Open Rail Data: The Journey

Peter Hicks - @poggs

OPENTRAINTIMES

About me

- Sector Lead (Rail) at Rockshore Ltd., experts in real-time information systems
- Former Network Engineer, 15+ years at big-name companies
- Passionate about data and making it available
- Runs the openraildata-talk mailing list

OPENTRAINTIMES

Why Open Rail Data?

- Rail user for 20+ years
- I’m inquisitive
- To help everyone understand the railway a bit better
- To make people feel happier and have a better time

OPENTRAINTIMES

Why “a journey”?

- “Open” is not binary
  - A sliding scale
  - Lots of small steps to get there
- “Open” is an evolution – this is new territory

OPENTRAINTIMES

The Players

- **Network Rail**
  - Operates the infrastructure, not the trains
  - Own TRUST, TD etc.
- **Train Operating Companies (TOCs)**
  - Run trains
- **Association of Train Operating Companies (ATOC)**
  - Provide shared services across the industry
- **National Rail Enquiries (NRE) – part of ATOC**
  - Provide passenger information across all TOCs
  - Own Darwin

The beginning

- No Open Rail Data
- I spread word of the NRE Live Departure Boards web service – lots of users!
- API was locked down to enforce the then-vague license conditions
- I felt responsible, wanted to do something to change it, but didn't know how

OPENTRAINTIMES

Working around the problem

- “Can we build our own?”
- Call it an experiment
- Ask Network Rail for timetable data
- “Yes”

OPENTRAINTIMES

TSDB Explorer

- A proof-of-concept – no problem if it fails
- Scary introduction to timetabling
- Working Timetable, Variation, Cancellation and Short-Term Planned schedules
- Great insight in to train planning

OPENTRAINTIMES

What about real-time data?

- Network Rail's TD.net platform
- First person outside the industry to ask for access
- Policy decision…
- "Yes"

OPENTRAINTIMES

Consuming the data

- Lots of it
- Incredibly detailed
- Not documented
- “Learn as you go on”

OPENTRAINTIMES

OpenTrainTimes

- Launched in January 2012
- Train timetable data wasn't yet 'open'
- Flushed out several problems with confidentiality – albeit painfully
- Focal point for getting data open

OPENTRAINTIMES

NETWORK RAIL'S FEEDS

OPENTRAINTIMES

Birth of the Feeds

- News on my work reached many influential people
- An opportunity to talk and ‘take the stage’
- Discussed at the DfT Transport Transparency Board
- Drove the Chancellor of the Exchequer’s decision to make Network Rail open up their feeds

OPENTRAINTIMES

Network Rail Data Feeds

- Built and operated by Rockshore for Network Rail
- Used by several hundred developers per day
- Free at the point of use
- https://datafeeds.networkrail.co.uk/

OPENTRAINTIMES

Network Rail Data Feeds

- Licence is very similar to the OGL
- Community-supported
- Community-written wiki at http://nrodwiki.rockshore.net/

OPENTRAINTIMES

Feeds available

- Timetable (CIF and JSON)
- Late notice timetable changes (VSTP)
- Movements and Amendments (TRUST)
- Network performance (RTPPM)
- Speed Restrictions (TSR)

OPENTRAINTIMES

Downsides

- Records events after they happen for a historical log
- Not designed for passenger information
- Required for a number of industry processes
- We still need Darwin!

OPENTRAINTIMES

THE MIDDLE YEARS

OPENTRAINTIMES

Post-Data Feeds era

- Network Rail became increasingly more open
- Growth in number of Data Feeds users
- The sky has not yet fallen
- The other parts of the industry have to play ‘catch-up’

OPENTRAINTIMES

Summer 2014

- NRE announced they were to open several of their feeds
- Semi-static XML feeds
- Live Departure Boards (original goal)
- Darwin Push Port

OPENTRAINTIMES

ALL ABOUT DARWIN

OPENTRAINTIMES

What does Darwin do?

- Takes TRUST and TD data to make forecasts
- Takes additional input from TOCs on changes not possible in TRUST
- Also takes alterations from Tyrell and information systems at stations
- “Single source of truth” (more in a moment)

What does Darwin feed?

- Live Departure Boards web service
- Online Journey Planner
- Some station information boards
- …and more

OPENTRAINTIMES

Why is consistency important?

- Forecasts are indicative never accurate 100% of the time
- Station board says train is 15m late, a non-Darwin powered app says train is 5m late… **who do I trust?**
- Station board says train is cancelled, non-Darwin powered app says it's running… **who do I trust?**

OPENTRAINTIMES

Consistency is not Accuracy

- Get the data open, then discuss its accuracy
- When it's open, we can report where it's inaccurate
- Fix the problem once for everyone, everywhere

OPENTRAINTIMES

DARWIN FEEDS

OPENTRAINTIMES

What feeds are there?

- Live Departure Boards API
- Push Port
- Darwin Timetable

OPENTRAINTIMES

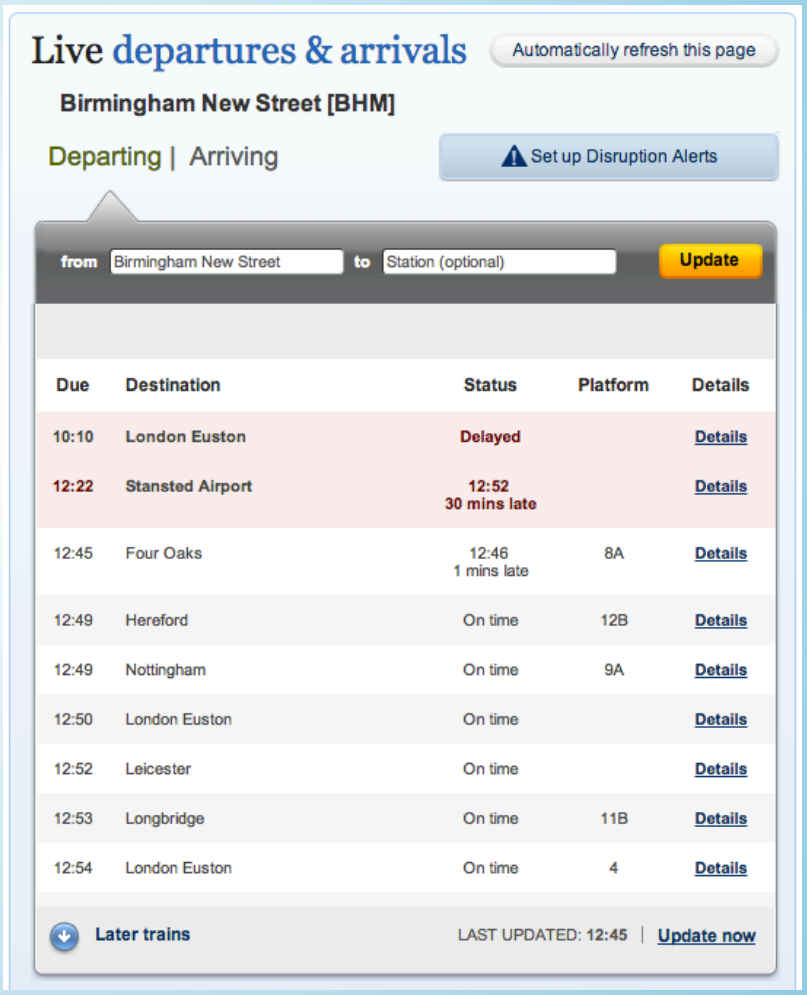
What's the LDB API?

- SOAP API
- Request/response
- 'Secured' via a per-developer token
- Self-service signup

OPENTRAINTIMES

What's the LDB API for?

- One-shot "trains at X"
- Standalone
- Reactive
- Suitable for the most common use-cases



OPENTRAINTIMES

What's the Push Port?

- A stream of data
  - Train movements and predictions
  - Cancellations
  - Timetables and changes
  - Station messages
- Not a silver bullet
  - Requires processing and knowledge
  - Easier than the Network Rail feeds
  - Less detailed than the Network Rail feeds

OPENTRAINTIMES

What's the Push Port for?

- Responding to events
- Enables 'push' services such as train cancelled notifications
- 'Big Data' analysis

OPENTRAINTIMES

What's the Darwin Timetable?

- Only the timetable data from Darwin
- 'Filtered push port'

OPENTRAINTIMES

What's the release schedule?

- LDB API – available now
- Push Port – available by 31st March 2015
- Timetable – available after the Push Port

OPENTRAINTIMES

The End...?

OPENTRAINTIMES

What else is left?

- Regular updates to fares data from RSP
  - Available to industry
  - Currently expensive
- Regular updates to the Routeing Guide
  - In machine-readable format, not PDF
  - Eyeballs will fix the bugs in it

What else is left elsewhere?

- Geospatial and mapping data
- Real-time data from OIS, TMS, LINX
- Train loading data
- Rolling stock data
- Ticket barrier data
- Disaggregated performance data
- Structured data from DfT, ORR

OPENTRAINTIMES

How do we get this data?

- Show that the community will voluntarily ‘do the right thing’
- Demonstrate tangible uses for the data as well as other possibilities
- Understand limitations and work with the industry on concerns

OPENTRAINTIMES

The End

OPENTRAINTIMES