Transport for the North

D-LOG & I-LOG DATA COLLECTION SYSTEM

Discovery Report

Transport for the North

D-LOG & I-LOG DATA COLLECTION SYSTEM

Discovery Report

TYPE OF DOCUMENT (VERSION) PUBLIC

PROJECT NO. 70071930

DATE: JANUARY 2021

WSP

8 First Street
Manchester
M15 4RP

Phone: +44 161 200 5000

WSP.com

QUALITY CONTROL

Issue/revision	First issue	Revision 1	Revision 2	Revision 3
Remarks	Draft_Final			
Date	12th January 2021	1st April 2021	7th April 2021	
Prepared by	C.Steer/A.Jones	TAME	TAME	
Signature				
Checked by	A.Porter	M.Smith	W.Cui	
Signature				
Authorised by	A.Porter			
Signature				
Project number				
Report number				
File reference				

CONTENTS

Acronym	Description
D-Log	TfN's Development Log. The nature of proposed Developments (D) could be residential and employment land uses that are planned for a local area, and would comprise location, development type, development scale and timescale for development/build out.
EFS	TfN's External Forecast System, part of NorMITs within TAME Analytical Framework – future year travel model based on changes exogenous to the transport system using trip end modelling techniques.
GBFM	MDS Transmodal's Great Britain's Freight Model, which forms the freight component of the UK Department for Transport's National Transport Model.
I-Log	TfN's Intervention Log. The nature of proposed transport Interventions, could be highway or rail schemes that are planned for an area and would comprise location, nature, scale and planned year of opening.
LPA	Local Planning Authority
NELUM	TfN's Northern Economy and Land Use Model, part of TAME Analytical Framework
NoHAM	TfN's Northern Highway Assignment Model, part of TAME Analytical Framework
NorMS	TfN's Northern Rail Modelling System, part of TAME Analytical Framework
NTEM	UK Department for Transport's National Trip End Model – model which forecasts the growth in trip origin-destinations (or productions – attractions) up to 2051 for use in transport modelling.
TAME	Technical Assurance, Modelling and Economics Team at Transport for the North
TfN	Transport for the North

1 INTRODUCTION

GENERAL

WSP were commissioned by Transport for the North (TfN) in October 2020 to undertake the development of a prototype web-based data collection system of the D-Log | I-Log data. The project comprises three distinct phases as outlined below:

- **Discovery** – understand end user needs to help define the prototype or Minimum Viable Product (MVP) for the data collection system. This work has focused on both the data contributors, TfN staff and it's Local Authority partners, but also the data end users, these currently being TfN teams and their supply chain, Highways England (HE), Department for Transport (DfT) & Ministry for Housing, Communities & Local Government (MHCLG).
- **Alpha Development** – development of a prototype/MVP web-based data collection system based on the prioritised needs of the end users to test that the system would be an improvement on the existing system.
- **User Research** – Undertake pilot testing of the Alpha product with a number of end users to determine whether the Alpha is developed further and that it meets the needs identified.

This report encapsulates our Discovery work which was undertake in accordance with Government Digital Service (GDS) Discovery guidelines. The work was commenced in early November 2020 and finalised in early January 2021, following a two week break over the 2020 Christmas period.

DISCOVERY PHASE

The Discovery Phase allowed WSP to recognise the valuable feedback that the existing prototype and way of working has generated, and it is critical that this is fully understood. This has included not only feedback from the contributing Local Planning Authorities but also within TfN, Technical Assurance, Modelling and Economics (TAME) and others, and the wider end-users of the data, such as the modelling community to ensure it meets their needs.

The Discovery Phase allows WSP to collate together the feedback from all of the stakeholders and from this consolidate a list of user and technical requirements needed to formulate the solution to be built in the Alpha Stage of this project.

It should be noted that the end user data requirements were yet to be fully evolved and defined and it is expected that development of the Alpha system and if successful in iteration through a Beta and live service these data needs will become better defined allowing the system/s for both D-Log & I-Log to become more valuable over time.

METHODOLOGY

The results in the section below were gathered using a variety of methods to reflect the various levels of detail required from them and the access to the stakeholder. Primarily the following methods were used:

1. **In-depth discussions and technical deep-dives** were used with TfN's D-Log and I-Log technical team to help understand how the current system operates and the challenges and opportunities these bring.
2. **Questionnaires** were used with D-Log data contributors as there was a large number of these stakeholders so group discussions would have been logistically too difficult. This allowed us to find the specifics of the stakeholders and the range of different challenges/needs of the stakeholders. A full list of the organisations involved is included within the Appendix.
3. **In-depth discussions** with external end-user stakeholders such as Highways England. As we were unsure of how they currently use the data, this more open setting allowed us to deep-dive into the particulars of their current and future requirements for the data products.

EXISTING D-LOG|I-LOG DATA SYSTEM

An existing system and set of processes have been developed by TfN to collect the required D-Log & I-Log data and has been used since 2019. Data has been collected annually since this time, which has highlighted the following challenges:

- Local Authorities have shown their interests and are keen to be engaged but most of them have limited resources to undertake the data updates.
- TfN spent a significant amount of time trying to find the appropriate contacts to liaise with within the Local Authorities and once established on-going comms.
- Most Local Authorities submitted the data using their own thresholds and classifications. TfN has reserved internal resources to re-format the data and log the data using TfN data schema.
- There are numerous gaps in the data, for example, of all the submitted data, only 60% of the sites have filled in with coordinates.
- TfN has realised the variation of thresholds specified for the reporting of residential and non-residential sites.
- Some Local Authorities have struggled with SharePoint as not every Local Authority's system allows them to log in.
- A range of technical software used across Local Authority Partners, especially GIS and Excel, which added to the amount of post-processing required for TfN to provide..
- While some Local Authorities have relatively detailed residential data, many do not maintain the non-residential data at a similar standard.

As can be seen from the above feedback our focus for the Discovery work is to:

- validate and understand the data contributor's context (technology, available data sources and work constraints)
- understand the end user's data needs, i.e. what data is needed to help them make their decisions
- simplify the collection and management of data from contributors to data end users

For clarity, we also provide a description of what D-Log & I-Log data constitutes.

D-Log : The nature of proposed Developments (D) could be residential and employment land uses that are planned for a local area, and would comprise location, development type, development scale and timescale for development/build out.

I-Log : The nature of proposed transport Interventions (I), could be highway or rail schemes that are planned for an area and would comprise location, nature, scale and planned year of opening.

PLANNING DATA MANAGEMENT LANDSCAPE

Currently the data that underpins the planning process, such as planning applications and the plan making process and its decisions are typically bound within proprietary planning application management systems, such as Idox and Northgate, and Local Authority tools, either specialist GIS products and /or productivity tools. The planning data, generally a public asset, is therefore not often able to be readily extracted, shared and understood.

Having access to the planning data would assist not only in making this information more accessible to the public but also provide decision makers at national and sub-regional level of the changing demands for services. Hence, a number of activities have been initiated by MHCLG, Connected Places Catapult and private software businesses as well as TfN to extract the planning data. We therefore outline some of the recent activities that have been undertaken in this space and explore the potential for them to be utilised to develop the Alpha stage.

Connected Places Catapult – Developed a prototype web service to visualise the status of Local Plans and housing requirements based on MHCLG available data. The tool doesn't identify the location of the proposed developments but visualises at Local Authority level.
https://sites.google.com/view/localplans/home

MHCLG – Local Plans Data – Discovery & Alpha - Commission by MHCLG with Discovery undertaken in 2019 with Alpha stage delivered in May 2020, undertaken by DXW. They worked with a number of TfN Partners; Liverpool City Region (CA) and Gateshead Metropolitan Borough Council. The aim of the project was to extract data from annual Strategic Housing Land Availability Assessment & Reporting. It focused on digitising the Local Plan reporting process and standardising data in accordance with MHCLG Digital Land data principles and developed initial prototypes. It recognised that the data collection process was a burden to teams, but identified value of the data. The next stage was to focus on understanding value for end users.

MHCLG – Central Register of Planning Permissions – Discovery & Alpha – SocietyWorks were engaged in 2019 to develop a prototype how to make planning applications available as a collection of data in order to show how a beta system with more complete data could work. Feedback gained during discussion with MHCLG (15/01/21) indicated that this Discovery assisted in gaining a deeper understanding of the challenges associated with the process, specifically, the lack of unique references used within the current planning system.

MHCLG Digital Land Team – Discussions were held with members of the Digital Land Team during w/c 11/01/21 on activity and progress within their team and wider digital transformation of the planning system. It was concluded that, as yet, no other party had focused on collecting future development data. However it was on the roadmap for the Digital Land team to explore later in 2021 and the team would be keen to understand our work, Discovery & Alpha feedback. Additionally, valuable insight could be gained from an agreed/tested data structure during this work that ideally would align with MHCLG Data Principles.

Private Organisations – PlanIt, Planning Parrot, Land Enhance & Planning Finder - Essentially web scrapers for Local Authority websites planning data built upon open source technology with the code available on GitHub. Data coverage and quality is limited and unlikely to be sufficiently granular to support the commission needs.

Local Authorities – LB of Southwark, Hampshire CC, GMCA, Leeds ODI, GLA Planning Data Hub – Various local authorities and public bodies have previously and are in the process of extracting planning data for various end uses, but generally to allow the data to be re-used, made available and transform processes. Of particular interest is the newly released GLA Planning Data Hub which is requesting that all London Boroughs provide planning data in a consistent and open format.

2 D-LOG FINDINGS

2.1 DATA CONTRIBUTORS

This section looks at the results of the feedback about the existing process from the analysis of the Correspondence Log with Local Authorities, the pre-survey questionnaire and the detailed questionnaire.

D-LOG20 DATA COLLECTION REVIEW

This section reviews the data collection process undertaken by TfN for the D-Log20 dataset, prior to engagement with WSP, to identify the lessons learnt from the experience, the outstanding challenges and the participation rates.

CORRESPONDENCE LOG

The Correspondence Log was put together by members of TfN's TAME team to manage the submission of D-Log20 data. The results below illustrate the completion rates of the dataset, the average number of submissions and the challenges faced in the collation of the data.

Table 1 Completion of Results D-Log20

Completion Results	Number of respondents	Percentage of respondents
Did not engage with comms	4	5.1%
Engaged with comms but did not submit data	3	3.8%
Submitted Data	72	91.1%
Total	**79**	**100.0%**

Table 2 Number of Developments (residential and non-residential) Submitted per Local Authority

Minimum number of D-Log submissions	2
Maximum number of D-Log submissions	1011
Average number of D-Log submissions (removing extreme outliers):	53

TIMING

- September does not seem a good deadline day due to lots of holidays taken in August/Early September which causes delays and potentially a deadline in mid-September from other projects.

DIFFICULTY PROVIDING CORRECT DATA

- Seems to be a disjoint between where GIS data and planning data are stored - needs better coordination.
- GIS areas to centroid eastings/northings seems to be a pain point - often needs to be processed manually by TfN.
- Residential are a lot easier to access than employment sites.
- Start/End dates and coordinates seem most likely to be missing.
- National Parks seem to have little resources/data available.
- It is quite a difficult undertaking for several Local Authorities as did not have technical abilities to collate and combine datasets quickly.
- Local Authorities are trying to provide partial data whilst finding missing data – no validation from the application, all human validation.
- It was anticipated that some Local Authorities may provide data in PDFs due to time constraints/lack of technical skills but the results show that PDFs weren't a very common format to send.
- Users are overwriting cells in provided Excel template with prescribed entries with their own categories.
- Quite a few Local Authorities had insufficient time so reverted to D-Log19 data.

COMMUNICATION ISSUES

- Initially in the D-Log collection process there was a big task to identify the correct correspondents within each Local Planning Authority. For those who still have not provided the correct data, it might reflect that they still do not have the correct contacts identified.
- Internal comms and people leaving/on holiday causes significant delays.
- Quite a few planners seemed unclear as to what is required from them (e.g. exact formats, etc).
- Help documentation within the spreadsheet was not clear enough to describe what to do with ambiguous data when the LA's schema did not match the TfN schema - improving this would help reduce errors/time taken/emailed questions.

OTHER COMMENTS

- One representative for some of the combined authorities in the North.
- Not too many technical issues with the portal itself – online portals are accessible.

PRE-SURVEY QUESTIONNAIRE

A pre-survey questionnaire was issued in November 2020 to get a grounding of the technical set-up of the Local Authorities that submit the D-Log data. A total of 25 Local Authorities responded to the questionnaire.

The majority of Local Authorities use an Idox based planning platform to store planning application data. The remaining Local Authorities use other large platforms or bespoke applications.

Table 3 Response Summary to Location of Planning Data

Planning Data Storage	Number of respondents
Idox's Acolaid	8
Idox Uniform	6
Northgate's Assure	2
Internally developed	2
Idox CAPS	1
Database; Excel Spreadsheet; GIS software	1
Fastplanning	1
Arcus	1
Information@Work	1
Salesforce	1
Tascomi	1

The majority of Local Planning Authorities surveyed hold their Local Plan data in GIS software however there is not one single software used, and includes Esri products, MapInfo, QGIS and bespoke software. A number of other Local Authorities store this data in planning software, but the majority seem to distribute this data across planning software, GIS software, databases and spreadsheets.

All respondents have access to a range of modern browsers and only one had access to only Internet Explorer. Therefore, we will progress to design the application for only modern browsers.

The average time taken to complete the data collation and submission process by respondents was reported at 6.5 hours, with a minimum of 1 hour and a maximum of 30 hours which reflects the large variation in how long it has taken Local Authorities to submit this data with some finding it an onerous task. This is a key requirement the future solution must address to reduce the time taken to complete the reporting and update it in future years. It might act as a resource for Local Authorities to be a single-source-of-truth for their own planning data.

DETAILED QUESTIONNAIRE

Following on from the pre-survey, a detailed questionnaire was sent to all of the existing D-Log contributors in December 2020. The aim of this questionnaire was to ascertain the specific data management practices that Local Authorities use in-house to define and store their planning data, both Local Plan and planning permissions, and to understand in greater depth the challenges they

faced with contributing D-Log data in the past to help refine the process. A total of 16 Local Authorities/National Parks responded to the questionnaire and their results are given below.

User Personas

The Local authorities were asked a set of questions to understand the characteristics of the people who undertake this data collation and submission so that we can cater the future application to their needs.

The results showed that the majority of the respondents were either a planner or a policy officer and only one was a GIS or IT administrator/specialist which suggests those partaking in this data collection have subject matter knowledge of the data.

Figure 1 Summary of Role of Respondent
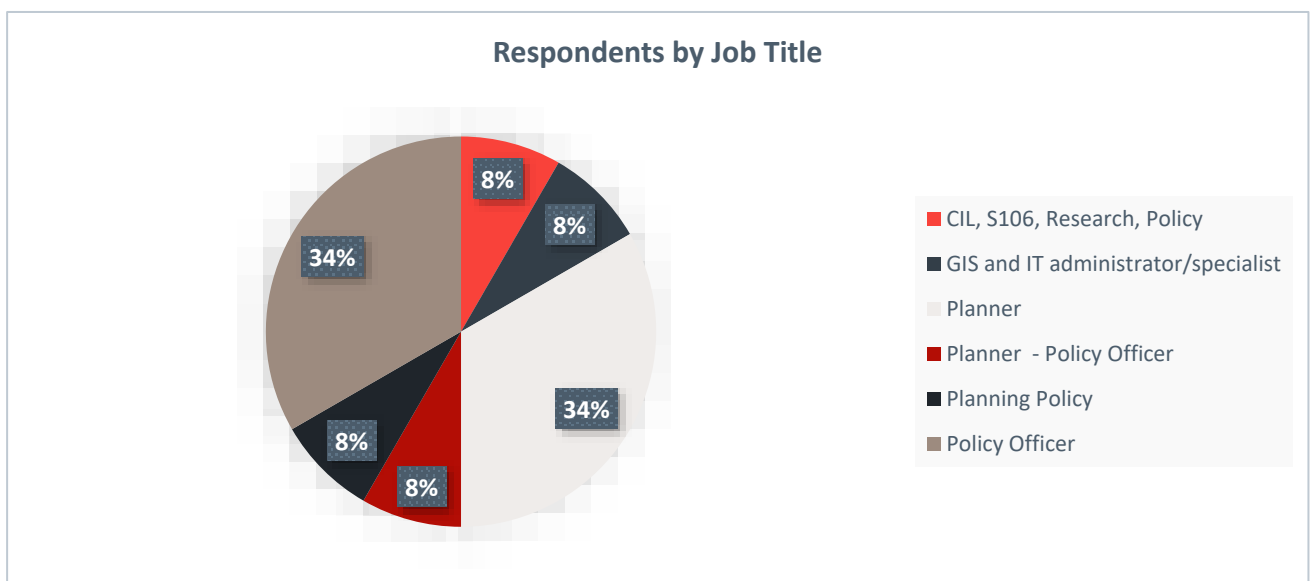
In order to understand the technical abilities of users we asked their comfort levels of using new technology, scored between 5 being very comfortable and 1 being very uncomfortable. Half of the respondents scored a 3 with the remainder scoring 4 or 5. This suggests that we do need to place an emphasis on making the application as intuitive to use in order to maximise uptake and reduce training pressure.

The D-Log data schema asks the user to provide some GIS data. In order to grasp the users' familiarity with such technology we asked them their skill levels with GIS. The results are shown below:

Table 4 Summary of Self-Reported GIS Skill Level by Respondents

Skill Level	Number of Respondents
Basic user of GIS desktop software	8
Comfortable user of GIS desktop software	2
I know how to find eastings/northings of a place using online/internal tools	2

This indicates that the majority of respondents had some level of knowledge of GIS and would understand basic concepts such as what a shapefile is, coordinates and plotting areas.

Data Access and Ease of Submission

In order to understand where data is stored, the questionnaire asked where the Local Authority stored their Local Plan data. The results are given below. This indicated that for many Local Authorities, the data is stored in multiple locations with many storing it in a GIS database and a different database to their planning permissions. This indicates that Local Plan data is treated differently to typical planning data which is discussed further on in this section.

Table 5 Summary of Location of Local Plan Data

Storage	Frequency
Different database/system as planning permissions	6
GIS database	9
PDF on our website	4
Same database/system as planning permissions	3
Spreadsheets	1

One aspect we were most interested in was the ease of collating and submitting this data as that would be one of the largest hinderances in the quality and coverage of the D-Log database. The self-reported answers on how difficult they found to collate the data are as follows, with no respondents choosing 'very easy' or 'very difficult'.

Figure 2 Reported Difficulty in Collating Data
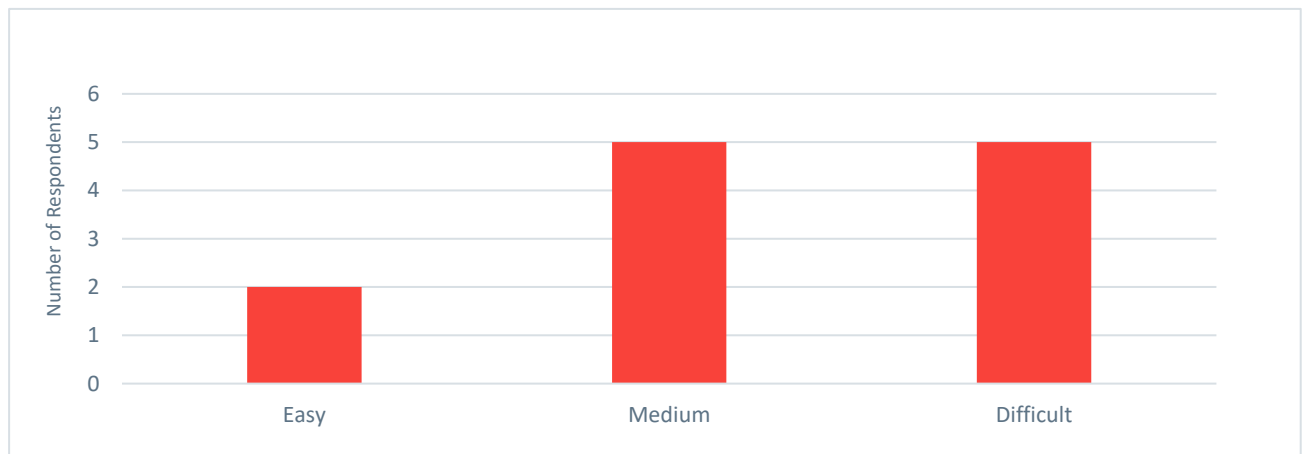
Figure 2 indicates that the process needs to be improved as currently users are finding it too difficult and this may result in them not completing the data or keeping it up-to-date. The following question gives some insight into the reasons why users found it challenging.

Table 6 Key Difficulties Identified by Respondents in Relation to Providing Data

Reason	Responses
I had to access multiple datasets and combine them together	10
I needed to extract data from another system to get the data	5
Needed to contact IT/other team to get access to data	2
Someone was on holiday/not available and I needed their help	1
I had issue accessing the data due to office closures/remote access issues	1
GIS officer left the Council	1
Extracting the data needed to fully complete the request was too time consuming and I was unable to do it	1

The most frequent reason given for difficulties with providing the data was the need to combine datasets together suggesting that the data required is stored in disparate locations within a Local Authority.

Another potential difficulty identified for providing the correct data or having to spend time adjusting it was about how similar the data schema held by Local Authorities are to those requested by TfN. The answers below show that the majority of respondents had some similarities between their data and D-Log however there was some level of adjustments required.

Figure 3 Reported Differences Between LPA and TfN Data Schema
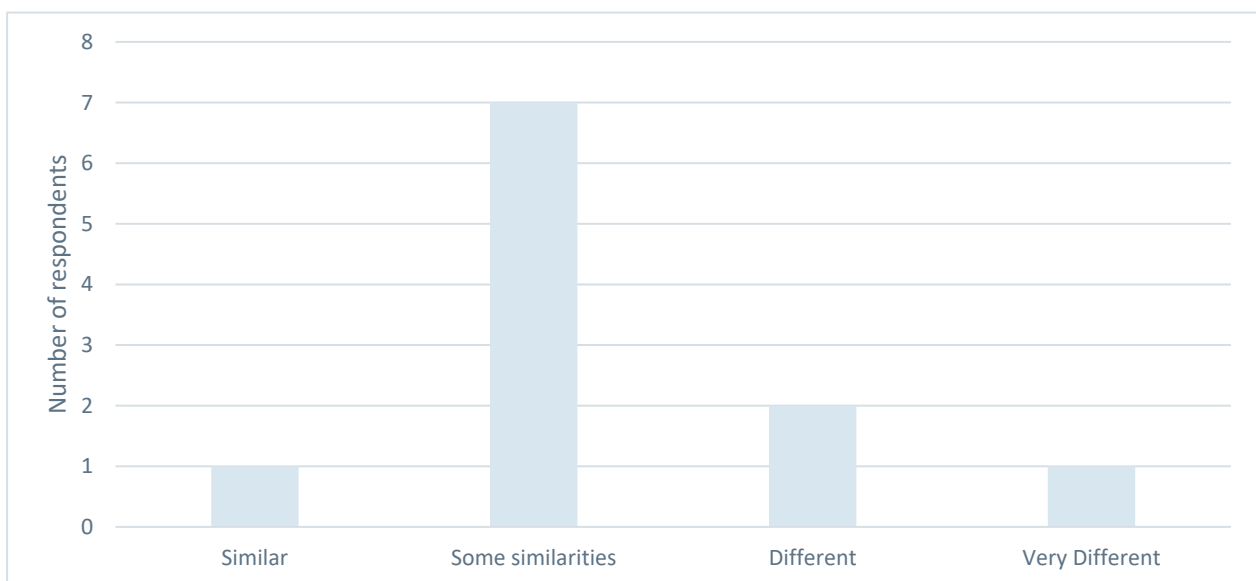
The largest differences reported by the respondents were as follows:

- Use classes – Local Authorities do not keep as detailed use classes as per [Planning Portal](#) required but instead use higher-level descriptions such as employment, retail, housing and waste/energy.
- Local Authorities may not store detailed data such as the number of bedrooms per dwelling or housing mix.
- One Local Authority mentioned that TfN asked for number of completions per calendar year where their recording is all done on the financial year.
- The planning permissions and Local Plan data were stored in different systems so needed to be extracted and combined together.

The questionnaire also aimed to identify which data categories were found the most difficult to complete. This raises questions over the balance over the level of detail requested and what can be provided. The answers to this are given below:

- Coordinates not stored by Local Authority – too time consuming to complete.
- Prompts for data categories are hard to remember and slows down filling in data.
- Sector types not aligning or being too early in the planning process to know.
- Completions dates are difficult to provide.
- Bedroom data is not available.

Following on from the feedback about some of the Local Authorities, a specific question about how they store their coordinates for the developments. The answers below show that there is a variety of methods used with some authorities using multiple. This indicates that the alpha solution will need to accommodate this range of location formats.

Table 7 Summary of Where Coordinate Data Stored

Option	Frequency
Don't store them - only have addresses/postcodes	2
Eastings/Northings stored within planning data	6
Had to find centre point of development boundaries using GIS software	5
Have GIS points already stored as a GIS layer	2
Sent over plans/shapefiles to TfN for them to process	2

In relation to how closely the data asked for by TfN matched the data held by Local Authorities, another question was asked to understand what the minimum threshold was they used to record the residential and non-residential proposals. The results are given below that show the majority record all residential and non-residential proposals and the remaining have relatively low minimum values.

Table 8 Summary of Respondents' Minimum Threshold for Residential Data Collection

Specified Quantum	Residential
No minimum	8
5 units	1
10 units	1
Unknown	2

Table 9 Summary of Respondents' Minimum Threshold for Non-Residential Data Collection

Specified Quantum	Non-residential
No minimum	7
50sqm	1
500sqm	1
1000sqm	1
Unknown	2

INCENTIVES

It has been recognised that it can be a significant task for Local Authorities to provide the D-Log data. Therefore, we were interested in their incentives for undertaking it as they were not legally obliged to provide it. Only 27% of respondents said that the data was useful for other purposes for their organisation however a majority of respondents said that they agreed or strongly agreed with seeing a value in providing this data to TfN.

Figure 4 Respondents Response to Whether D-Log Data was Useful for Other Purposes in their LPA

Legend: Disagree, Neither agree/disagree, Agree, Strongly Agree
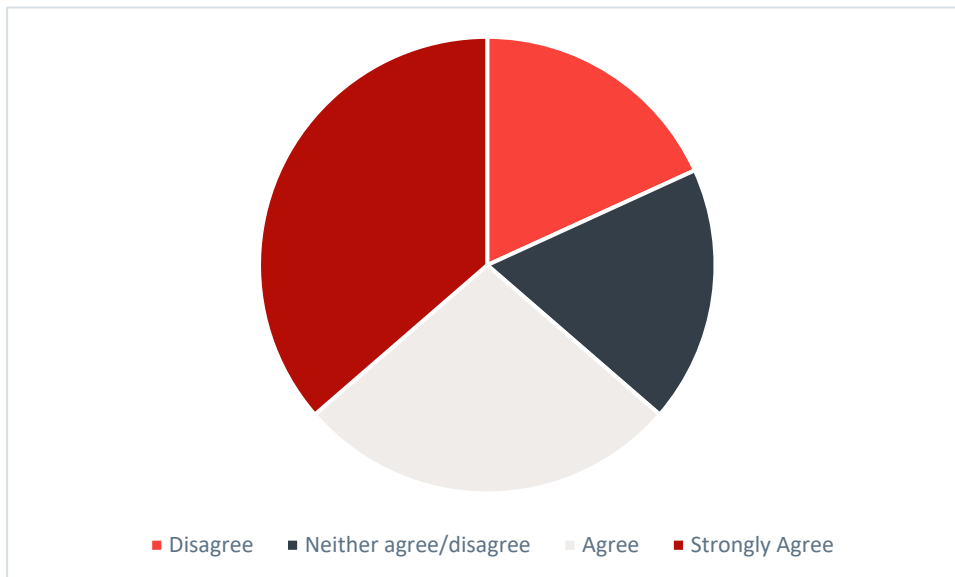
LIMITATIONS

The response rate for the pre-survey questionnaire and the detailed questionnaire was 32.9% and 14.4% respectively. As both questionnaires were undertaken via an online survey tool it is possible that this may have acted as a barrier for completion for some less technical users. As a result of this and the proportion of total users who answered that these results may not be representative of all users. Furthermore, the survey was undertaken in the lead up to Christmas 2020 and alongside other surveys requested by TfN, the respondees may not have had the capacity to respond.

2.2 EXISTING TECHNICAL ARCHITECTURE

POSTGRESQL

D-Log data is currently stored within a PostgreSQL instance hosted on Microsoft Azure. Three tables contain development site information, and there are additional tables for relating Local Authority information and use classes to development sites. Within the database, domains exist for some fields so that only allowed values can be entered, thus improving data consistency.

Each development site is recorded in the development_site table, with additional data stored in the residential_development_site or non_residential_development_site table depending on whether the development site is residential or non-residential. Each development site has a local planning authority assigned (from the local_planning_authority table), and additional local authorities can be flagged as being associated with the development (associated_authority table).

The UML diagram below illustrates the relationship between the existing D-Log tables and domains (green).

Figure 5 Existing D-Log Database Structure

EXISTING D-LOG DATA PROCESS

The current process for populating the D-Log database with D-Log data is as follows:

1. Local Authorities enter data into their D-Log Excel file template.

2. Local Authorities upload the Excel files to the TfN Sharepoint.

3. The Excel files are downloaded by TfN and are manually checked to ensure they meet column and sheet requirements for further processing. Some manual tidying up of the file may be required at this stage.

4. A python script is run to tidy up some names and formatting in the Housing and Employment tabs, and remove hidden sheets.

5. Once all Excel files have been received from local authorities, a FME workspace is run which will take the data from each of the local authorities and merge it into a new Excel file which has all the correct columns for residential and non-residential D-Log data.

6. Another FME workspace (see screenshot below) is then run to load the data from the merged Excel file into the database.

Figure 6 Existing D-Log FME Process
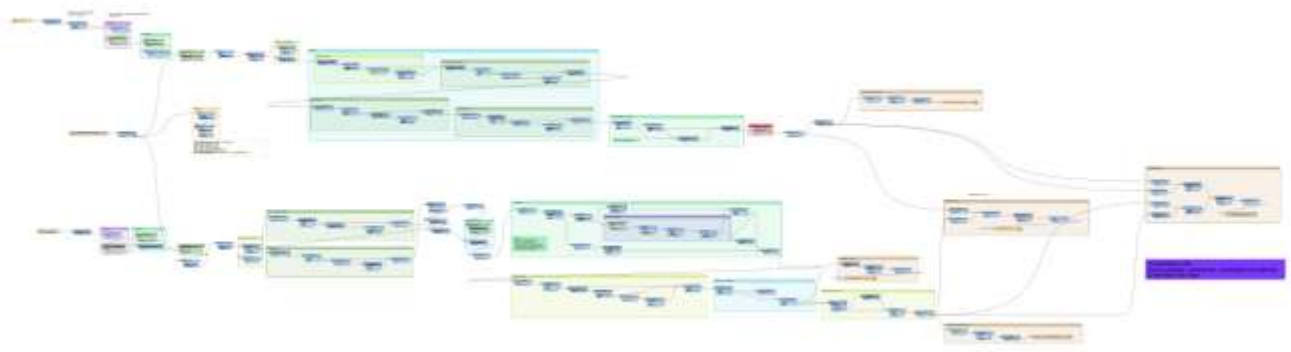
Key Issues

The following issues with submitted Excel files have been identified. The FME workspaces are necessarily complex to deal with the following:

1. Some 2020 data has been submitted using the 2019 schema. To handle this, further functionality needed to be added to the FME D-Log Processing.fmw workspace.

2. A high degree of manual validation and checking is still required on submitted Excel files e.g.

   - Are the required sheets present and named correctly?

   - Are the column names as required, or is it easy to map/copy over the correct column names?

3. Some combined authorities have separate schemas to other Local Authorities. Workarounds are employed in the FME workspaces to account for this.

4. The Local Authority name entered in the Excel file often doesn't match up with name in the local_planning_authority_2019 look up table. Fuzzy matching is employed with the FME workspace to mitigate this.

5. D-Log Processing.fmw contains lots of workarounds (validation and converting) to account for different text formatting/content that is placed in each sheet cell by different local authorities. For example:

   1. Site area is requested in hectares, but may be filled in any number of ways such as "<number>" (correct!), "<number> ha", "<number> m^2", "<number> m2"….

   2. Use classes: should ideally be a comma separated list of classes such as "B1,B2,B8", but the following have been submitted:

- "B1b,B1c,B2 and B8" - this example would need to strip the B1 "b" and "c" suffixes, then remove the extra B1 entry, and also remove/ignore the "and".

- "A mix" which is not "a mixture of classes" but "a mixture of A classes"

- "A4 2000sqm, A1-5" which has an area value for the A4 use class, and a range which equates to "A1,A2,A3,A4,A5"

2.3   INTERNAL DATA USERS

WSP set up a meeting with members of TAME team to discuss the previous D-Log data schema design, the collection process and the feedback and challenges from both the end-users' and TfN's perspective. This included sharing of the Correspondence Log of all the interactions with D-Log data contributors, their personal experience of the data collection process and key internal contacts. Whilst the results of the Correspondence Log largely feature in the appropriate section, a summary of their experiences collecting the data are given below:

- A number of data contributors just did not engage with the process at all and would not respond to any emails, which is likely to be related to workload or staffing issues in Local Authorities.

- It took quite a lot of time to establish the correct contacts within each organisation however data collection became easier once they were established.

- There were quite a lot of contributors who sent multiple messages back and forth for both clarifications or because they were only able to provide data a bit at a time. This was time consuming and quite difficult to monitor, hence the creation of the Correspondence Log.

- There have been communication emails issued to explain the project before to data contributors.

Question posed - What is the data used for by yourselves and your partners? – (we would like to understand this so that we request the correct quantum and quality of data to reduce burden on the LPA's and data collection activities and also provide value to your stakeholders in accurate and consistent data. We undertook a brief review of the data collected there are some inconsistencies which could generate over reporting, under reporting and duplication.)

- We established the data will feed into EFS.
- We established that a higher threshold of major developments would be acceptable but recognise that multiple planning consents for sub major development within a LA may affect transport needs, for instance densification of suburban/peri-urban locations.
- We established that polygon extent is not critical but area is useful to understand density.
- We established that non-residential mixed use developments are critical to understand all components and classes and Gross Floor Area (GFA) to determine likely trips.
- We established that start and end year are important but linear build out rates, especially for residential, are usually undertaken rather than address specific time periods.
- We established that actual delivered and planned developments from a 2018 baseline need to be accurate.
- We established that the D-Log data is used in conjunction with NTEM data or other TfN/DfT scenarios for population and employment growth, with the latter to spatially distribute the

development, but capped at the relevant growth predictions, which may or may not predict likely delivery.

■ The adjusted population and employment projections from the EFS that use the D-Log as an input will be used within other TfN models, such as the Northern Economy and Land-Use Model (NELUM) and the Great Britain Freight Model (GBFM).

■ We established that EFS does not take into account removal of trip generators or origins through demolition activities. This may be captured through the planning process.

■ The raw D-Log data is also likely to be used by TfN in detailed business case work, such as Northern Powerhouse Rail (NPR), where some of the economic benefits could be related to further development, over and above that specified in the D-Log. For example, development in the vicinity of new rail stations, not currently accounted for in spatial plans.

2.4 EXTERNAL DATA USERS

Highways England

WSP arranged a meeting with Highways England (HE) to discuss about how HE currently uses the D-Log data they request, if it could be improved and whether they had any potential use for the I-Log dataset. The summary of the discussion is given below:

■ Haven't defined granularity on what data they would require for D-Log however other STBs only send them developments of over 1000 residential units.

■ For their route strategies modelling they only use Local Plan data and not permitted planning applications to estimate increases/changes in traffic flows within their transport model.

■ The maximum year they model to is 2050.

■ In terms of the locations of developments, all developments should be included that are located on the Strategic Road Network (SRN), larger ones on the Major Road Network (MRN) (3 miles away from SRN excluded) and exclude any on the Local Road Network (LRN) with the exception it had a significant impact.

■ They did not have any awareness of the I-Log dataset. If they were to use it in future they would only be interested in large railway interventions of changes to rail stations.

■ HE collects lots of data about network performance and I-Log would be helpful to see if projects have already been put forward to address these issues.

Department for Transport (DfT)

WSP and TfN undertook a meeting with Modelling and Appraisal team from DfT to discuss how DfT use data very similar to the D-Log in their revisions of the National Trip End Model (NTEM) and how this data could potentially be improved to provide more value. The summary of the discussion is given below:

■ NTEM is a land-use travel demand transport model that largely uses housing predictions from Local Plans and job forecasts from the Department for Education's Working Futures publications. The 5-10 year forecasts are extrapolated to form long-term projections.

■ NTEM is revised roughly every 5 years, primarily influenced by the release of census and mid-census estimates for population (years ending in 1 and 6) with an offset for the time taken to release and process the census data. The maximum year that is modelled is 2051.

- The housing data required is the aggregated change in the number of residential units by Local Authority by 5-year period (see above). DfT are not interested in the number of developments that make up the total number of residential units however are only interested in those that are likely to go ahead and not highly speculative developments. The residential mix, size of house and location of residential development are not included in their analysis and so are not required.

- DfT do not analyse how accurate the Local Authorities have been with their delivery of their forecast housing developments. DfT undertook an internal study over 2 years ago on the Oxford-Cambridge Corridor to compare the forecast housing developments in the Local Authorities within that vicinity to the planning permissions that were permitted and built. This found that Local Authorities varied significantly with their ability to delivery housing developments, with many overestimating delivery and a small number underestimating. As a result, no single factor could be used to refine the forecasts.

- NTEM 6 and 7 experienced a number of small revisions as Local Authorities were not willing or able to engage initially with providing the housing development quantum and signed off numbers which when later incorporated into the NTEM lead to the Local Authorities questioning the outcomes of the model and requesting changes.

- At present the NTEM is formulated at Local Authority District geography for statistical significance however there is pressure to potentially recalculate this at the MSOA level in future.

- More on how Local Authority planning data is applied to NTEM can be read within the NTEM Discovery report: [Department for Transport NTEM discovery report (publishing.service.gov.uk)](publishing.service.gov.uk)

Ministry of Housing, Communities and Local Government (MHCLG)

TfN arranged a meeting with MHCLG Development Plans, Planning Directorate to have an initial discussion about the work each team has been working on, and whether there are overlaps in the work and efforts the teams have experienced. This initial meeting has highlighted the mutual interest in structuring and using the spatial planning data. It has been agreed that further discussions will be arranged.

Homes England

TfN joined a collaboration workshop that organised with Homes England Strategy and Markets team. The focus of this workshop was mainly on understanding what each team does and seeking any mutually beneficial ways of working in future. Given the busy agenda of this workshop, it was agreed that further workshops for D-Log and I-Log programmes will be arranged in the future to seek ways to collaborate.

3   I-LOG FINDINGS

3.1   DISCOVERY

The information presented has been derived from the following discussions:

■ Major Roads Team, TfN
■ NPR, TfN
■ Strategic Rail, TfN
■ TAME team, NoHAM transport modelling
■ TAME team, NoRMS rail modelling

3.2   I-LOG CONTENT

An intervention was defined as being any change or addition to the transport network and are generally physical alterations rather than policy changes. However, there could be non-physical changes such as increased train frequencies that would be incorporated into I-Log.

Given this definition there could potentially be an overwhelming number of intervention records. As such TfN staff are compiling a list that have been suggested through a number of studies or from partner organisations. There is currently no formal procedure for deciding what should and should not be included in I-Log. Although some interventions like those that have been through Investment Programme Benefit Analysis (IPBA) are clearly more likely to proceed and high priority ones that need to be captured.

However, it was also felt that allowing TfN staff to act as a 'filter' for all those interventions that are suggested is beneficial. If partners were to submit directly to I-Log it could produce a less targeted and useful list. As such a network of contacts to directly contribute towards I-Log has not been pursued.

3.3   DATA CONTRIBUTORS

As TfN staff are performing this 'filtering' role they are considered to be the data contributors for I-Log who work with those partners and ultimate delivery bodies. This includes staff in the Major Roads, NPR and Strategic Rail teams.

3.4   DATA USERS AND USES

Internal users of intervention data may be those same teams that contribute data. The latest information about an intervention could be shared with partners, be subject to a media query or feed into transport modelling. As such TfN would benefit from a single source of truth regarding interventions to access the latest information and service these needs.

Many of the users and use cases are expected to change and develop in the future as I-Log itself develops from its current form. For example, it was believed that Local Authorities would benefit from access to a regional view of interventions to explore influences around their borders.
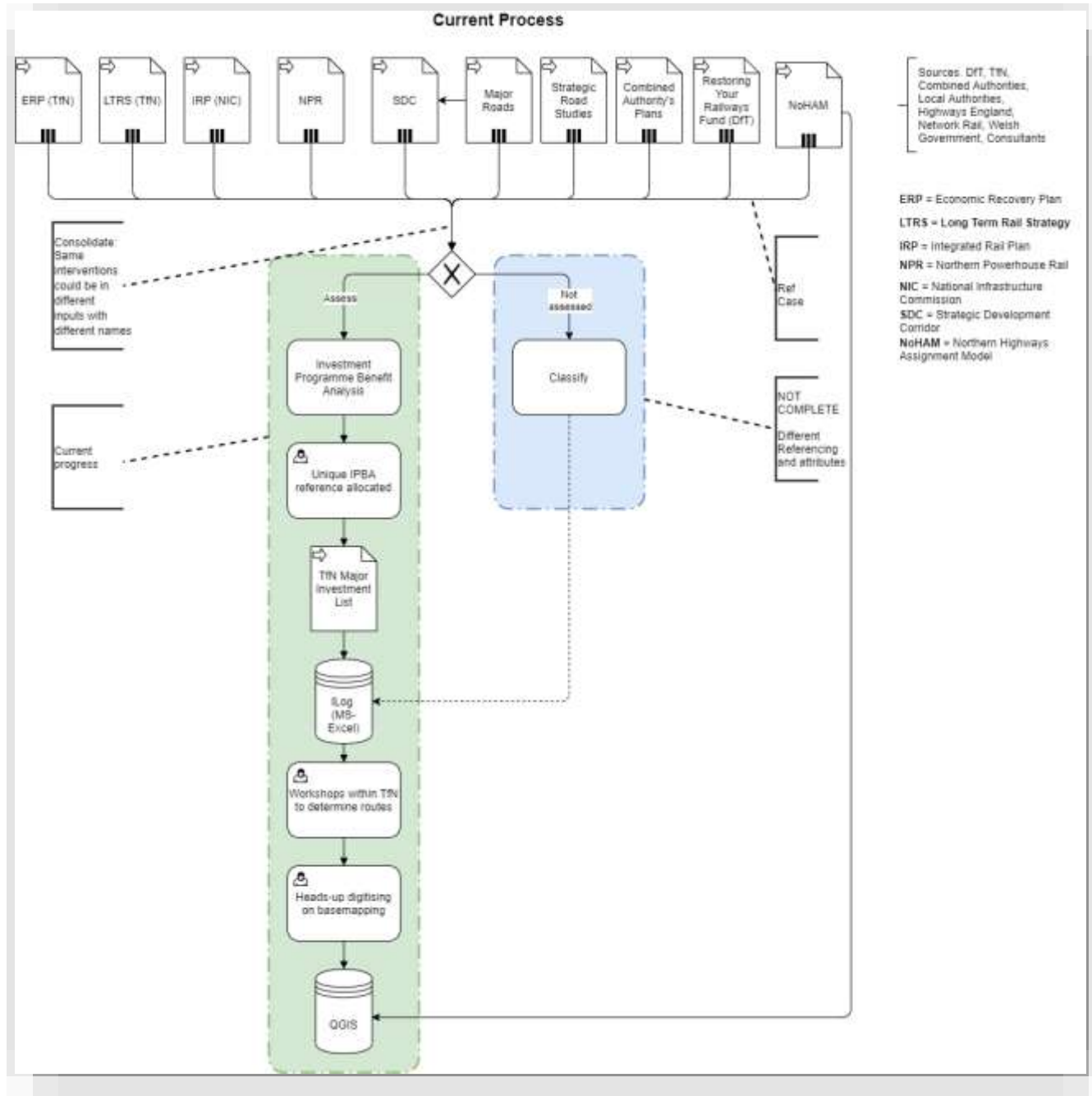
As the data is at an early stage of development the users and use cases are yet to be clearly defined.

3.5   I-LOG DATA

The understanding of the processes and progress of I-Log are illustrated in **Error! Reference s ource not found.**. During conversations with the Major Roads, NPR and Strategic Rail personnel the following points were determined:

- The Major Roads team are leading the I-Log development in liaison with TAME TfN colleagues, but this is likely to change in the future.
- At the time of writing, the I-Log spreadsheet contains the IPBA interventions (around 300) and non-IPBA rail schemes (approximately 300 also).
- An issue in the data collection process is that the same intervention is known by different names in different sources. TfN are 'merging' the interventions from these sources.
- Although the acquisition of data is a very manual process, once a definitive list is produced, there is no need to repeat such a scale of data collection.
- Instead, the attributes about existing interventions need to be managed (e.g. status) with modest numbers of new interventions needing to be added.
- The current attributes include:

  – Reference number
  – Scheme name
  – Scheme reference
  – Delivery body
  – Status
  – The sequencing information for business as usual, Strategy 1 and Strategy 2 (which is only relevant to the IPBA interventions).
  – Some indication of source of rail schemes through flags (added in v0.1)

- The I-Log should be the single source of truth for interventions across TfN. At the moment, interventions can be repeated in various lists.
- There needs to be an owner of I-Log to maintain the data.
- A record of modifications should be maintained to inform users of the record's status.
- Several conversations stated that it would be useful to readily access any associated documents for the intervention (e.g. pdf documents, CAD plans).
- There will be a need to protect the sensitivity of the data. Although the list of interventions is public domain information, certain details about an intervention could be sensitive (e.g. an exact route alignment)
- It would be helpful to record the source of the intervention (e.g. which study it was an outcome of).

D-LOG & I-LOG data collection system   PUBLIC | WSP
                                                                                        January 2021
Transport for the North   Page 20 of 42

Figure 7 Existing I-Log Data Process

Figure 7 Existing I-Log Data Process

3.6 TRANSPORT MODELLING

Following reviews on NoHAM and NoRMS about transport modelling the following points were raised:

- Road modelling would benefit from having any design assumptions documented in I-Log. These necessary attributes would need to be defined and could potentially require numerous attributes.
- It was recognised that this would make the data collection more onerous (although it would not apply to all interventions).
- Design assumptions may be different to model assumptions.
- Potentially any model assumptions could be fed back to I-Log
- Rail modelling would benefit from having details of the current and proposed stations, frequency and journey times.
- It would also help to identify where there had been a change to an intervention since it was first published.

3.7 I-LOG MAPPING

Following discussions with a member of Major Roads Team, TfN who is gathering the GIS representation of the interventions, we established that:

- The I-Log spreadsheet and GIS mapping are two separate systems.
- Alignments are digitised in consultation with staff familiar with the interventions onto a basemap.
- WSP suggested using road and rail vector data (such as the vector OS OpenMaps) to trace the alignments and provide faster more consistent digitising.
- Some interventions, like those on the road network, are represented by the start and end junctions. These two map features relate to just the one set of attributes.
- WSP believe that using desktop GIS remains the most appropriate place to perform such complex digitising.
- Some interventions may not yet have decided on the exact alignment and should be represented differently (e.g. different accuracy categorisation / dashed line).
- Some interventions do not have any mapping data associated with them. There is a plan to categorise interventions and only digitise those that fall into particular categories. These categories are not yet defined. It is important to recognise that only a subset of interventions will have accompanying mapping.
- Some interventions are comprised of many supporting schemes. For instance, NPR will occupy substantial corridors across the North, but there will be many smaller interventions that only exist to support this strategy.
- The mapping needs to consider the display of these overarching schemes to ensure they do not mask the smaller ones. However, the user needs to be aware that some smaller schemes are related to an overarching one.

3.8 WSP OBSERVATIONS:

- There is a limited description of the intervention in the current spreadsheet. This has been expanded for the rail interventions added in the 0.1 version of I-Log
- It does not include proposed start date and end date where interventions are in progress.
- As the current data only contains IPBA interventions, any other interventions will need a different referencing method (as the current ones are in the form IPBA_*nn*). The version provided does not include a unique ID for non-IPBA interventions.

- The 'Status' field, if it is populated, is really more indicative of the funding mechanism (e.g. RIS1, RIS2).
- As such a status that describes the certainty of progressing would be preferred (confirmed during discussions). A list of valid statuses would need to be defined.
- There is no data that would assist with road or rail modelling.
- Having a map of interventions during the time when new interventions are being considered and then reviewed before being added would greatly assist the consolidation activity.

4 SUMMARY

The Discovery Phase has allowed WSP to understand the current processes undertaken to collate, process, update and use the D-Log and I-Log datasets. It has shown that the previous versions of the datasets have provided a solid foundation for future iterations and have identified key challenges faced by the different stakeholders associated with the datasets. This chapter identifies these key challenges which will be addressed in the following requirements and proposed Alpha Phase chapters.

4.1 D-LOG CHALLENGES

The Discovery Phase has identified some key challenges for the following three groups: the data contributors, typically Local Authorities, the data processors which are members of TfN and for the end-users of D-Log which include both TfN and external stakeholders. These challenges are summarised by group below.

DATA CONTRIBUTORS

Table 10 Details of Challenges Faced by D-Log Data Contributors

Challenge	Details
Time taken	The process of collating together the correct data, entering it into the spreadsheet and submitting it to TfN was relatively time onerous for most Local Authorities. This may have explained those organisations that did not want to engage in the project and those that provided incomplete data
Providing eastings/northings	Local Authorities store their location information in a number of different locations and formats. A number of them do not store eastings and northings with their planning data and this either needs to be gathered from other sources or derived from the data itself. The solution would need to be able to accommodate these different formats.
Data schema differences	Local Authorities reported that whilst their data schema for planning data was similar to what TfN were requesting, there were differences and this was particularly evident for residential data in the level of detail included for development build-outs and number of bedrooms/housing sizes and detailed land use types for non-residential data.
Combining datasets	The previous iterations of D-Log have requested both Local Plan data and planning permissions which for many Local Authorities are stored in different locations and therefore they have had to spend time combining these datasets which often have different schema and this may also involve contacting other colleagues to gain access to these systems.
	In addition, planning permissions were more likely to change on a frequent basis which requires data contributors to ask for data to be exported out of planning systems which may require additional assistance compared to Local Plan data which could often be completed by publicly available documents that updated infrequently.
Lack of guidance	The number of contributors attempting to overwrite drop-downs, missing out data or asking questions to TfN indicate that there was a lack of guidance

DATA PROCESSORS

Table 11 Details of Challenges Faced by D-Log Data Processors

Challenge	Details
Missing attributes	There were a number of attributes that were commonly missed which required TfN to either complete manually or provide estimates. This was time intensive and, in some cases, could introduce inaccuracies when estimates are made.
Missing GIS data	As a result of some Local Authorities storing their location data only as GIS polygons, TfN had to convert these to points using GIS software in order to extract eastings and northings. This is time intensive and relies on TfN's GIS skills.
Receiving data in piecemeal manner	Many of the Local Authorities provided their data in parts due to time pressures, uncertainties over the data and the data being stored in different systems. This increased the burden on TfN to collate together datasets from a Local Authority and check which datasets may still be missing.
Hard to identify project progress	Due to the large number of Local Authorities and the piecemeal giving of data, it was difficult to monitor the progress of each Local Authority.
High levels of communication with data contributors	There was a large administrative overhead of collecting the data due to the receiving of data in a piecemeal manner as discussed above, queries about the data to be provided and chasing individuals to provide data.
Maintain and update complex Extract-Transform-Load (ETL) workflows	In order to overcome data provision issues such as attribute drop-downs being overwritten, text being provided instead of numbers and different units being provided, a large FME workbench was required. This was a large and complex ETL process to provide standardisation of data and fix errors. The maintenance and updating of this system is a large overhead and still did not catch all errors which then have to be handled manually.

DATA USERS

Table 12 Details of Challenges Faced by D-Log Data Users

Challenge	Details
Missing bedroom/housing sizes	Internal TfN modellers require the estimate of housing sizes as a proxy for trip rates. Many Local Authorities were not able to provide this data as either Local Plan data is too early in the development process to provide estimates or the Local Authority does not store this data.
Missing build-out of developments	Similar to the above, Local Authorities often cannot provide build-outs of developments which is required data for both TfN modellers and external stakeholders who use this for strategic level modelling.

Uncertainty over progress in delivery of developments	TfN have previously asked for both Local Plan and planning permission data however was not specific about which comes from each source and therefore some Local Authorities provide data from one source or another. It is therefore difficult to understand the total quantum of development prescribed by the Local Plan and through planning permissions, to what extent these have been achieved.
Includes irrelevant sized developments	For external stakeholders, they are only interested in strategic level developments and therefore they must filter out large amounts of the data. When this is shown on a map, it can be misleading as to the quantum of development that is of strategic significance.

4.2 I-LOG CHALLENGES

The I-Log data collection is at an earlier stage of development compared to D-Log. As the data has been collected within TfN there has not been a separate data processing stage. The issues below, therefore relate to data contributors within TfN.

DATA CONTRIBUTORS

Table 13 Details of Challenges Faced by I-Log Data Contributors

Challenge	Details
Time taken	Creating the initial list of interventions is a time-consuming manual process
Consolidation	There are a number of sources of intervention data creating overlaps. Consolidating these data sources is difficult (often only a name of an intervention to work on). There can be hundreds of interventions to review.
Providing route alignments	Deriving the GIS data has involved workshops to agree the correct alignments with heads up digitising onto basemaps. This requires personnel with GIS skills and others with extensive rail and road knowledge.
Ongoing data management	Ongoing maintenance of the data (i.e. updating records, adding new records) needs to be facilitated. Roles, responsibilities, procedures will be formalised as part of the TfN's Assurance Framework, which will be developed within the first quarter of the 2021 business year

DATA USERS

Table 14 Details of Challenges Faced by I-Log Data Users

Challenge	Details
Database schema	A schema is yet to be defined.
Frequency	There is currently no formal submission of data on a defined interval from data contributors. The submission interval decision together with roles,

	responsibilities mentioned above will be formalised as part of the TfN's Assurance Framework, which will be developed within the first quarter of the 2021 business year.
Confidentiality	If shared outside TfN, some data could be confidential. Route alignments can be especially sensitive.
GIS disconnected from data	The attribute data and GIS data are recorded on two separate systems which have to be cross-referenced to get the full information.
Associated documents	Several discussions raised the benefit of having any associated pdf documents or plans linked with interventions (where available) and where an individual has responsibility of ensuring these are up-to-date.

5   USER REQUIREMENTS

This section details the user requirements for the development of a solution and includes all functionality and interface requirements. Each requirement states whether it will be included within the Alpha Stage of development or future iterations.

5.1   D-LOG SPECIFIC REQUIREMENTS

The list below gives the user requirements specific to the D-Log dataset. This is to be used in conjunction to the user requirements in common with I-Log given later in this chapter

Table 15 D-Log Specific Requirements

Requirement	Details	Stage of Development
Add a new Local Plan record	A Local Plan record is the Local Plan as an entity unto itself and not each individual planning allocation site. Add a new Local Plan record including start and end date, adoption status etc	Alpha
Upload a PDF of a Local Plan	Upload the corresponding PDF of the Local Plan	Alpha
Upload a GIS shapefile of Local Plan sites	Upload a GIS shapefile of Local Plan sites and select unique identifier for areas	Alpha/Future Versions
Delete Local Plan record	Delete an existing Local Plan record	Future Versions, not required until new Local Plans are adopted
Edit Local Plan record	Edit a Local Plan record including uploads to ensure data is kept up-to-date especially the status of a plan	Future Versions, details unlikely to change once adopted
Add a new Local Plan Site record	A Local Plan Site record is an individual planning allocation site specified within a Local Plan. Add a new Local Plan Site record including all attributes with drop-downs and data validation	Alpha
Provide postcode in place of easting/northing	Ability to type in postcode to get easting/northings in the absence of the data being provided	Future Versions, could be found via external websites
Revise Local Plan Site record	Revise a Local Plan Site including attributes so the data is kept up-to-date	Alpha
Delete Local Plan Site record	Delete a Local Plan Site to ensure the list is kept up to date	Alpha

5.2 I-LOG SPECIFIC REQUIREMENTS

Table 16 I-Log Specific Requirements

Requirement	Details	Stage of Development
Add new intervention	Add main new intervention attribute details including name, description, source, status. An intervention can exist in the system without geometry information.	Alpha
Optionally, add sequencing information	This is optional as sequencing data does not apply to all interventions. It is required for those interventions that have reached this stage.	Alpha
Optionally, add rail modelling parameters	This is optional as it only applies to rail interventions where this information is known.	Alpha
(Optionally), add transport modelling parameters	Only required for those interventions where the information is known.	Future. Parameters are not yet defined.
Upload documents	Add pdf's, documents and images about the intervention	Alpha
Upload shapefile	Upload a shapefile of the route(s) of one or more interventions and identify the field(s) used as unique identifiers. The use of a corresponding unique field in the GIS and attribute data will enable a bulk upload of interventions.	Alpha
Optionally, generalise shapefile	Generalise shapefile if the route alignment is sensitive	Future. Not required for internal use.
Modify attribute data	Change any associated attributes	Alpha
Delete / replace shapefile	Remove with option to replace an existing alignment with a new one	Future. High priority item for the backlog

5.3 ADMIN

This section identifies the user requirements in relation to administrative tasks within a potential solution for both D-Log and I-Log.

Table 17 Details of Admin Requirements

Requirement	Details	Stage of Development
Define network location for document store	For linked pdf's, plans and other documents.	Future Version. An agreed network location can be used for the Alpha version.
Update lookup table	Add, rename or delete options used within drop-downs	Future Versions, this could be edited manually by development team.
Add new data attribute	Add a new data attribute to either the D-Log or I-Log schema. This includes an associated required lookup tables.	Future Versions, this could be edited manually by development team.
Delete a data attribute	Delete an existing data attribute from either the D-Log or I-Log schema. This includes an associated required lookup tables.	Future Versions, this could be deleted manually by development team.

5.4 USER EXPERIENCE

This section identifies the user requirements in relation to user experience within a potential solution for both D-Log and I-Log.

Table 18 Details of User Experience Requirements

Requirement	Details	Stage of Development
Modern design/clear interface	Ensure the design of the interface is clear and easy to read and that it is easy to navigate to increase uptake and reduce questions from stakeholders	Alpha
Branding	Branding for TfN to allow users to recognise they are in the correct solution	Alpha/ Future Versions, not necessary to have completely branded solution for Alpha release
Intuitive workflows	To ensure the workflows through the solution are intuitive so users have a clear idea of what they need to do and reduce questions to TfN	Alpha

Search for record	Ability to find specific records so data is easy to find	Alpha
Filter records	Ability to filter records by Local Authority	Future Versions, most Local Authorities do not have enough data to require filtering
Identify records that need review	Ability to identify which Local Plan Sites need to be reviewed according to a specified time frame	Alpha
Tool Tips/Help	Tool tips/help information to ensure users understand what data is required, to address any ambiguities especially where the user's data may vary from TfN's schema	Alpha

5.5 ACCESS REQUIREMENTS

This section identifies the user requirements in relation to users accessing a potential solution for both D-Log and I-Log. For the Alpha stage, it would require a small set of users (less than 20) to be able to test the application. Once the tool has developed further and rolled out to a production environment, then all stakeholders would be granted user logins.

Table 19 Details of Access Requirements

Requirement	Details	Stage of Development
Add User	Add user (normal and admin) so they can access the solution	Future Versions, a set list of users already exists and could be uploaded in the Alpha version
Delete (archive) User	Delete any user from the solution so they cannot access the data and for GDPR as we cannot store personal information for longer than required	Future Versions
Log in/out	To ensure the solution and data within it are secure and only accessed by users with credentials	Alpha
Change Password	To allows users to change their passwords so users can ensure their accounts cannot be compromised	Future Versions, could be undertaken by an admin
Forgot Password	To allow users reset their passwords, especially as the solution is likely to be used on an infrequent basis and stop TfN having an administrative burden	Future Versions, could be undertaken by an admin

5.6 REVIEW MECHANISM

This section identifies the user requirements in relation to the review mechanism of data within a potential solution for both D-Log and I-Log.

Table 20 Details of Review Mechanism Requirements

Requirement	Details	Stage of Development
Review Date	Show the review date for a record	Alpha
Reviewer	Show the person who made edits to a record	Alpha
Revise Data	Ability to change data, this may involve new inputs rather than editing existing depending on archiving requirements	Alpha
Remove Data	Remove records/archive them to ensure data is kept up-to-date	Alpha

5.7 EXPORT MECHANISM

This section identifies the user requirements in relation to the export mechanism of data within a potential solution for both D-Log and I-Log.

Table 21 Details of Export Mechanism Requirements

Requirement	Details	Stage of Development
Export list to spreadsheet	Must be able to export the data to allow for them to use the data in other systems. For example DfT need to be able to export the data aggregate it to Local Authority level and then put it as an input to their transport models.	Alpha

6 DATA REQUIREMENTS

This section details the data requirements for the development of a solution and includes all the data schema for the D-Log and I-Log and data formats. Each requirement states whether it will be included within the Alpha Stage of development or future iterations.

6.1 D-LOG STRUCTURE

The Discovery Phase has identified that Local Authorities tend to store their Local Plan data and planning permission data in different systems. In addition, the majority of Local Authorities when asked to provide planning data in previous D-Log versions have provided only Local Plan data and not planning permissions, often because they could not access planning databases. Therefore, for the Alpha development stage it is proposed that the most valuable dataset to collect is the Local Plan Site Allocations for each of the Local Authorities. This allows the TfN modellers to have a maximum development quantum and is the data that is required by external stakeholders.

Planning permissions which have similar structures but often more detail can be added in future iterations. A table of potential additional attributes has been included in a table below. In addition, there are a number of technology companies and external stakeholders who are working to address the challenge of collating together all planning permissions nationally and these may significantly develop in the near future. Developing the Alpha stage just for Local Plan data allows TfN to learn from other organisations' experiences and potentially build on their work in future.

The figure below illustrates a potential database structure for the D-Log dataset. This would be compromised of the two key elements:

- **Local Plan Upload:** the metadata about the Local Plan itself including adoption status and dates, a copy of the PDF and a shapefile containing the locations of the allocation sites.

- **Local Plan Site Allocation:** attributes about a specific site allocation including land use types, development quantum, proposed development years and likelihood of development

As shown in the figure, there are a number of look-up tables used in order to provide drop-downs for users to instil data conformity. In addition, most fields have a property of 'NOT NULL' meaning users must explicitly specify if they do not have that data and cannot leave it blank.

LOCAL PLAN DATA REQUIREMENTS

Figure 8 Proposed Local Plan Database Design
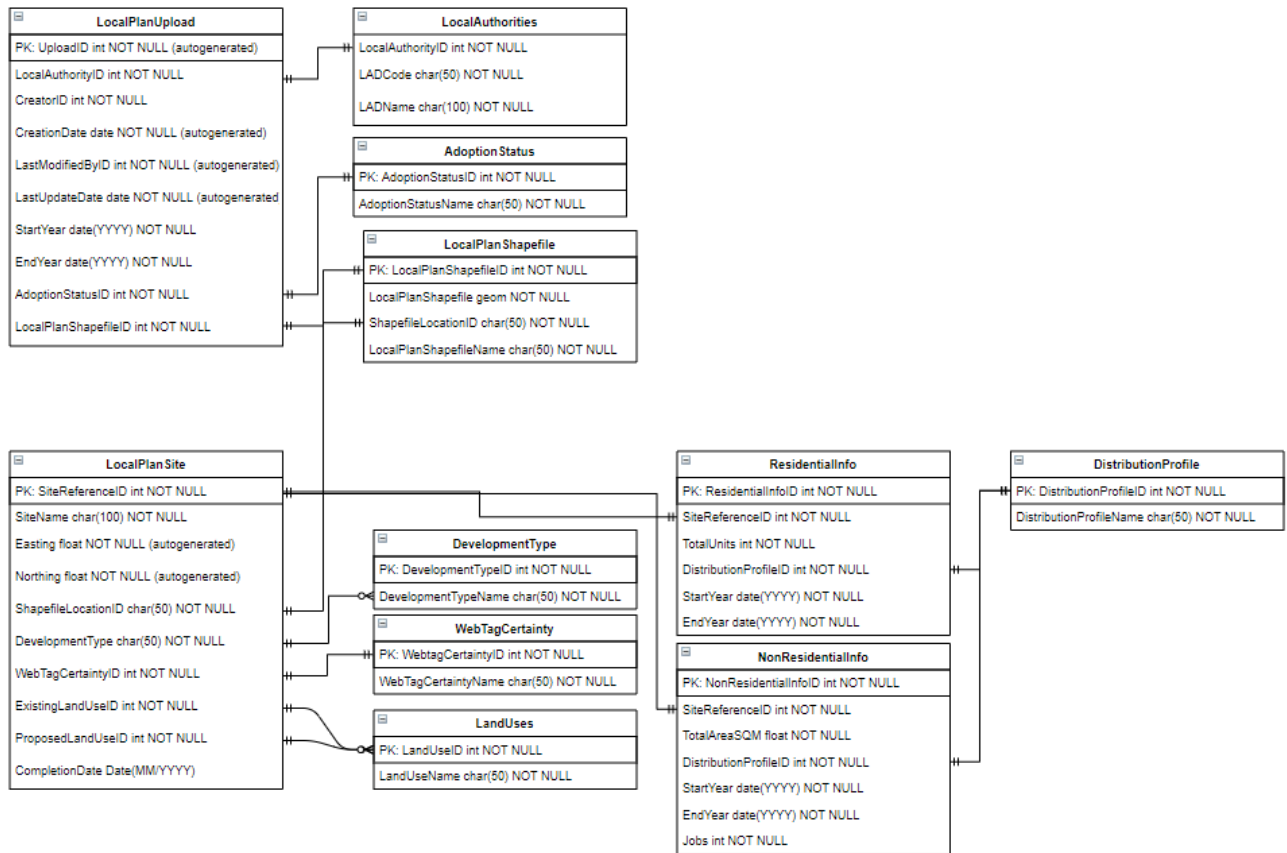
Planning Application Data Requirements

The Planning Application data shares many of the same data attributes as the Local Plan data but would include more detailed information as planning applications tend to be more detailed than higher level strategic Local Plan Site Allocation data. In addition, in future versions it would be beneficial to be able to add developments specified in Local Authority's Strategic Housing Land Availability Assessments (SHLAAs). The attributes would include the following and there would be a distinction made in the database as to whether the source of the data was a Local Plan, SHLAA or planning application.

Table 22 Proposed Additional Attributes for Planning Application Data Schema

Attribute	Details
PlanningRef Char(50) NOT NULL	Planning reference of application so can be cross-referenced
PlanningApplicationType Char(50) NOT NULL	Type of planning application eg. Outline, detailed etc as a proxy for level of detail expected

|	
Bedroom1Units Int NOT NULL	Proportion of total units with 1 bedroom
Bedroom2Units Int NOT NULL	Proportion of total units with 2 bedrooms
Bedroom3Units Int NOT NULL	Proportion of total units with 3 bedrooms
Bedroom4Units Int NOT NULL	Proportion of total units with 4 bedrooms
Bedroom5Units Int NOT NULL	Proportion of total units with 5 bedrooms

6.2 I-LOG STRUCTURE

An outline data model for I-Log data is presented in **Error! Reference source not found.Error! Reference source not found.**. There are 3 main tables:

- ■ **Interventions** – the basic information about an intervention, its status, and dates (if known)
- ■ **Sequencing** – this is optional data which allows the sequencing information to be entered just for those interventions that have been assessed.
- ■ **Rail Parameters** – optional information that can be added to rail interventions that have known details about the changes.

The other tables of **Mode**, **DeliveryBody**, Intervention**Status**, RouteStatus, **SequenceDates** and **Stations** are lookup lists to enforce consistency in data entry.

Road transport modelling parameters (discussed within Section 3.6) have not been included for the Alpha phase as they are not yet clearly defined and are likely to be more onerous to populate.

Figure 9 Proposed I-Log Data Model

Figure 9 Proposed I-Log Data Model
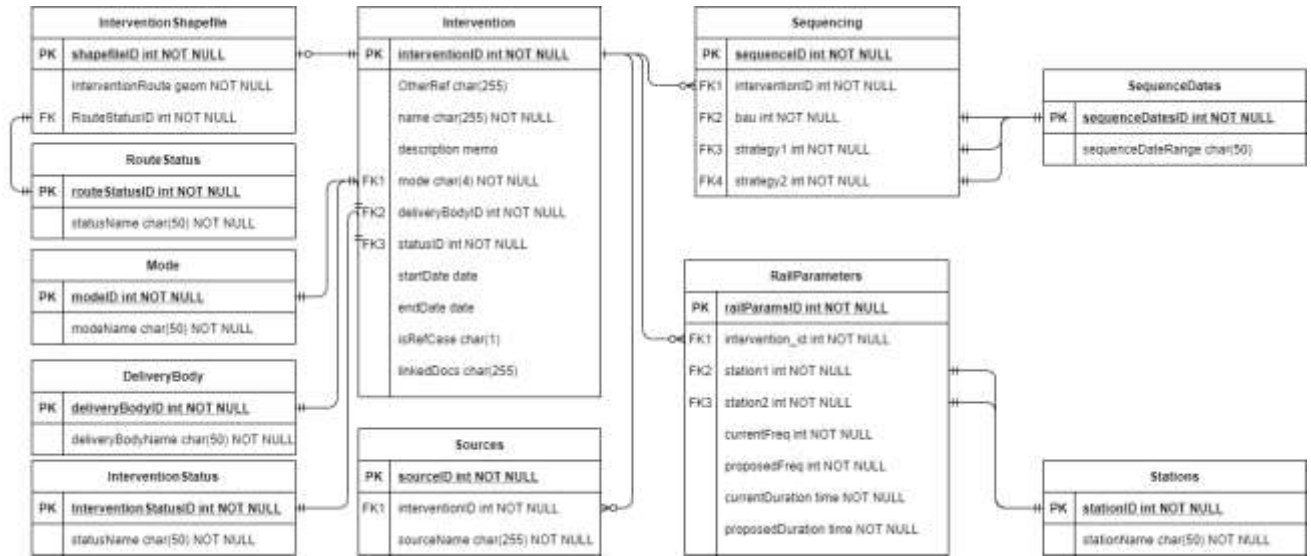
6.3   USER MANAGEMENT

A user management system is required to allow the solution to store personal information, for users to login to the system and be allocated a normal user or administrative user role. An initial database design for the user management element of the solution is given in the figure below and would be applicable for both D-Log and I-Log.

Figure 10 Proposed User Management Database Schema
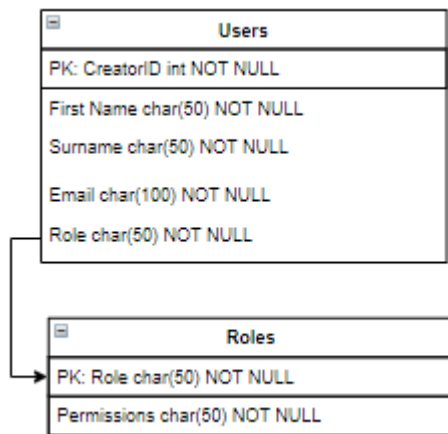
6.4   DATA MANAGEMENT

DATA HISTORY

The Discovery has identified a need to clearly audit changes to the data so that users can see a history of changes. This will provide important information to relay confidence in the data. The proposed table below identifies who, when and what has been created or changed. It is recommended that the modification description is auto-populated to ensure accurate and consistent descriptions.

Figure 11 Proposed Data History Database Schema

	Data History
PK	historyID int NOT NULL
FK1	dataID int NOT NULL
FK2	modifiedUserID int NOT NULL
	tableName char(50) NOT NULL
	modificationDate date NOT NULL
	modifcationDescription char(255) NOT NULL

DATA COMPLETION

The Discovery Phase has highlighted that the current data collection process leads to inaccuracies in the data or missing data which has led to a complex data cleansing process being required to ensure data integrity. Whilst efforts were made in the previous D-Log to use drop-downs to enforce data standards, due to the flexible nature of Excel spreadsheets, these data standards could be overwritten. This culminated in often having to manually manipulate the data which has been time intensive.

To improve on this, a form-based structure is proposed with strict data format controls including drop-downs, calendars and minimum values to ensure that data is in the required formats. This will reduce the need for data cleansing and ensure the data is what is required.

MISSING DATA

As part of the data integrity requirements, all values will be mandatory however it is recognised that some Local Authorities do not store some of the requested attributes and therefore these attributes will have 'unknown' or 'N/A' value options. This will ensure that there are no missing data records in the database and that all NULL values indicate a conscious recording that the data is not available.

DATA DISTRIBUTION FOR TEMPORAL DATA

The Discovery Phase indicated that temporal data has proven quite difficult to collect and manage. This is as a result of Local Authorities not producing their Local Plans to a uniform timetable and many did not have a clear idea of how large developments would be built out. In addition, from the conversation with DfT their internal evidence indicated there is significant inaccuracies with this profiling. It is therefore proposed that instead of users putting in specific housing units or areas for

non-residential land uses, instead to use a start year, end year and a distribution profile to help users give an indication as to how they think the development will be built out.

DATA OWNERSHIP

It is clear the data is dynamic and will need to be updated on a frequent basis to be of value to data end users. However the responsibility for the data updates and its accuracy would need to be agreed with both the data contributors and data users, once this has been agreed and understood this can be shared with all parties and it can then be taken into consideration when the data is used.

7 NON-FUNCTIONAL REQUIREMENTS

This section details the non-functional requirements for the Alpha Stage development and includes all the browser requirements and legal considerations.

7.1 DATA PROTECTION

GDPR

WSP as a Data Processor on behalf of TfN, the Data Controller, would need to implement the GDPR policies in accordance with TfN's policies. We would need to gain consent from the user to retain their Personal Information (First Name, Last Name, Email Address) for the purposes of accessing and using the tool. This would also include the need to delete users' details after a given period and this would need to be included as a technical requirement for the future solution.

SECURITY

High levels of security will be required for both the database containing personal information and operational data as well as the application itself. This is so that the solution is in accordance with Data Protection laws as well as maintaining adequate up-times and good reputational status.

7.2 PLATFORM

ACCESS

Any future solution would be required to be accessed by multiple different organisations and therefore must be an online application that is accessed from any network.

BROWSERS

The survey of Local Authorities has shown that all but one had access to modern internet browsers and therefore it will be developed for these. It is noted that Internet Explorer is no longer supported by Microsoft and therefore will not be accommodated.

7.3 EXISTING TFN INFRASTRUCTURE

HOSTING

TFN IT infrastructure is currently built upon Microsoft's Azure cloud platform.

DATABASES

The D-Log data is currently stored in the TfN corporate database, a hosted PostgreSQL service on Azure. The database is spatially enabled with PostGIS version 2.4.

TfN also have an installation of ArcGIS Enterprise to enable the viewing, querying and sharing of the spatial data held within the PostgreSQL database.

8 ALPHA PHASE REQUIREMENTS

8.1 INTRODUCTION

This section outlines the requirements for a minimum viable product (MVP) for D-Log and I-Log. A MVP is a version of a solution with just enough features to be usable by early customers who can then provide feedback for future solution development. This approach is being taken to address the following:

■ The existing D-Log and I-Log datasets are in their infancy and are not currently used by stakeholders for operational purposes and therefore stakeholders are not able to provide all potential requirements or future improvements

■ It allows both internal and external stakeholders to become comfortable with using the solution before potentially increasing the complexity

■ It gives TfN the opportunity to gain feedback from stakeholders to ensure the solution is intuitive and useful before rolling out a solution into production

■ There are a number of technology companies and external stakeholders who are working to address the challenge of collating together all planning permissions nationally and these may significantly develop in the near future. Developing a MVP allows TfN to learn from other organisations' experiences and potentially build on their work in future

8.2 PROPOSED SOLUTION

The Discovering Phase has identified a large number of user, technical and data requirements, which are detailed in the chapters above. For the Alpha Phase, the requirements which provide the maximum value to end-users have been prioritised which can then be tested by users, revised and built upon in future versions. The section below outlines the proposed solution which would address the Alpha Phase requirements.

It has been recognised that the key users of D-Log and I-Log are different stakeholders and that the contents of each dataset are different however large overlaps have been identified as to their user and technical requirements. Fundamentally they both require a data collection and review mechanism, a named user login system to prevent non-stakeholders accessing the data and provide accountability and the ability to export data or use the data within other systems. Therefore, the proposed system is for both D-Log and I-Log with a recognition that users may only have access to one out of the two datasets.

INTERFACE

■ Online application accessible via modern internet browsers

■ Intuitively designed application focused on the key workflows of adding, reviewing and updating data records

■ Form based approach to data entry with high levels of validation by means of drop-downs, date pickers and format restrictions

■ Ability to upload PDFs, documents and shapefiles

■ Ability to review existing data, make updates and delete irrelevant data

- Ability to export data for use in other systems

- Admin ability to add users

- Admin ability to review data update progress

ADMIN

- All users must be named, have a login, be granted access to one or both of the datasets and a role

- Two roles required – Data Contributor and Admin

DATA

- All data to be stored in databases

- Where possible all data should be a 'NOT NULL' attribute whereby all missing data must be explicitly identified

- Database is secured

- Database is interoperable with other systems and support batch exports

- Database structures to be based on indicative designs given in Data Requirements

9 RISK REGISTER

This section notes the risks that have been identified throughout this Discovery Phase which may have implications for future phases of development.

Table 23 Risk Register

Risk	Details	Probability	Mitigation
Pre-populated D-Log20 data	To minimise the effort required by data contributors, the working version of the application for D-Log21 should include the existing data from D-Log20	Low/Medium	ETL process to be developed to convert the existing database to the new schema
Register of development data being addressed by MHCLG	MHCLG are currently focused on standardising the planning data provided by Local Authorities. This is currently at the designing of the data schema stage and is not due for implementation till 2022/3.	Medium	Continuing communication with MHCLG to understand their progress and see opportunities to collaborate

8 First Street
Manchester
M15 4RP

wsp.com