitimately close control points on the same map. Guardrails calibrated for one error distribution quietly misbehave under another; each had to be made scale-aware rather than absolute.

A third pass returned the vision model to the work under better management: it inventories the map regions on each slide (catching multi-panel layouts and insets), names the place independently of the earlier triage, describes the drawn route's color and style to guide extraction, and harvests labeled features on overlapping high-resolution tiles, with the same geocoding and robust-fitting gates downstream. It contributed one further overlay-grade alignment (a second Johannesburg coverage view) and one instructive failure: a Dhaka metro map whose local exchange names geocoded, unbounded, to random places worldwide — four of which colluded into a fit with innocent-looking residuals. The review practice of inspecting the trace and the fit separately caught it, and geocodes are now rejected outside the target place's bounding box. The lesson generalizes: with enough junk points, some subset will agree; consistency alone is not evidence.

Maps that still resisted: satellite-photo basemaps whose only anchors are town and campus centroids (a Delft–Den Haag quantum-fiber route stayed just short of consistency), engineering charts with a single geocodable feature, and web-portal screenshots where the drawn network is dense coverage rather than a traceable route. The final tally: nine of the 54 candidates aligned to draft KML — three fully automated, five with human control points, one model-assisted overlay-grade — and the accompanying tour file carries the five with distinct, verified geometry.

6 Cautions on use

These maps are historical documents, not engineering data. A route on a conference slide is what an operator chose to present, at a moment in time, at whatever simplification served the talk. Alignment accuracy is bounded by the slide's own drawing accuracy well before it is bounded by our fitting. The KML files are drafts produced by the process described above; nothing here has been field-verified, and the material should not be relied on for anything operational. Each map in the index links back to its source talk in the reading room so the geometry can always be read in its original context.

7 What is in this bundle

- `METHODOLOGY.pdf` — this document.
- `ihi-fiber-tour-2026-07.kmz` — open in Google Earth: the aligned maps, each slide ground-overlaid at its fitted position with the traced fiber on top, linked to its source slide in the reading room.
- `candidates.csv` — the 54 candidates with event, year, talk, model-proposed place (advisory), pipeline status, and source URL.
- `maps/`*`name`*`/` — per map: the slide image; the traced route in pixel coordinates (`route_px.json`) with its overlay check image; and, where alignment passed, `route.kml`, the fitted transform with control points and residuals (`georef.json`), and a plot of the route against OSM road geometry (`*.qcgeo.png`).

8 Where this could go next

A light human-in-the-loop step — clicking three or four intersections per map — would carry most of the traced-only maps to aligned status; the automation already does everything after that. Older material (scanned print maps, hand-drawn national backbones) needs projection handling beyond the north-up screenshot assumption. And the natural destination for verified geometry is an open schema for fiber data, where these traces could sit alongside present-day open infrastructure datasets as historical layers. We would welcome collaboration on any of this.