

# The Great Data Challenge Our Least Valued but Most Valuable Asset

Nick Illsley  
14<sup>th</sup> April 2011

# Joining-up the modes .....

- Transport Direct joins up the Modes
  - ◆ Rail
  - ◆ Bus
  - ◆ Coach
  - ◆ Light Rail
  - ◆ Car
  - ◆ Ferry
  - ◆ Air
  - ◆ Walking
  - ◆ Cycling
- Building on Individual Systems
- Providing the Glue



Train  
2 hours 57 mins



Walk  
4 mins



Bus Q1  
6 mins



# ...and location-based travel planning

- The complexity of life requires awareness of location and the ability to travel between places
  - ◆ Where I live
  - ◆ Where I work
  - ◆ Where I shop
  - ◆ Where my children go to school
  - ◆ Where I get healthcare
  - ◆ Where my family and friends live
  - ◆ Where I go for leisure
- Where we are is inevitably a compromise, and mobility is the key to reducing hassle!



# A Truism?

- We all know where we are?
  - ◆ Great Minster House
  - ◆ 76 Marsham Street
  - ◆ SW1P 4DR
  - ◆ Pimlico (nearest tube)
  - ◆ Westminster Borough
  - ◆ London
  - ◆ South East
  - ◆ England
  - ◆ United Kingdom
  - ◆ Europe etc

# User “Language”

- How does a user define their needs?
- Place gazetteers
  - ◆ Addresses / Postcodes
    - 33 million individual entries
  - ◆ Places of interest
    - Hotels, hospitals, sports grounds, etc, over one million
    - Translated into Ordnance Survey Grid References (co-ordinates)
  - ◆ Transport nodes
    - Over 330,000 From NPTG or NaPTAN gazetteers
- Maps
  - ◆ Select points from locality maps

# Key Data Challenge

- Universally seen as key issue
  - ◆ Data Standards
  - ◆ Data Collection
  - ◆ Data Management
  - ◆ Effective Feedback from Fault Identification
- Consensus Required
  - ◆ Adopted NaPTAN as industry standard
  - ◆ Stop level timing is reasonable 21st century requirement
  - ◆ Electronic process from data creation through to customer
  - ◆ Standard languages and formats across all modes
- Maintenance Important as well as Collection

# The Accuracy Challenge

- Measurement and Correction
  - ◆ Every Issue=Error until resolved
- 30 million to 30 million = 1 trillion
- 300,000 to 300,000 = 100 billion
- 99% Accuracy = 1 billion wrong answers
- Early Press Comment
- Data Chain Management
- End of Paper Timetables
- **Bad Data = Wrong Answer**

# Initial Data Sources

- Data Agreements with all crucial suppliers
  - ◆ including
    - Traveline regions
    - Rail Industry (including real time information)
    - Coach Industry
    - Highways Agency
    - Ordnance Survey
    - Car Parks
  - ◆ Hundreds of Individual Sources
- Formal data agreements and schedules
- Retail Agreements also required for fares etc

# Data consistency

- Data formats and exchange protocols are not enough
- Information to be shared must be consistent
- NaPTAN - National Public Transport Access Nodes
  - ◆ bus/ coach/ tram stops
  - ◆ rail/ metro/ underground stations
  - ◆ ferry ports/ airports/ taxi ranks
  - ◆ entry per stop post/ multiple entrances to stations
  - ◆ at 1m accuracy
- NPTG - National Public Transport Gazetteer
- Move from timing-point to stop-level data

# Format & Exchange Standards

- TransXChange – establishing mechanisms to share and exchange data between transport operators / authorities
- JourneyWeb - a communications protocol to support a federal system of electronic journey planners
- RTIG - Real Time Information Group Standards e.g. SIRI
- Consistent with Transmodel, EU Adoption
- Global Standards – Google Transit
- Fit with other standards, GIS, Communications, Internet etc

# Complex airport



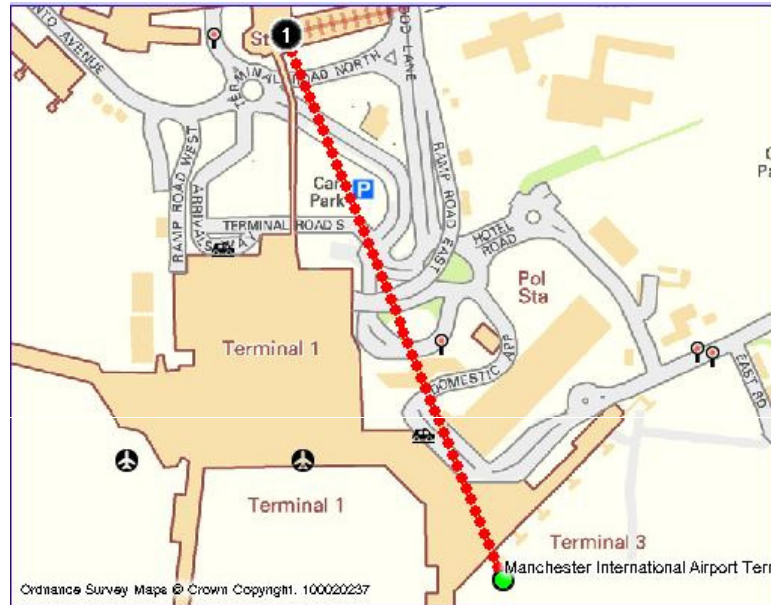
Each terminal/ bus stop & rail/ metro station is included

# Simple airport



Remote airport – no nearby bus stop – so no public transport connection

# Air interchange



Connection between terminal/ bus stop & rail/ metro station is included as part of journey

# Flight/ ground interchange



Specific walk times between arrivals halls & each bus stop/ rail station

# Missing bridge at Paddington

transport direct

The screenshot shows a web browser window displaying the transport direct website. The page title is "Location map" and "London Paddington Rail Station". The main content area features a map of the Paddington area in London, with a red line indicating a missing bridge. The map includes labels for "Paddington Station", "London Paddington Rail Station", and "Bayswater". The map is surrounded by a blue border with navigation controls like "Zoom in", "Zoom out", and "New search". To the right of the map is a "Map Symbols" panel with a list of categories and their corresponding icons. Below the map is an "Overview map" showing the location of the main map area within the United Kingdom. The bottom of the page has a blue footer with navigation links and the transport direct logo.

Location map Printer friendly

London Paddington Rail Station

Plan a journey | Select new location | Help

Zoom in | Zoom out

New search

Choose one of the options (above), change the map view or select symbols to show on the map.

Clicking on the map will zoom and centre

**Map Symbols** | Help

- Transport**
  - Rail stations
  - Underground/Metro
  - Bus/Coach stops
  - Airports
  - Ferry terminals
  - Taxi ranks
  - Car parks
- Accommodation, eating & drinking
- Sport, entertainment & retail
- Attractions
- Health
- Education
- Public infrastructure

Show selected symbols

Overview map

Map legend 1:8000

transport direct

Map legend

1:8000

[Home](#) | [FAQ](#) | [About us](#) | [Contact us](#) | [Sitemap](#) | [Welsh/Cymraeg](#) | [Related sites](#)  
[Terms & conditions](#) | [Privacy policy](#) | [Data providers](#) | [Accessibility](#)

v9.10.2

Connecting People to Places

# Bridge is now in ITN

transport direct

Outward journeys for Mon 10 Mar 08 leaving after 10:35

Option	Transport	Changes	Leave	Arrive	Duration	Select
1	Walk	0	10:35	10:45	10 mins	<input type="radio"/>
2	Car	0	10:35	10:45	10 mins / 0.7miles	<input checked="" type="radio"/>

Maps: Outward journey 2 (Car)

Map of entire journey

+ [Progress Bar] -

Clicking on the map will zoom and centre

**Key**

- High traffic
- Medium traffic
- Low traffic
- Traffic unknown
- Start location
- End location
- Toll booth
- Ferry

**Map Symbols**

- Transport
  - Rail stations
  - Underground/Metro
  - Bus/Coach stops
  - Airports
  - Ferry terminals
  - Taxi ranks
  - Car parks
- Accommodation, eating & drinking
- Sport, entertainment & retail
- Attractions
- Health
- Education
- Public infrastructure

Overview map

Ordnance Survey Maps © Crown Copyright. 100020237

Start | Novell GroupWise... | Transport Direct - ... | Transport Direct... | Removable Disk (F:) | Microsoft PowerP... | 10:42

# Missing new road (A4146)

transport direct

Location map Printer friendly

Wing, Buckinghamshire

[Plan a journey](#) [i](#) [Select new location](#) [Help](#)

[Zoom in](#) [Zoom out](#)

[New search](#)

Choose one of the options (above), change the map view or select symbols to show on the map.

+ [Progress bar] -

Clicking on the map will zoom and centre

**Map Symbols** [Help](#)

You can only view map symbols on the map in the highest five zoom levels. (outlined in zoom scale)

**Symbols can show:**

- Transport stops
- Accommodation, eating & drinking
- Sport, entertainment & retail
- Attractions
- Health
- Education
- Public infrastructure

**Overview map**

Map legend 1:80000

Home | [FAQ](#) | [About us](#) | [Contact us](#) | [Sitemap](#) | [Welsh/Cymraeg](#) | [Related sites](#)  
[Terms & conditions](#) | [Privacy policy](#) | [Data providers](#) | [Accessibility](#)

v9.10.2

# New A4146 is in ITN

transport direct

Outward journeys for Mon 10 Mar 08 leaving after 10:45

Option	Transport	Changes	Leave	Arrive	Duration
1	Car	0	10:45	10:57	12 mins / 8.3miles

Maps: Outward journey 1 (Car)

Map of entire journey

You can view different stages of the journey by selecting them in the list and clicking 'Show Route'. Choose one of the options (above), change the route view or select symbols to show on the map.

+ [Progress bar] -

Clicking on the map will zoom and centre

**Key**

- High traffic
- Medium traffic
- Low traffic
- Traffic unknown
- Start location
- End location
- Toll booth
- Ferry

**Map Symbols**

You can only view map symbols on the map in the highest five zoom levels. (outlined in zoom scale)

Symbols can show:

- Transport stops
- Accommodation, eating & drinking
- Sport, entertainment & retail
- Attractions
- Health
- Education
- Public infrastructure

Overview map

Map legend

1:92107

# Good Locational Data

transport direct



transport direct

transport  
direct.info  
Connecting People to Places

Oops!



# So What's So Hard?

- National Level
  - ◆ Global Issues
  - ◆ Climate Change
  - ◆ International Competitiveness
  - ◆ Exports and Imports
  - ◆ 24x7x365
- Driven by Governments and Groupings e.g. EU
  - ◆ Airports and Slots
  - ◆ Low Cost Airline Growth
  - ◆ Relative Subsidies and Costs
- Broad Policy Objectives
  - ◆ Road Pricing
  - ◆ Taxation
- Data and Knowledge Important?
- How hard at EU Level!
- Individual Decisions
  - ◆ Millions of Choices
  - ◆ Modal Decisions
  - ◆ Accessibility
  - ◆ Personal Mobility
- Driven by Local Authority, Operators and Culture
  - ◆ Local Policy
  - ◆ Commercial Activity
  - ◆ Town Centre Policy
  - ◆ Peer Group Pressure
- Detailed and Minor Factors
  - ◆ Car Ownership
  - ◆ Subsidised Travel
- Information and ITS Important?
- Individuals are still the target audience!

# EU Directives

## **INSPIRE**

- Covers the Infrastructure for Spatial Information across Europe
- Transport Networks in Annex 1
- Supports many things including journey planning
- Linear Elements with Nodal Connectivity
- Intermodal/Cross Border Connectivity
- Freely Available (not necessarily free)

## **ITS and Urban Mobility Action Plan**

- Deployment of ITS across EU
- Targets Road Traffic and Travel Information Services
- Builds on INSPIRE and Digital Mapping Services

# DfT Key Initiatives

- Deficit Reduction
- Localism
- Creating Growth, Cutting Carbon
- Smart Ticketing
  - ◆ Most Journeys Enabled by end 2014
  - ◆ Ten Largest Urban Areas
- End-to-end Journeys
  - ◆ Value of End Legs, especially when Unfamiliar
- Transparency Agenda
  - ◆ Holding Departments to Account
  - ◆ Value for Money from Services
  - ◆ Data Free of Charge
  - ◆ Enable Users to make the Right Choices

# Data, data everywhere

- Everything is becoming a sensor
- Great opportunity for economic filling of data gaps
- Web 2.0, mobile phone tracking, vehicle tracking, CCTV, loops, sensors.....
- BUT
  - ◆ Can we identify what is useful?
  - ◆ Can we assess quality?
  - ◆ Can we merge and mash and interpret?
  - ◆ What do we archive, what do we throw away?
  - ◆ Do we charge or give away?
  - ◆ What can users have access to and amend?

# My Potential Future Model

- Framework
  - ◆ DNF Based as Attributes to Links
  - ◆ Searchable Repositories and View Service (semantic web based)
  - ◆ Associated Tools and Facilities
- Data Needs Identified
  - ◆ Major drivers assessed
  - ◆ Potential secondary and tertiary usage enabled
  - ◆ Who has what? Who needs what?
- Governance and Ownership Established
  - ◆ National Body
  - ◆ Associated Codes of Practice
  - ◆ Level of Guarantee

# Public and Private Sector Roles

- Need to set some general guidelines that foster joint working and maximise efficiency and effectiveness:
- Public Sector
  - ◆ Standards
  - ◆ Collation and audit of data
  - ◆ Spine of core data
  - ◆ Addressing market failure
- Private Sector
  - ◆ Customer facing services
  - ◆ Customer care
  - ◆ Revenue generation
- Individuals
  - ◆ Personalisation
  - ◆ Innovation
- Not fixed but general areas of expertise and interest

# Efficiency and Re-use

- Create Data Once – Use Many Times
  - ◆ Local Authorities
  - ◆ Transport Operators
  - ◆ Performance Measurement
  - ◆ Real-time Systems
  - ◆ Journey Planning Systems
  - ◆ Third Parties and Data Mashing
  - ◆ Data Accuracy therefore Crucial
- Exploit Back Office Value
  - ◆ Open interfaces and common protocols
  - ◆ Engage with wider public and private sectors
  - ◆ National virtual dataset
  - ◆ World class not just world first delivery

# What Others May Do With It

- Data Ownership and Knowledge = Power
- Little Incentive or Desire to Share Data
- IT did not always aid the Correlation of Disparate Data
- New Approach
  - ◆ Past Failures
  - ◆ Power of Information
  - ◆ Web 2 – the Semantic Web
- Primary User – creates for own purpose, describes and Values (metadata)
- Secondary and Tertiary User – combines, assesses, builds

# The Goal?

- Many Hundreds of Data Sources
- Common Data Architecture and Standards
- Efficient and Effective Processing of Data to Knowledge and Information
- Open and Published Interfaces
- Many Hundreds of Channels and Services
- Appropriate Rules and Governance
- Build on existing, federate together, work cross borders, relate to user preference, integrate with wider information services

# Conclusions

- Data Matters – Wrong Data = Wrong Answer
- 21<sup>st</sup> Century Front End with 19<sup>th</sup> Century Data Creation
- Create Once Use Many Times
- Comprehensive Public Transport Data and Protocol Architecture
- Roads Data – Strangely Behind the Game
- Significant Improvement and Development
- Electronic Creation is a Major Step Forward
- Share it with Others to Gain Maximum Potential Value!

# Conclusions 2

- These are challenging times
- Government Policy says
  - ◆ There is less money
  - ◆ Localism
  - ◆ Big Society – focus on the individual
  - ◆ Sustainable support for new technology
- New Business Models and Partnerships Required
- Overarching Data Policy Framework
  - ◆ Transparency
  - ◆ Open Data
  - ◆ Support for Innovation