The Data Challenge: Collecting rail statistics in the UK
UNECE Working Party on Transport Statistics
7th June 2017
Ashley Goddard, ORR
Office of Rail and Road (ORR)

- ORR is a non-ministerial government department
- Independent regulator for Network Rail and Highways England
  - Economic regulator for mainline rail and major road network
  - Safety regulator for all forms of on-rail transport
- We work with infrastructure companies, devolved governments, railway undertakings and other industry bodies to protect the interests of rail and road users, improving the safety, value and performance of railways and roads today and in the future
- We publish data and information about the railway industry as part of the UK government's commitment to transparency, and to improve railway performance.
A complex industry

An overview of the rail industry in Great Britain

- **Government departments provide strategic direction and funding to the railways and procure rail franchises and projects.**

- **European Parliament/European Commission**: "The European Commission are also funders."

- **ORR**: Accountable to Parliament and also to the European Parliament and Commission. "ORR is the economic regulator for the British half of the Channel Tunnel, ARAF the French half."

**The Intergovernmental Commission (IGC)**

- ORR regulates/works with all of these bodies - including safety bodies - across the whole rail industry. As well as TOCs' and FOCs', track access agreements with Network Rail (and, on occasion, with other infrastructure managers, e.g., HS1, ports and terminals).
Publisher of UK rail statistics
Information & Analysis

- ORR has a Memorandum of Understanding with the Department for Transport covering the production, dissemination and publication of statistics about the usage and performance of the National Railways system.

- ORR’s role is to develop a consistent set of data and information for use by Government and the industry for decision making.

- Therefore, ORR collects data from a range of industry bodies, including Network Rail, train operating companies and freight operating companies.

<table>
<thead>
<tr>
<th>Internal use</th>
<th>External use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network regulation, monitoring</td>
<td>Statistical releases, data portal and other publications</td>
</tr>
<tr>
<td>franchise agreements and consumer obligations</td>
<td></td>
</tr>
</tbody>
</table>
Dissemination of data

- Excel data tables (Data Portal)
- Statistical releases (quarterly and annual)
- Analytical reports and industry publications
- User engagement articles
- Dashboards and infographics
- Social media
- Seminars
- Press releases and News By Emails
ORR Data Portal

http://dataportal.orr.gov.uk/
National Statistics

Quarterly:

- **Passenger Rail Usage** *(passenger journeys, passenger kilometres, passenger revenue)*
- **Passenger & Freight Rail Performance** *(PPM by TOC, CaSL by TOC, FDM)*
- **Freight Rail Usage** *(freight moved (ntkms), freight lifted (tonnes), freight delay per 100 train kilometres)*
- **Rail Passenger Complaints** *(Complaints by TOC, Complaints by category)*

Annually:

- **Regional Rail Usage** *(passenger journeys within and between regions)*
- **Rail Finance** *(government subsidy, private investment & rail fares index)*
- **Rail Safety Statistics** *(fatalities, injuries, train accidents on all networks (underground, heritage, mainline))*
- **Rail Infrastructure, Assets and Environment** *(CO₂ emissions, average age of rolling stock, route length (electrified), number of stations)*
Data processing
Data Sources

- Three main mechanisms through which data is collected by ORR
  - Method of supply typically varies by organisation / organisation type

1. Infrastructure and network data
   - Mostly Network Rail (Main Infrastructure Manager – regulated entity)
   - Automatic data supply and processing (mostly)

2. Operational data from railway undertakings
   - Passenger and freight operators
   - Manually requested and collated using data collection templates (Excel)

3. Health and Safety statistics
   - Includes non-mainline and heritage
   - Collected though online web form
Network Rail – What and When?

- ORR can request under a license condition any data or information it needs from Network Rail to carry out its regulatory functions.

- Approx. 60% of the data processed by ORR.

- Data received at regular intervals:
  - Periodic (every 28 days)
  - Quarterly
  - Annually

- Very large volume of data:
  - 1000 data files per year
  - 39 million rows of data
Network Rail – How?

- Memorandum of Understanding (MoU) agreed between ORR and Network Rail

- These provide
  - Agreement of what data is to be received, when, delivery medium
  - Prescribed format for data supply (Excel, csv)

- Review every 12 months to ensure business needs still met
1.1.4. File format: Excel file (.xls)
1.1.5. Data format: Spreadsheet with multiple worksheets
1.1.6. Frequency of provision: Periodic (4 week period)
1.1.7. Date/s file supplied: Received approx 21 days after period end.
1.1.8. Data delivery medium: XLS file uploaded via SharePoint
1.1.9. Layout:
   1.1.9.1. 104 Delays - Header range is A1:L1. See figure xxx below for range/data type information.
   1.1.10. 101 Incident Count - range A1:AJ46. See figure two below for range/data type information:
   1.1.11. Known future activity relating to dataset: None at present
   1.1.12. Number format:
      1.1.12.1. Percentages to be shown as numbers for example, 94.77% should be 0.9477,
      1.1.13. Decimal figures to have the precision format of decimal for example decimal (16,14)
   1.1.14. Sources and notes: Supply by email to ORR beforehand
Network Rail – Challenges in data supply

- Data quality
  - Implausible data e.g. 101% punctuality (validation)
  - Correctness (verification)

- Revisions and changes (expected and unexpected)
  - Data improvements/refinements over time
  - Change in franchise / new operators
Network Rail – Solutions to challenges

Past: R script to check files against a series of business rules
  - Time consuming
  - Extra complexity

Future: Business rules built into the loading process through SQL Server Implementation Services (SSIS)
  - No manual intervention (SSIS part of existing warehouse infrastructure)
  - Ease of pinpointing errors and warning flags
Railway Undertakings

- Some data mandated within franchise agreements - operators are required to provide this data
  - Additional data above and beyond those specified are essentially provided through ‘good-will’

<table>
<thead>
<tr>
<th>Passenger</th>
<th>Freight</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Memorandum of Understanding (with most operators) for particular data sets specific to ORR’s needs and duties</td>
<td>• No MoUs with freight operators</td>
</tr>
<tr>
<td>• Data templates distributed via e-mail to each organisation (Excel)</td>
<td>• Send in their own data (no template)</td>
</tr>
<tr>
<td>• Uniformity of data supply across industry</td>
<td></td>
</tr>
</tbody>
</table>
### Section B - Case totals and response times

(CHP indicators 1 and 2)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Contact method</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of complaint correspondence</td>
<td>Letter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Email/Webform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telephone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meet the Manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Online Forums</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of complaints responded to within complaints handling target</td>
<td>Letter (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Email/Webform (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telephone (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meet the Manager (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Online Forums (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All contact methods (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - Letter</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - Email/Webform</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - Telephone</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - Meet the manager</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - Online forums</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaints handling target - All contact methods</td>
<td>Target (working days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Railway Undertakings - Challenges

■ Consistency between operators
  – Are all understanding the guidance and rules correctly?

■ Timely delivery
  – Defined timescales for supply -> to feed into publication schedule
  – What to do if data not supplied -> what regulatory powers can be used?

■ Processing challenge
  – Staff and time resource for collating and maintaining data sets from different sources
  – E.g. freight data sent in by each operator has to be collated into a template by ORR staff, which can then go into standard data warehouse process
RIDDOR data – Online webform

- ORR (as the railway safety executive) needs to be aware of any events that fall under Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (RIDDOR)

- Webform used by heritage and non-mainline railway undertakings to supply RIDDOR data
  - Data from Network Rail and other major operators collected via other mechanisms
Why use a web form?

- Regular reporting can be a burden for heritage or smaller operators
  - Very few events – most returns would be Nil
  - No other regular data supplies to ORR

- Allows reporting by exception

- Enables consistency at the point of entry
  - Best way of getting good quality data from minor data suppliers and small operators

- Data is fed straight through into ORR Data Warehouse
The RIDDOR webform
Webform data - Challenges

- Pre-defined format does not always capture all information needed
  - Specificities for certain operators may not apply to the data fields
  - ORR can add more to the webform as needed

- Take up of the use of the webform
  - Getting small organisations to move away from paper methods of working
  - Users not always used to filling in the form
  - Confusion about the guidelines or how to complete
  - Skill challenge rather than technical
  - Offer 1:1 support to organisations/users to help with entry of data
The End

■ Any questions?

Ashley Goddard
Senior Statistical Analyst
Office of Rail and Road | One Kemble Street | 2nd and 3rd Floors | London | WC2B 4AN
ashley.goddard@orr.gsi.gov.uk
Tel: 020 7282 3947
Mob: 07714 46 00 33