**Objective**

A UK rail company has commissioned a new database system to track contact from customers through to resolution and allow data analysis to help identify areas requiring service improvement. It is intended to record contact from various communication methods including direct customer input into the company’s website will require validation and cleaning. Open-source data feeds from National Rail should be integrated to aid reliable categorization of the data. The system is expected to replace and incorporate a spreadsheet currently used to log contact.

**Logical Design**

A database solution will collect data and store it in a highly structured format, making it accessible for users managing cases or searching the system, aiding data retrieval for analysis (Bookshear & Brylow, 2020). The database is intended to collect data on public contact flowing into the organisation and store an end-to-end view of the type and subject; organisational ownership of contact; and outcomes/ resolutions of the contact.

The system will hold data in a structured format but the inputs into the system come from multiple sources in different formats. Data directly from National Rail regarding train services and stations is available in semi-structured machine-readable XML format. The legacy customer service data is held in an XLSX, a highly human-readable format that is not easily read by database applications and requires processing before being usable. The data from other internal organisational datasets, such as HR systems, is retrievable in CSV format and considered semi-structured because it does not have the same level of organisation as a database system (EMCS 2015; Olhurst 2012)

The system is intended to follow best practice normal form rules of database design. This involves storing data in related groups per table to avoid duplicating rows within tables and repeating data across the system. The approach improves data consistency; avoids storing redundant data; and minimises insertion, deletion and modification errors as changes are made in one location (Harrington, 2009). The tables in the database each contain IDs, which operate as primary keys allowing association with related tables to join data across the system (Kazil & Jurmul, 2016).

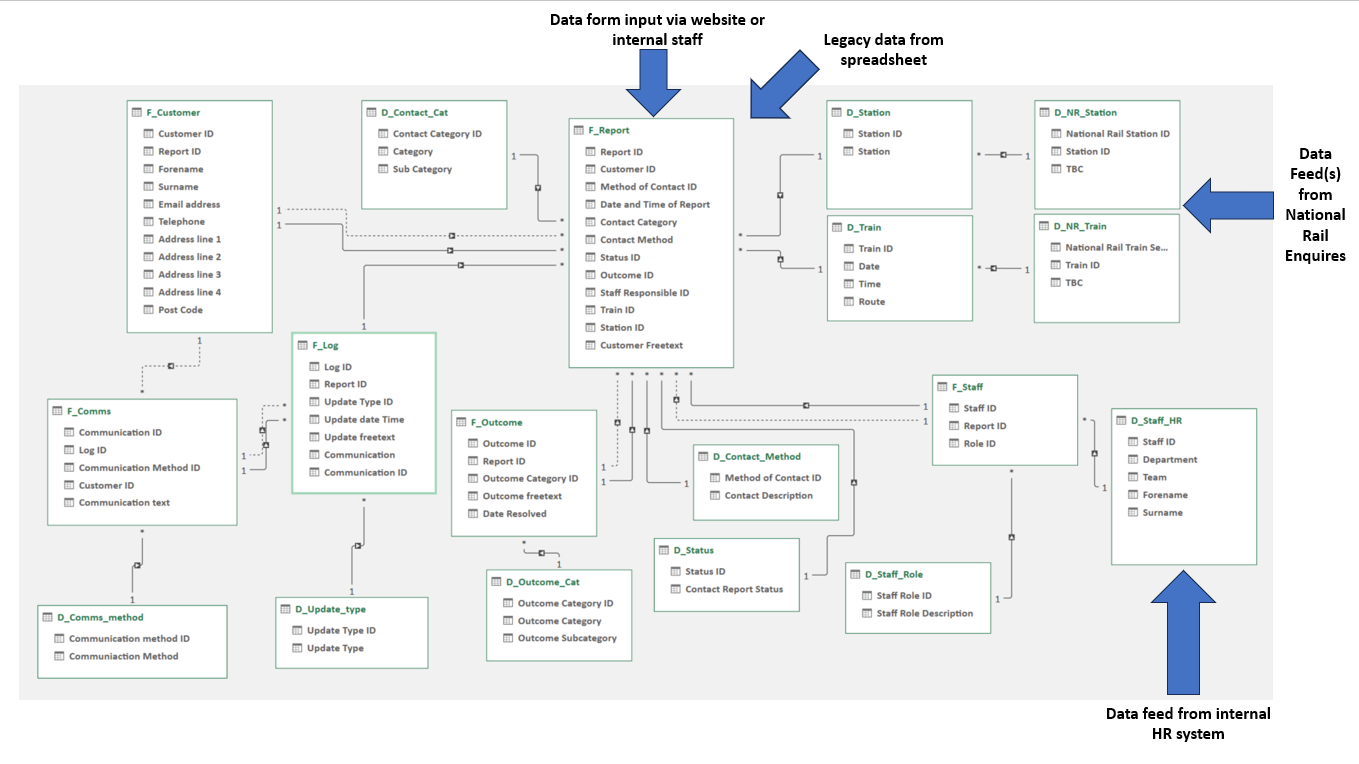


Figure 1 Database logical structure and data collection

The system will hold various data types within each table. IDs will be unique integer numbers. Varchar fields can hold strings of characters of various lengths, this will allow direct text entries as well as descriptive strings relating to data values to make them human-readable. Date Time fields record when key events occurred. (Harrington, 2009).



Figure 2 Data Types

**References**

Brookshear, J. G. & Brylow, D. (2020) Computer science : an overview. Thirteenth, global edition. New York, New York: Pearson.

EMCS. (2015) Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley Professional Development. Available via VitalSource Bookshelf [Accessed 17 September 2023]

Harrington, J. L. (2009) Relational database design and implementation clearly explained. 3rd ed. Amsterdam : Morgan Kaufmann/Elsevier.

Kazil, J. & Jarmul, K. (2016) Data Wrangling with Python. 1st ed. Sebastopol, CA: O’Reilly.

National Rail (2023) National Rail data feeds. Available at: https://www.nationalrail.co.uk/developers/ [Accessed 17 September 2023]

Ohlhorst, F.J. (2012) Big Data Analytics: Turning Big Data into Big Money. Wiley Professional Development.

Wiki.openraildata.com. (2023) KnowledgeBase. Available from: <https://wiki.openraildata.com/KnowledgeBase> [Accessed 16 September 2023)

Pandey, P (January 19, 2021) Extracting information from XML files into a Pandas dataframe, *Towards Data Science.* Available from: <https://medium.com/p/11f32883ce45> [Accessed 16 September 2023).

PyNative.com (2022), Python ISO 8601 Datetime, Available from: <https://pynative.com/python-iso-8601-datetime/> [Accessed 17 September 2023)

Connolly, Thomas, et al. (2014) *Database Systems: a Practical Approach to Design, Implementation, and Management, Global Edition*, Pearson Education, Limited, 2014. *ProQuest Ebook Central*, [Accessed 17 September 2023]