May 2022

Department for Transport – Active Travel Discovery Report

CGI IT UK Ltd

14th Floor,

20 Fenchurch Street,

London

EC3M 3BY, UK

Table of contents

1 Introduction

The Department for Transport (DfT) Active Travel Team is providing significant investment in Active Travel (AT) Infrastructure, largely through Local Authorities.   Data is needed to understand the outcomes of AT funding, to understand where the AT infrastructure that has been built, what benefits it provides to whom and how it links with other forms of transport.

This report is the output from an 8-week Discovery Phase, aiming to answer the following questions

- Is there a service that can be built that meets user needs?
- Roughly how long will it take and how much will it cost to develop that service?
- What should DfT be exploring in Alpha to consider options, evaluate opportunities and mitigate identified risks?

The discovery team is cross-functional team including representatives from the DfT Active travel Statistics team and from CGI as the IT partner for this phase.

They have undertaken interviews with over 20 key stakeholders including local and combined authorities, data providers and consumers to establish a consensus view of what information and processes already exist. As well as considering the potential users and contributors to such a service, the wider digital architecture has also been reviewed to ensure that a service will align to the standards and guidelines.

The team have used this information to build a view of the gaps and opportunities which exist to answer the questions posed above. This discovery report describes those findings.

2 Executive summary

2.1 Introduction

The Active Travel Statistics team launched the Discovery with the purpose of exploring spatial data of cycling infrastructure in England. This was refined to the core question of **'Where is the segregated/unsegregated cycling infrastructure in England at any given point in time, and how was it funded?'**. In the project brief on the Digital Opportunities and Specialists (DOS) marketplace, four initial questions were posed as objectives; the findings are summarised below. Whilst completing this project, the team considered related historic and ongoing work undertaken by other teams, taking account of lessons learnt to ensure a successful end product that meets user needs.

Objective 1) What is the best approach to collecting and updating the data? Five different data collection options have been identified: OpenStreetMap, algorithmic methods, surveying of the infrastructure, augmentation of data on recent developments and review for completeness with a survey to enhance data. Each of these options have their own strengths and weaknesses, the best approach will depend on the customer, to balance their priorities for cost, timings and accuracy of data. Section 3.5.2.4 and 3.5.3.3 provide more detail.

Objective 2) What data already exists? Two groups of data have been identified. The first is spatial data on the location of the cycle infrastructure. Some Local and Combined Authorities collect this data as paper maps, online static maps or GIS layers. The second group is information about the type and benefits of the infrastructure, to derive understanding about the needs and successes of schemes. It should be noted that across both these data groups, there are varying levels of completeness and data collection, and storage methods are inconsistent. Section 3.5.1 provides more detail.

Objective 3) What are the key technical challenges? Most data collected by local authorities does not have licencing constraints, but considerations will need to be made for third party data such as Ordnance Survey data and use of the Public Sector Geospatial Agreement. Some specific datasets which have been specially commissioned by DfT and Local Authorities, do have intellectual property rights and therefore have licensing constraints. Section 3.5.4 provides more detail.

Objective 4) What are the barriers to a solution? The solution will need to take into consideration the capacity of Local Authorities and benefits of using the solution will need to be clearly demonstrated to all key stakeholders, to ensure the tools use. Many different capabilities already exist and are widely used by stakeholders, such as the propensity to cycle tool. The solution will need to interoperate capabilities that already exist without reinventing them. Section 3.5.5 provides more detail.

2.2 Approach

Four research activities were conducted as part of the Discovery: a literature review, an initial workshop with core stakeholders, wider stakeholder interviews and a survey. These activities informed the creation of five products, that were used to answer the core question and the four DOS objectives, these products are:

Stakeholder personas

The research was used to develop four stakeholder personas: Local Authorities, Active Travel policy, DfT Active Travel analysts and DfT Active Travel statistics. They provide a base understanding of the core stakeholders. They include information on who the user collaborates with (Objective 1), the data they use (Objective 2) as well as their ambitions and challenges for the solution (Objective 3, Objective 4).

Stakeholder map

The stakeholder map shows in greater detail how the stakeholder personas connect to each other. This was created to understand the links, flows and exchanges between stakeholders, in the creation and exchange of data. This develops understanding of what data exists, who holds it (Objective 2) and helps to understand what the best approach to collect and share data might be (Objective 1) as well as what any barriers to a solution might be (Objective 4).

User journeys

The user journeys provide greater detail to the stakeholder map, summarising the 20 interviews conducted into the four stakeholder personas and converting interview transcripts into narratives of daily tasks, flows and connections. The user journeys develop understanding of stakeholder's roles, needs and pain points of the current processes, to generate opportunities for the tool. In doing so, the user journeys address all four DOS objectives.

Data catalogue

The data catalogue lists the available data to understand what data currently exists (Objective 2) but it also informs the technical challenges (Objective 3) as it documents the format, geographical extent, availability and status of the data.

Service blueprint

The service blueprint draws together the detailed information from the user journey and presents it alongside the connections between different organisations and users, from the stakeholder map. The service blueprint details the current policies and processes LAs carry out and how these feed into DfT processes, the needs of the tool are identified and the opportunities for the tool are highlighted. As the service blueprint summaries all the information above, it addresses all four objectives.

2.3 Summary of Options

2.3.1 Data collection

Five different options have been identified to collect data on the location of cycle infrastructure and associated metadata. Across all the options, the following factors are broadly similar: no significant licencing issues, good support of service standards, relatively good ease of correction, relatively good breath of data and relatively good documentation.

The best approach will depend on the customer's priorities for balancing low cost, fast implementation and ease of obtaining high accuracy data, as these vary the most between the options. For example, collecting data by improving reporting methods is the least expensive solution, requiring the least effort but will take longer to obtain full coverage of the cycling network, compared to a LA or third-party survey, which would likely be more expensive but quicker to complete the dataset.

	Data Collection Options				
	OpenStreetMap	3rd Party Survey	LA Survey	Algorithmic methods	Improved reporting during developments
Effort to collect	Moderate	Low to high: Can scale dependent on spend	High	Low to high: Variable, dependent on attributes	Low
Time to collect	Moderate	Moderate to high	High to Very High	Moderate	Very high
Cost to get full coverage	Low	Very high	High	Low to moderate	Low

2.3.2 Technology

Five different technological options have been identified to host the potential solution. All five options have low maintenance costs, relatively good alignment with architecture guidelines and good ability to accommodate up-to-date data. The option of replacing LA systems with a single system is included for completeness and less so for consideration, as it is likely to be very disruptive and is therefore it is shown in italics in the table below.

The best approach will depend on the customer's priorities for balancing cost and ease of implementation. The desire for an open-source tool and the need to consider capacity of local authorities were often mentioned throughout the interviews from a range of different stakeholder types.

Total costs vary significantly between the options, a new open-source tool would be the least expensive option and replacing LA systems with a single system is likely to be the most expensive.

Not all the options allow for an open-source tool and some options would have impacts on organisations, in particular option five which is included for completeness.

	Technology Options				
	New Commercial GIS (ArcGIS Online)	New Open-source	Expand on Existing Tools	Integrate and re-platform existing tools	*Replace LA systems with single system*
Initial licencing cost	Moderate-High	Low	Moderate	Moderate	*High*
Ongoing licencing cost	Moderate to high	Low	Low	Low	*Moderate to high*
Initial implementation cost	Low	Low to moderate	Moderate	Moderate to high	*Very high*
Maintenance cost	Low	Low	Low	Low	*Low*
Currency of data	Very high	Very high	Very high	Very high	*Very high*
Degree of business change on LAs	Low	Low	Low	Low	*High*
Impact on other parties	Low	Low	Moderate to high	High	*Low*
Complexity of implementation	Low	Moderate	Moderate to high	High	*Very high*

3 Architecture approach

The approach to collecting and processing data covers a number of areas which are broken out in the document. The structure used to define and document all of these elements conforms to an IT Architecture Framework

3.1 Architecture framework

Using a standard approach provides consistency, allowing those familiar with the approach to understand where particular elements are defined, and also ensures that relevant aspects are considered in a logical order.

The following excerpts[i] from The Open Group architecture framework (TOGAF), tackle the question of why the use of any framework is desirable:

> ***Why do I need a Framework for IT architecture?***
>
> *Using an architectural framework will speed up and simplify architecture development, ensure more complete coverage of the designed solution, and make certain that the architecture selected allows for future growth in response to the needs of the business.*
>
> *Architecture design is a technically complex process, and the design of heterogeneous, multi-vendor architectures is particularly complex. TOGAF plays an important role in helping to "demystify" the architecture development process, enabling IT users to build genuinely open systems-based solutions to their business needs.*
>
> ***Why is this important?***
>
> *Those IT customers who do not invest in IT architecture typically find themselves pushed inexorably to single-supplier solutions in order to ensure an integrated solution. At that point, no matter how ostensibly "open" any single supplier's products may be in terms of adherence to standards, the customer will be unable to realize the potential benefits of truly heterogeneous, multi-vendor open systems.*
>
> ***What specifically would prompt me to develop an architecture?***
>
> *Typically, an architecture is developed because key people have concerns that need to be addressed by the IT systems within the organization. Such people are commonly referred to as the stakeholders in the system. The role of the architect is to address these concerns, by identifying and refining the requirements that the stakeholders have, developing views of the architecture that show how the concerns and the requirements are going to be addressed, and by showing the trade-offs that are going to be made in reconciling the potentially conflicting concerns of different stakeholders.*
>
> *Without the architecture, it is highly unlikely that all the concerns and requirements will be considered and met.*

The Open Group Architecture Framework (TOGAF) Architecture Definition Method has been used because it is the pre-eminent non-proprietary industry standard for defining and documenting IT architectures.

3.2 The TOGAF standard

The analysis included here broadly aligns to the first 5 areas of the TOGAF Architecture Model , shown as A-E on this diagram.

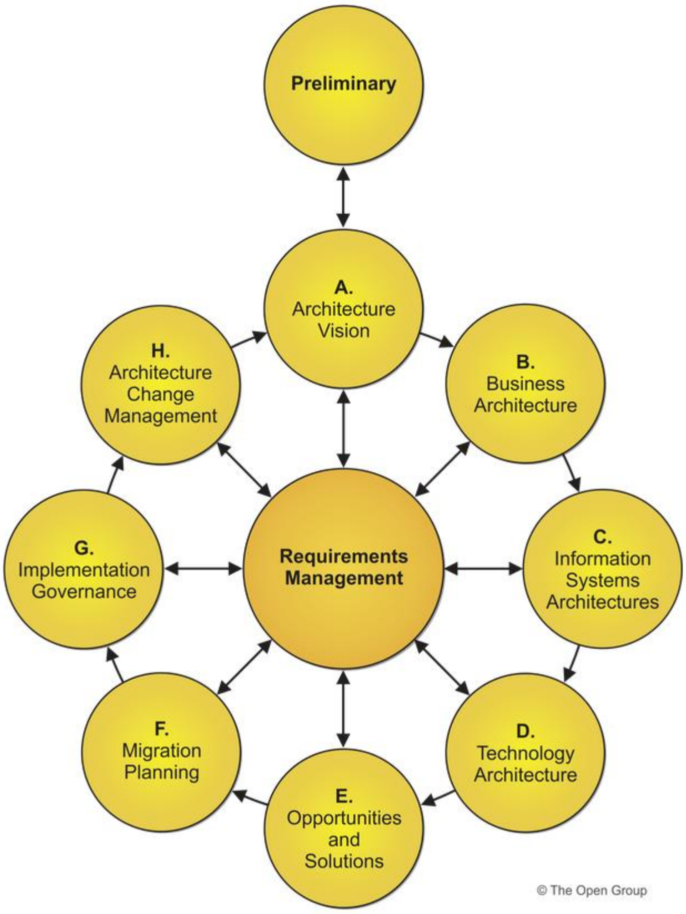

Figure 1 - TOGAF Architecture Definition Method

TOGAF is an industry-standard framework for defining and documenting architectural aspects of systems and processes.

Further details of the TOGAF can be seen at https://www.opengroup.org/togaf

3.3 Architecture Vision

The scope of work identified the key problem as being this:

"There is currently no definitive spatial data source of active travel infrastructure at a local level in England, so the active travel statistics team want to explore solutions for collecting this data in order to produce a map for England. We are looking to understand where (and what types) of infrastructure exists and how this links with other transport modes. We are looking to understand what data already exists from LAs, industry bodies and commercial providers and identify gaps and challenges in the use and further improvement to this data.

Using this knowledge, we wish to understand how best to define a baseline of active travel spatial data and a methodology to keep this updated to enable us to create a map and analyse the data."

Over the course of the discovery, this was refined to a single immediate question:

"Where is the segregated/unsegregated cycling infrastructure in England at any given point in time, and how was it funded?"

This focus was validated during the interview and survey phase, where 67% of respondents agreed that if they had information on this it would make completing some of their work tasks easier.

The focus on cycling[ii] infrastructure is because this is considered a simpler and more defined scope. However, it is understood that other travelling modes (mostly walking) are also of interest, and that the proposed solution should be suitable for including that information as well.

The primary business requirements of the application to get to the Minimal Viable Product are:

- To gain a clear view of what relevant cycling infrastructure exists
- To understand, as much as possible, the defining attributes of the infrastructure per the LTN1/20 guidance (size, position, whether segregated etc)
- To build a data maintenance and collection process which keeps the information current while not imposing onerous and expensive work onto the local authorities, who are currently one of the main information providers

The business drivers of the potential application are outlined in section 0, and the stakeholders are documented in section 3.4.

Beyond that initial objective, there are two further questions which are expected to be answered through further development:

"What is the quality/safety/convenience/accessibility of cycling infrastructure (including specific compliance with LTN 120, but also other criteria)"

"What are the benefits/outcomes derived from the cycling infrastructure (and ultimately from the funding), in terms of usage and health, carbon footprint, air quality, traffic congestion, economy etc"

These lead to subsidiary objectives for the application, beyond the initial MVP

- To fully understand the cycling infrastructure with all the defining attributes of the LTN1/20 guidance
- To enable the efficient gathering, analysis and objective comparison of projects undertaken to deliver active travel infrastructure, including their cost/benefits analysis
- To extend the infrastructure definitions to include walking route

3.4 Business Architecture

At a high level, the basic data and process needs of the application are described in this section, beginning with a high-level summary of the interactions between various stakeholders and then going into detail of the work processes and key stakeholders involved.

3.4.1 High-level Business Architecture

The diagram below summarises the main flows of information and organisation between the various interested parties. These are then described in more detail in the following sections

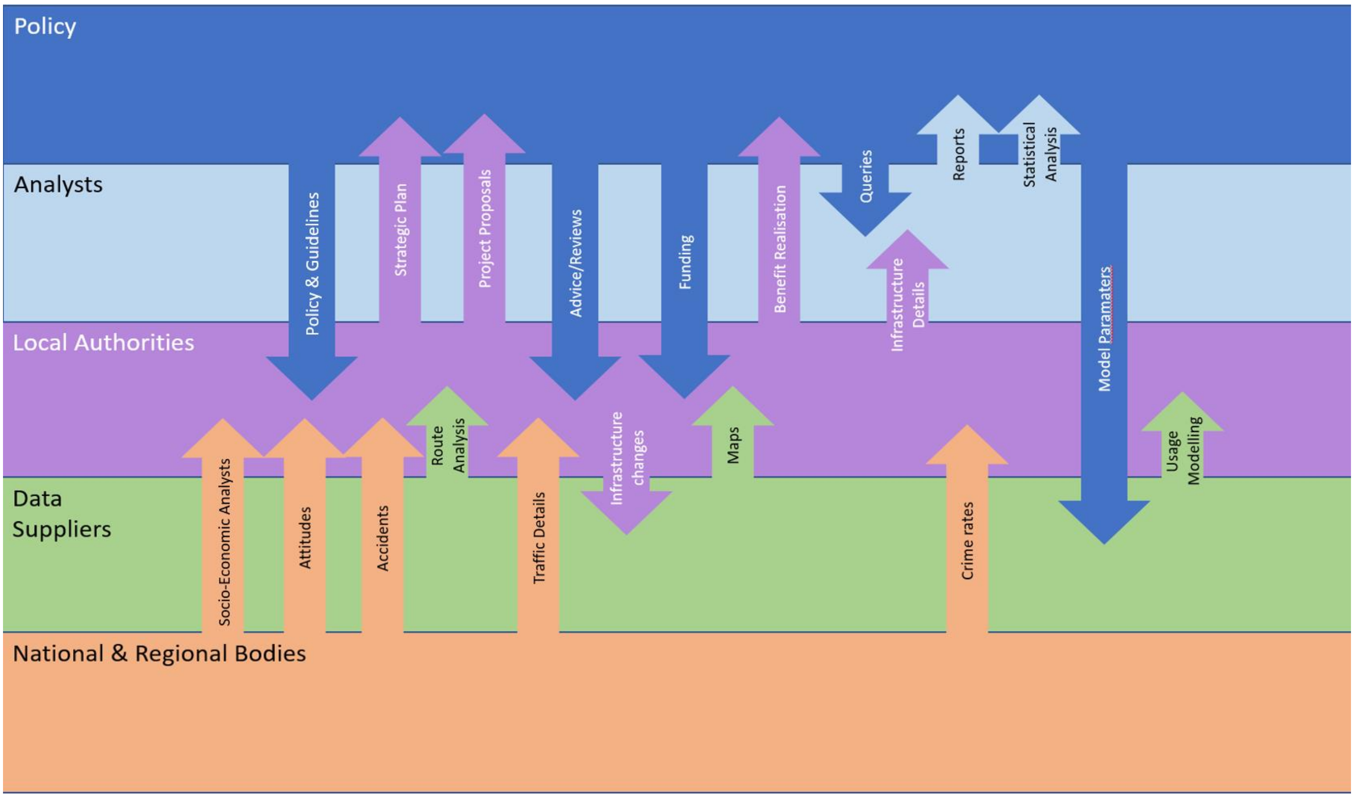

Figure 2 - High-Level Business Process

3.4.2 Stakeholders

3.4.2.1 Stakeholder Map

The first element of the user research was to establish a stakeholder map, which is a visual representation of the ecosystem of stakeholders involved in the potential usage of the future AT data tool. It helps to understand who is involved, to reveal existing formal and informal relationships and exchanges between stakeholders, to identify first blockers between them, and to find unknown relationships, fostering existing ones, or creating alternatives if needed.

We needed to identify all important data users and providers. As well as identifying these actors, it is important to map their interactions. Through the interviews we were able to understand which items and services are exchanged and between whom.

The stakeholder helps map shows a high-level diagram of the most important actors and their exchanges. It is designed with proximity of use to the potential future tool.

As a result, we can communicate to the team and other stakeholders the complex information exchanges and high-level data relationships, and we were able to prioritize and focus on the most important data users and providers as well as their needs and expectations on the tool and data sets.

3.4.2.2 Contributing Stakeholder Locations

During the Discovery, a series of interviews were undertaken with a number of Local Authority stakeholders. These were primarily selected to be as representative a sample as possible considering geography, stage of active travel adoption, and economic base of their populace, though this was also influenced by availability of authorities during the discovery process.

Following on from these interviews, a survey was conducted which collected information from a further set of Local Authorities. The geographic spread of these contributors is shown in the map below.

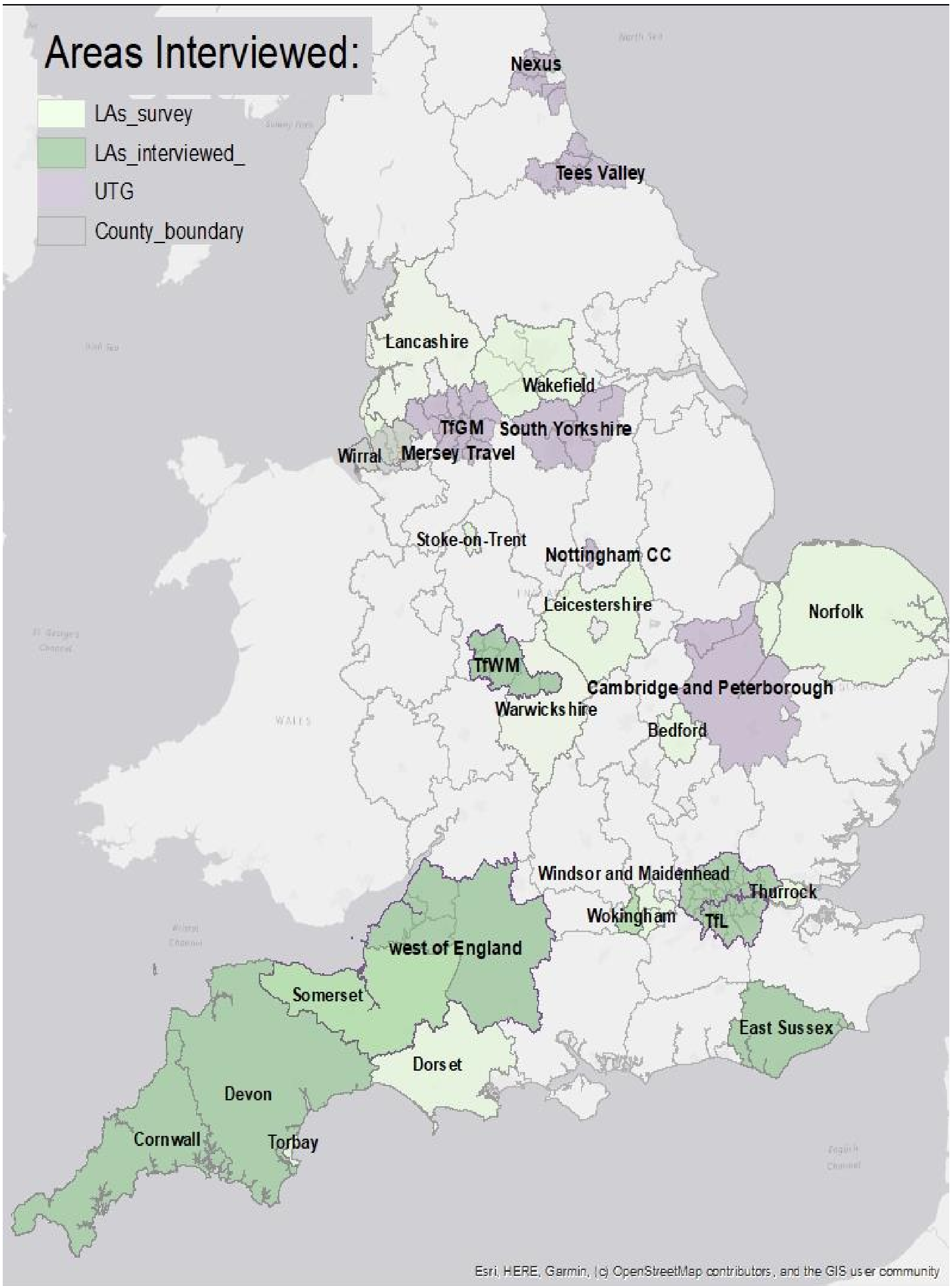

Figure 3 - Contributing LA stakeholders

3.4.3 Stakeholder Profiles

A stakeholder profile is a visual document used to identify specific stakeholders and describe them in terms of their influence and interest in the project. It's crucial to know your stakeholders, as their interest can have a positive or negative impact on the project execution.

The stakeholder profiles are an extension of the stakeholder map and give more insight into who these are. Creating these profiles helps the team to deepen our understanding of users' needs, experiences, behaviours, and goals.

We have collected an array of information that is related to each stakeholder's AT journey and involvement, to paint a complete picture. As well as understanding their current work, we also captured their dreams and aspirations as well as challenges they face.

3.4.4 Blueprint diagram

A service blueprint is a diagram that displays the entire process of service delivery. In our case the 'service' is provided by the potential data tool to all stakeholders. Since the tool does not exist yet, we have mapped data challenges and data opportunities into the diagram.

The service blueprint is built by first listing all the key stakeholders involved in the data tool usage process on a vertical axis, and all the processes where the data tool is used on the horizontal axis. The resulting matrix allows us to represent the flow of actions that each stakeholder performs along the process in relationship to the other stakeholder's processes. For deeper understanding we have included the 'tool needs' extracted from the user journeys.

The Service Blueprint tells us a high-level story of how all involved stakeholders interact and how at each step they could potentially benefit from the data tool. Through mapping the data challenges and opportunities we can see that all stakeholders will equally have the same data challenges and benefit from a tool in a similar way. The blueprint also shows clearly where the stakeholder processes interact. Those interaction could be made easier and faster as outlined in the data opportunities.

3.4.5 User Journeys

User journey maps are a visualisation of a stakeholder's relationships with a product (potential data tool) over time. We have gained deep insights through over 20 interviews and 21 survey responses.

We have included different aspects in the user journeys and therefore created rich and in-depth documents. The top rows show infographics explaining each stakeholder's main process with some more details in the rows below. This section is mostly followed by a section which looks at the AT tool usage within each step of the process. The bottom part is analysis and most important. It outlines pain-points, user needs and opportunities.

The user journeys are a main tool to identify the opportunities which will ultimately influence the further User Experience Design and defines important features.

We have learnt that even though processes are different many pain points, user needs, and opportunities align between the main stakeholders. We were therefore able to identify the main data challenges and main data opportunities.

3.4.5.1 Main Data Challenges

There was a considerable degree of commonality across users in their perception of the challenges that they face:

- Getting a consolidated view of the infrastructure, both centrally and for LAs
- Getting a complete view – being able to differentiate between completeness of representation of the infrastructure at all, and then completeness of attributes (shared/segregated/road surface etc). Although these both represent 'completeness' of the data, they are two different aspects, the first defining the

existence of AT infrastructure in a particular location, and the seconds providing richer details about the characteristics and quality of that infrastructure.

- Getting a view of the cost and benefits of projects, in a way which is easily comparable across projects. This is important both to central analysts and policy makers to determine the value for money of past and proposed projects but could also help local authorities to understand best practise and similar projects in other localities.

3.4.5.2 Main Data Opportunities

Similarly, the view of the main data opportunities was similar across user groups as well:

- Accessing data from a shared platform would allow all users to know where data is stored and have self-service access to it rather than being dependent on asking experts to derive data for them.
- Centrally stored data would also allow users to apply unified data formats. This would allow all stakeholders to use meta data in their own tools and process the data sets according to their needs.
- Getting a complete view: Knowing where the infrastructure is would allow all stakeholders to save time in their processes. LAs could use this data to develop their bids and case studies, statisticians would be able to answer requests quicker, policy and analysts would be able to provide funding and form the right and supportive policy documents.
- Cost and Benefits: Knowing what has been spent where and which impact the scheme had would allow all stakeholders to define success, prioritise in the right way, provide support with the right policies and schemes where needed and support the LA in their ATE journey

3.4.6 Key User Needs

User needs are one element of the user journeys, these needs were relatively consistent across different user types, including those interviewed and those responding to the survey.

Over 50 different data and tool needs were identified from the interviews and surveys. The needs listed below are the ones that were mentioned the most. Although not specifically mentioned here, understanding the gaps and ‘dead ends’ in the current cycle network was a key theme through the interviews.

3.4.6.1 Data Needs

Item reported	Percentage of survey respondents that would find this data “incredibly useful”
Condition of cycle path (eg whether LTN1/20 compliant)	88%
Number of people cycling and wa king	61%
Footway width	65%
Carriageway width	59%
Traffic levels (especially identifying low traffic streets)	85%

3.4.6.2 Tool Needs

Item reported
Access to underlying data (socio-economic segregation, health levels, index of multiple deprivation, etc)
Public facing and office version
Ability to share with other organisations
Ability to contact data providers for clarifications or corrections
Ability to query the data at different geographical levels (region, county, ward, postcode)

The survey found that most respondents would value the ability to segment their data by multiple dimensions. It also identified that currently 80% of respondents could only share data by email attachments.

3.5 Information Systems Architecture

This section details the current and planned data and application architectures and outlines the strengths and weaknesses of each of the design options.

3.5.1 Baseline Data Architecture

The current data architecture consists of two elements – the data stores and data flows, and the extent to which they satisfy the needs expressed in the business architecture, above.

3.5.1.1 Key Data Flows

This diagram shows the current key data flows between the central bodies and local authorities, and some of the providers for the data, at a conceptual level.

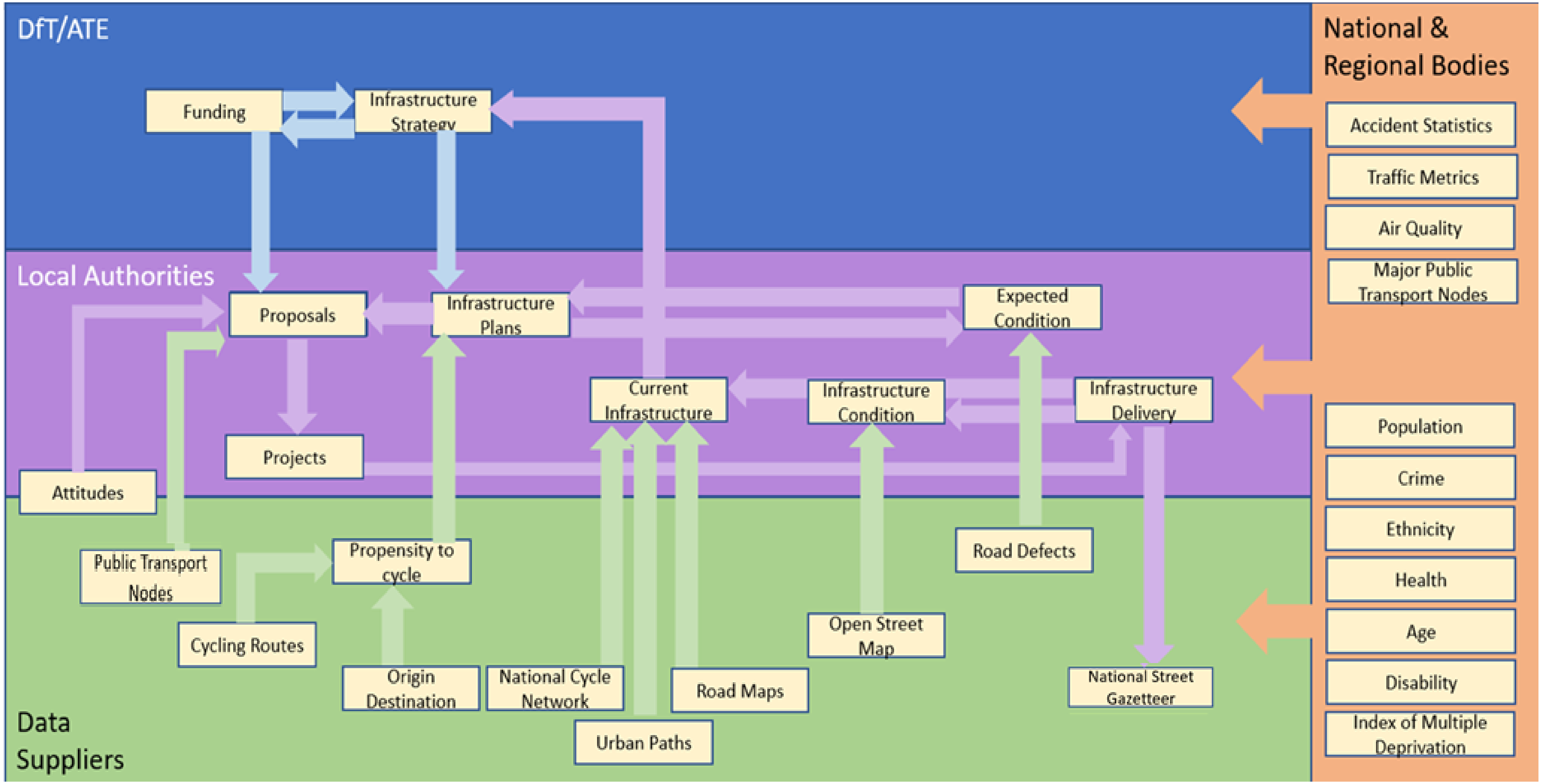

Figure 4 - Key Data Flows - Baseline

3.5.1.2 Data Completeness: LA data

Currently, most local level cycling infrastructure data is held by local authorities (LAs), although some third-parties are beginning to explore ways to collect this data. It is clear that there are varying levels of completeness of data currently held by LAs, broadly corresponding to the age of the active travel infrastructure, but also influenced by the amount of investment in that infrastructure and in mapping it.

- For ongoing and planned projects, it is understood that data needs to be collected along the lines described in LTN1/20, to include attributes describing the quality of the infrastructure, and that this is largely built into the projects.
- For maintenance of existing routes, it is more likely that the full breadth of quality information is captured, but this is not necessarily embedded into the maintenance processes.
- For infrastructure which has 'recently' been delivered, that is since the LTN 2008 guidance from DfT was issued, more detailed information on the infrastructure is likely to have been captured, but to a lesser degree than specified in LTN1/20.
- For older infrastructure it is *likely* that information is captured, at best, only in a fairly superficial level – that the infrastructure exists in a particular location.
- Other methods are currently used to supplement data created by the LAs themselves – including Strava, surveys, crowd-sourced data (particularly OpenStreetMap) and using other resources which intersect with their network (for example National Cycle Network).

These are necessarily generalisations, and the specifics of how populated each class of data is will vary between local authorities, but the fundamental point is that there are wide disparities in the amount and quality of data available on the existing infrastructure, which leads to choices on how and whether this is all to be brought up to the maximum level of standards and completeness to help establish a baseline of existing infrastructure

3.5.1.3 Digitisation

Although most (80+%) of the LAs that have responded to interview or survey have all held this data in some form of GIS system, some LAs do not currently have this capability – where they have the information at all, it is likely that it exists on paper or non-GIS systems (Excel or Word documents or drawing tool).

Even the respondents that had digitised their data, 90% of them still used static maps as one of their methods of data storage.

This is obviously a challenge for creating both a centralised view of what exists and of embedding a reliable maintenance process into the LAs so that this data is kept up to date.

3.5.1.4 Local knowledge

One of the key learnings from discussions with Local Authority stakeholders, is that some important information is not currently recorded in their records of active travel infrastructure. Rather, this knowledge is held by specific people who have experience in a particular area – examples include things like:

- Where certain types of obstacles (stiles, steps, bollards, width restrictions) exist on older infrastructure, or on rights of way
- Current ownership of disused bridges, tracks etc which were formerly managed by Network Rail or Highways Agencies

This kind of institutional knowledge is invaluable but is dependent on the continuity of specific people in their roles.

3.5.1.5 Licencing

The licensing conditions of the key data which could support the aims of identifying the infrastructure fall into one of the following categories:

- Data on the local infrastructure created or managed by local authorities – this data is freely shareable, subject to the caveats about sharing with the general public, discussed in the next section.

- Data provided by Ordinance Survey in the form of various maps – these are widely usable internally by government agencies under the terms of the Public Sector Geospatial Agreement (PSGA) but may have some restrictions on usage for tools that are public-facing.
- Data derived from open-source repositories – examples are OpenStreetMap and the data provided by CycleStreets; this data is open and freely shareable by definition.
- Data which is collated and published centrally, generally from data.gov.uk – this should all be freely usable for the purposes envisaged for this application.
- Data which has been explicitly commissioned by LAs, DfT or other government agency, which the intellectual property rights may reside with the provider.

There do not appear to be significant licensing issues which would prevent the target architecture being achieved.

3.5.1.6 Data Sharing

A recurring theme during the discovery interviews with both local authorities and central users, is the need to differentiate between data at three different stages of publication:

- Data about delivered active travel infrastructure
- Data about existing projects – the public will have been consulted, where appropriate, and proposals will have been published
- Data about aspirational projects – potential projects which have not been socialised outside planning and active travel departments, and which may or may not progress to projects.

The first two of these are widely shareable, both with local and combined authorities and the central planners / analysts and with the wider public. However, the third set will need to be controlled so that it is not accessible outside strictly controlled users.

There do not appear to be significant privacy or security issues with consolidating the data into a single repository, provided that it contains the capability for selective access based on role.

It is notable that although the LAs often have a need to share data of this type[iii], some reported that they did not have a good mechanism for sharing the data, with 80% of survey respondents relying on email attachments. This is an obvious opportunity for process improvement, if all relevant participants can see the same data in the same system.

3.5.2 Target Data Architecture

This section describes the target data architecture for the basic implementation - it includes the shared repository for active travel infrastructure which constitutes the proposed Minimal Viable Product (MVP), as well as some possible options for other capabilities which could be developed onto this platform.

3.5.2.1 Key Data Flows

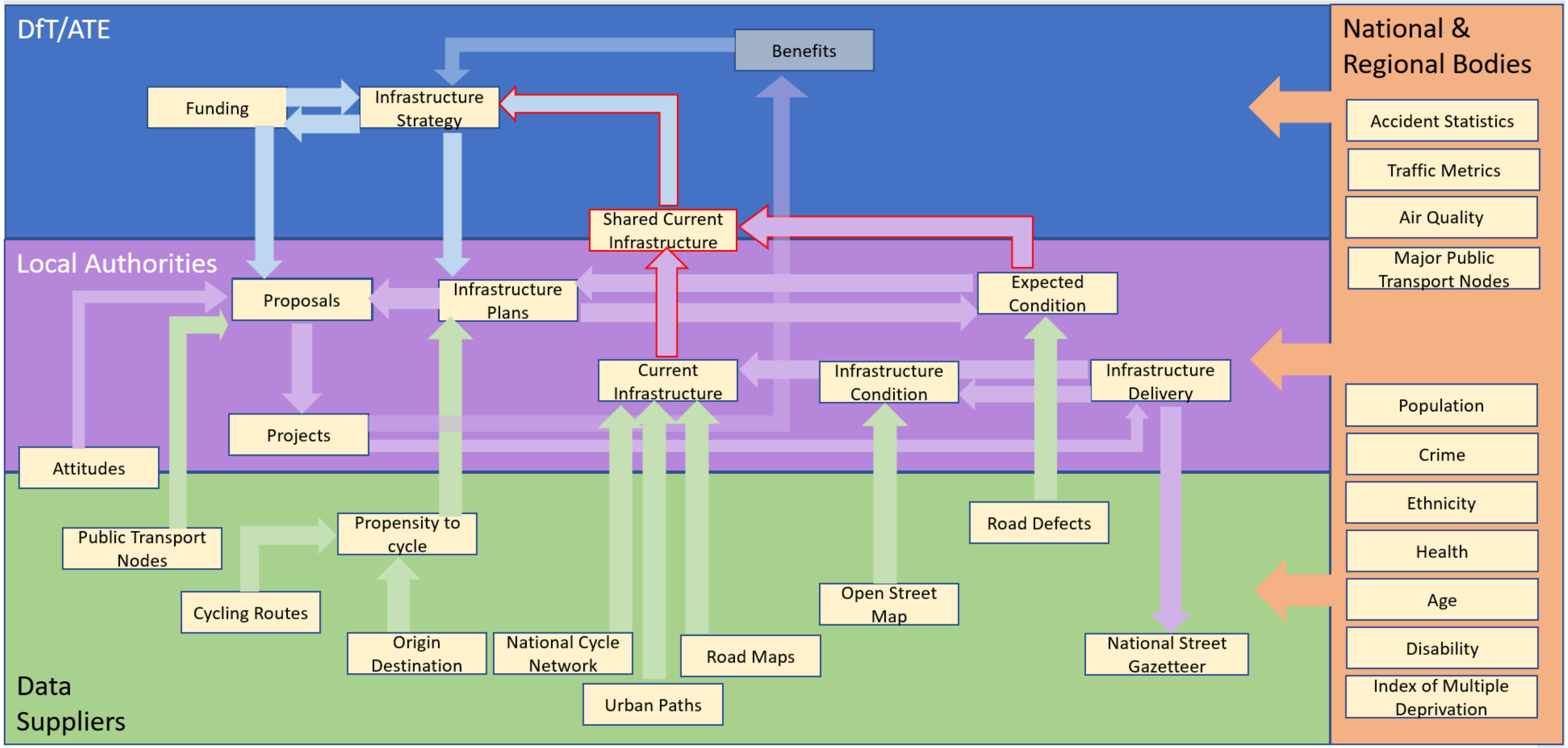

Figure 5- Key Data Flows - Target

The data flows diagram shows some small but significant changes from the equivalent base diagram (Figure 4 - Key Data Flows - Baseline) – these are all outlined in red. The most important is the addition of the shared repository of current active travel infrastructure, being updated from the equivalent systems of the local authorities. This is one of the possible configurations, which are covered in detail later (see 3.5.3.1).

This diagram also includes one of the possible additional features, shown as partially transparent – the benefits management loop, which would provide standardised upload, analysis and comparison of implemented schemes.

3.5.2.2 Basic IT requirements

This is covered in more detail in section 3.5.3, but it is useful at this point to identify the basic elements which are needed to support the single view of the data.

The key elements are:

- A shared GIS system and database – this is the repository for the national view of the infrastructure, with access to central DfT / ATE staff as well as other stakeholders like LAs and authorised partners, and for suitably released data, the general public and other data consumers (for example journey planner tools)
- Replication mechanism(s) – assuming that LAs will continue to maintain their data locally in their own GIS (if they have one), this facility would provide method(s) for updating the central tool from changes made in the local system should data be collected in this way
- Data entry mechanism – for LAs and others that need to contribute data, but who don't have a GIS system, method(s) of submitting the data in other forms. This could also be achieved by providing update access to the core shared system.

3.5.2.3 Infrastructure Data Replication and Synchronisation

At the moment, where the AT infrastructure exists in digital form, it is only present in 'official' form in local authority systems. Some LAs may have the data only in the form of non-GIS documents, or physical maps.

If the shared central database is implemented as per the diagram above (Figure 6 - Key Data Flows - Target), there will need to be a mechanism:

- to replicate changes from the local authority or other data provider systems to the shared repository.
- and for those LAs and other providers without a local application, a capability to enter the necessary information to describe their infrastructure.

This mechanism is dependent on the other options chosen; therefore, it has not been investigated in depth at this stage but will be one of the key questions to be investigated in Alpha (see section 4.2).

Considerations in this area include:

- Whether to automate the updates or include a review step before accepting updates
- Whether to immediately accession changes from the local environment or whether these should follow a schedule (the trade-off being stability of data to allow predictable results of querying, against immediacy of updates). There may be conflicting requirements for this[iv] – with local users or route planners using the data, wanting to have immediate updates but statistical analysts possibly requiring stability of data at certain points.

3.5.2.4 Infrastructure Data Completion

Ultimately, the target of the application is to achieve completeness of the dataset describing active travel infrastructure, in a shared repository. In order to achieve this, there are three main options:

- Bring all data up to the LTN1/20 level immediately before launching the application (the application would be available in a restricted way to allow the data to be loaded, but not widely available)
- Launch with whatever incomplete data is initially available, then gradually bring all the data up to standard
- Bring the available data up to a minimum standard (existence of infrastructure) with additional data from areas which have it, then launch the application followed by a gradual improvement in the data

The characteristics of the three options are summarised in the table below.

Task	Immediate LTN1/20	Launch with whatever is available	Launch with minimum standard, then enhance
Resourcing / effort to get to initial state	Very high	Low	Moderate

Task	Immediate LTN1/20	Launch with whatever is available	Launch with minimum standard, then enhance
Time to get to initial state	Very High (2-4 years)	Low	High (1-2 years)
Time to get to data-complete state from initial state	N/A (complete at previous step)	High (2-4 years)	High (2-4 years)
Quality/completeness of data initially	High	Low	Moderate
Quality/completeness of data eventually	High	High	High
Cost to get to initial state	Very high	Low	High
Incremental cost to get to end state	N/A	High	High
Resourcing required	Very high – probably 3rd party	Low	High
Business process change in local authorities	Low (if 3rd parties used), Very high otherwise	Moderate	High
Benefits	Gold-standard Data	Quickest launch	Fastest route to minimal view (existence of infrastructure)
Disadvantages	Very long lead time to launch High expenditure before any benefits	Data very incomplete until much later System may not be regarded as delivering worthwhile benefit	Data not enhanced until much later Availability of limited data may diminish enthusiasm to enhance data

In essence, the first option would involve an aggressive high cost/high speed survey of the whole infrastructure in as much detail as is required – either by the local authorities or, more likely, by 3rd parties. This approach is risky because 24% of LAs surveyed do not currently have plans to confirm LTN1/20 compliance, so support for this approach, particularly if the LAs are expected to do the research, is uncertain and would require careful business change management.

The second option would not have the immediate benefits of getting to the gold-standard quality of data but would get the IT infrastructure in place with whatever data was available and then, gradually bring the data up to standard – lower cost/faster and less disruptive, but would delay the availability of complete data, and the acceptance of the system may be questionable with incomplete data.

The third approach is something of a hybrid of the first two. Rather than going immediately for the ultimate set of data, the idea would be to survey the infrastructure in enough detail to get to the “where is the infrastructure” stage, and then progressively enhance the data to the required level. This is probably the route which gets to a usable state quickest, but the availability of that data set may impact the enthusiasm to then enhance the data to the full level.

Finally, it would also be possible to take a hybrid approach – for example, target some priority areas with the higher speed approach and bring the rest up to standard more slowly.

The flowchart below shows the high-level sequence of steps required to get the system up and running and then enhanced to its desired state.

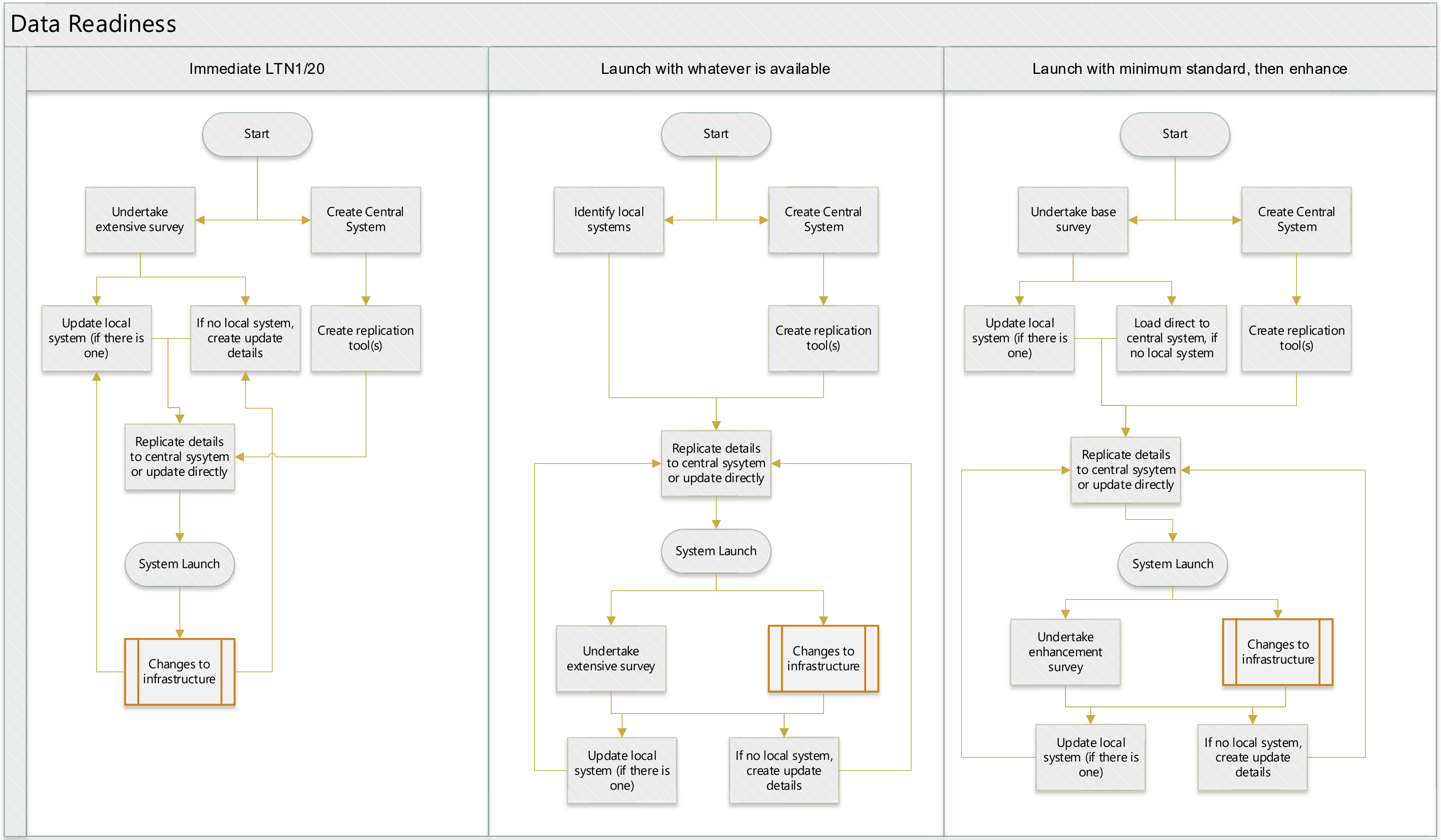

Figure 6 - Infrastructure Data Completion Flowcharts

3.5.2.5 Achieving Data Completion

Regardless of which strategy is taken, to get to a comprehensive picture some approach(es) will need to be undertaken to fill in the gaps of the existing LA data. These options would then also be used to maintain and update data regularly. The following lists the possibilities and identifies the strengths and weaknesses of each.

To use any of these options, creating a technology platform to receive the data is a pre-requisite. In most cases, this needs the creation of the shared central infrastructure system to be in place, though for some LAs where they already have a GIS system, the data could be updated in their local system first and then replicated to the shared system or alternately the data could be created in some other format which can be consumed by the central system.

The options are:

- Adopt the data present in OpenStreetMap, an open-data/crowd-sourced mapping tool which already holds substantial amounts of information about cycling infrastructure
- Commission a 3rd party to survey the infrastructure. Ordnance Survey are an example, but other vendors have capability and experience in this space also
- Ask Local Authorities to perform the survey
- Use computer processing of other information (maps, images, traffic levels, etc) to infer the location of elements of infrastructure. This might require multiple contributing groups with expertise in different aspects. Note that the 3rd party survey may also be using elements of computer processing to undertake their survey.
- To require increased reporting of infrastructure during any projects which touch on AT infrastructure – e.g., road changes, construction of cycle ways

Task	Open Street Map	3rd party survey	LA survey	Algorithmic methods	Improved reporting during developments
Initial available coverage	High in well-used areas, Moderate in sparsely used	Very good for NCN, variable in other areas depending on local initiatives	Very varied – some high, some non-existent	Dependant on map/image source	N/A
Effort to get to full coverage	Moderate – very good coverage in highly-used areas, concentration needed in other areas	Cost dependant – can scale out depending on spend.	Very varied – for most, probably High	Variable, depending on attributes	Low, by-product of works

Task	Open Street Map	3rd party survey	LA survey	Algorithmic methods	Improved reporting during developments
Time to get to full coverage	Moderate, if incentives provided, otherwise very unpredictable – Low in well used areas, High in others	Potentially Moderate-High, but cost dependant – can scale up resourcing for faster surveying depending on spend.	High to Very High (probably 2+ years), and very dependent on available resourcing	Moderate Very dependent on how susceptible to algorithmic processing the attribute is	Very high – dependant on projects
Cost to get to full coverage	Low	Very high	High	Very dependent on how susceptible to algorithmic processing the attribute is	Low
Licensing	N/A	Dependent on commercial agreement. Probably OK for DfT use, may be an issue for public use	N/A	TBC – Probably not an issue[v]	N/A
Support of service standards	Very good	Dependent on commercial agreement	Good	Dependent on the specific groups involved	Good
Ease of correction	Good	Good	Moderate	Highly dependent on cause of issue	Moderate
Breadth of data	Very good	Good – Very Good	Very Good	Moderate - Good Very dependent on how susceptible to algorithmic processing the attribute is	Moderate – Good, dependant on skill of staff reporting
Supportability	Very good	Very good	Good	Highly dependent on cause of issue	Moderate - Good
Documentation	Good	Very good	Good	Good – Very Good	Moderate – Good
Comments	Reluctance from some LAs to trust this data as it is not "official" Conversely, data is, to some extent, self-correcting as errors are recognised and fixed by public consensus"	Survey of quality of infrastructure may need pre-survey of location of infrastructure	Very dependent on available resourcing within LAs, and existing GIS knowledge and commitment to AT	Potentially, very effective, but may not be able to provide full solution for all attributes	Depends on the knowledge/skills of the persons reporting the details Least disruptive to existing processes or staff, but much slower to get to complete results.

Task	Open Street Map	3rd party survey	LA survey	Algorithmic methods	Improved reporting during developments
	Quality and certainty of sparse areas could be rapidly improved by incentivising visitors to those areas either through paying directly or competitions, prizes etc.				

These options are not necessarily mutually exclusive – for example, the data in OpenStreetMap (OSM) might be used in conjunction with data from specific LA or 3rd Party surveys – this might be particularly appropriate for areas on OSM which are not heavily used, as the quality of data in OSM is dependent on the size of the crowd supplying each data item.

3.5.3 Application Architecture

The application architecture describes the software to achieve the desired capabilities. It also includes some aspects of data architecture, where that architecture is driven by the software option.

There are a number of different ways to approach this solution, each with a different mix of strengths and weaknesses. These are all set out in the tables below.

These options only include delivery of the MVP, to understand what infrastructure exists – they do not include further capabilities for benefit management or cost reviewing. These should be assumed to be additional cost, bespoke development, regardless of which option is chosen, but the cost and effort to deliver them will not be significantly affected by the MVP option(s).

3.5.3.1 Technology Options

3.5.3.2 There are 5 technology options to be considered, which are briefly described below

- Commercial GIS, using ArcGIS Online from ESRI. This is a very well-known and widely used GIS, with existing implementations in both DfT and many LAs (25% of those surveyed, 100% of those interviewed)
- Best of breed open-source, using GeoNode GIS and Postgres database. Again, very widely used tools with a good array of support from the community of users and developers.
- The next option is to utilise some of the tools which have been funded by DfT, for example the Propensity to Cycle Tool[vi], Cycling Infrastructure Prioritisation Toolkit[vii], Active Travel Routes in Development[viii], and implement an API to allow data to be shared amongst these[ix]. All of these tools use similar data and consolidating the data provision would allow the existing investment to be leveraged, while enabling a more robust updating mechanism, to keep data current across all of them. In this option, either one of the existing GIS platforms would be extended or a new tool (probably GeoNode) would be implemented.
- The fourth option is more radical – this would involve creating a data platform to be shared between a set of tools (see examples in previous bullet point) which could depend on that dataset for current information. This is similar to the 2nd (open-source) option but would also involve some refactoring of the existing tools to use the new shared platform. As in the previous option, either one of the existing GIS platforms would be extended or a new tool (probably GeoNode) would be implemented.

  The comparison table, below, assumes that this work would all be done as part of the same implementation project, it could also be done as a phased implementation, with the provision of the shared platform in the first phase and the migration of the other tools in subsequent phase(s).
- Finally, a variant of the first option but in this case the single repository doesn't supplement the GIS systems in the LAs, it replaces them with a single shared instance. There are some advantages to this, in that it would remove the need for maintenance and management of all the local systems and would not need replication or synchronisation to the shared repository. However, it would be very disruptive to implement and is included here mostly for completeness – this would be a more attractive solution if starting from a clean sheet of paper.

3.5.3.3 Technology Option Comparison

	New Commercial GIS (ArcGIS Online)	New Open-source GeoNode)	Expand on Existing Tools	Integrate and re-platform existing tools	Replace LA systems with single system (ArcGIS online)
Initial Cost (licensing)	Moderate-High[x]	Low / None	Moderate[xi]	Moderate	High
Ongoing Cost (licensing)	Moderate – High, depending on the number of users	Low / None[xii]	Low / None	Low / None	Moderate – High, depending on the number of users
Control over data location	No	Yes	TBC – may require multiple[xiii]	Yes	No
Control over hosting	No	Yes	TBC – may require multiple[xiii]	Yes	No
Availability of support / development staff	Very good	Good	Good	Good	Very good
Degree of custom development	Low	Low - Moderate[xiv]	Moderate-Fairly High[xv]	Moderate – Fairly High[xvi]	Low
Initial Cost (implementation)	Low	Low - Moderate	Moderate	Moderate – Fairly High	Very High Replacement of all LA GIS systems
Maintenance cost (resource cost)	Low-Moderate	Low-Moderate	Low-Moderate	Low-Moderate	Low[xvii]
Complexity of Replication of LA data	Low	Low – Moderate	Low – Moderate	Low – Moderate	N/A
Currency of data	Very good	Very good	Very good	Very good[xviii]	Very good
Leverage existing investment	No	No	Good	Very good	No
Alignment with architecture guidelines	Moderate De facto standard, but not aligned to open-source	Good	Good	Good	Moderate - Good De facto standard, but not aligned to open-source

	New Commercial GIS (ArcGIS Online)	New Open-source GeoNode)	Expand on Existing Tools	Integrate and re-platform existing tools	Replace LA systems with single system (ArcGIS online)
					Reduces number of discrete local instances and of different GIS systems to single shared instance.
Alignment with Service Standards	Low (not open-source)	Good	Good	Good	Low (not open-source)
Complexity of implementation	Low	Moderate	Moderate – High	Fairly High	Very High Involves the replacement of all local GIS systems and data migration from all of them
Degree of business change on LAs	Low	Low	Low	Low	High
Impact on other parties	Low	Low	Moderate – High May require changes in existing tools l ke PCT and other CycleStreets apps	Fairly High Likely to required changes in the other tools, but conversely also likely to make the data more current in them	Low
Key Risks	Change of commercial offering	Abandonment of key tool(s) by developers	Unexpected Complexity of integration	Complexity of integration and migration	Significant Business change Significant & complex migration
Key Benefits	De Facto industry standard, and in use in DfT/LAs Software as a Service platform so no responsibility for hosting	Low cost/high quality Adherence to Architecture & Service standards	Improvement in currency and maintenance of all tools – builds in update mechanism for all tools Avoids an element of reinventing the wheel where parts of the solution already exist Might reduce the amount of new development required	Improvement in currency and maintenance of all tools Avoids an element of reinventing the wheel where parts of the solution already exist Potential for new capabilities to be developed easily from shared data pool Might reduce the amount of new development required	Single shared platform Provides full GIS capability to all LAs, including those currently without that capability. Software as a Service platform so no responsibility for hosting

Figure 3 - Technology Option Comparison

3.5.3.4 Application Option Costing

The costing of the data collection will be heavily influenced by the method(s) used, the data quality standards required, the speed by which the data is required to be collected and commercial conditions related to ownership of the data – see section **Error! Reference source not found.** for more details of this. Whichever method is chosen, this is independent of the application option, so does not affect the relative costing, so has been added separately.

3.5.3.5 Application Cost

Based on the standard Alpha-Beta-Live model, we would expect the Rough Order of Magnitude cost of the project to break down as shown in the table below.

	New Commercial GIS (ArcGIS Online)	New Open-source (GeoNode)	Expand on Existing Tools	Integrate and re-platform existing tools	Replace LA systems with single system (ArcGIS Online)
Alpha Phase	Low	Low	Low	Low	Low
Beta Phase[xix]	Low	Low	Low	Low	Low
Support, Maintenance & Enhancement[xx]	Medium	Medium	Medium	Medium	Medium
Total Resource Cost (A)	Medium - high	Medium - high	Medium	Medium	High
Initial License Cost	Low	Very low	Very low	Very low	Low
Ongoing Maintenance Cost	Included	Low	Very low	Very low	Included
Hosting Cost[xxi]	Included	Low	Low	Low	Included
Total Platform Cost (B) over duration of Alpha, Beta & Support	Low	Low	Low	Low	Low
TOTAL A+B, rounded[xxii]	Medium - high	Medium - high	Medium	Medium	High

3.5.3.6 Data Acquisition Cost

The data acquisition cost is extremely challenging to quantify, as it is heavily dependent on factors like the speed required, how many resources can be allocated, whether tasks can be done in parallel, commercial considerations around the ultimate ownership of the intellectual property identified. However, in order to get some basis for this, the following have been used as benchmarks:

- Sustrans surveyed the whole of the National Cycle Network to a high level of detail in approximately two years.
- The cost of the Sustrans work has been shared and other commercial estimates allow a very rough cost approximation to be assigned.
- The full AT network is believed to be approximately 8 times the size of the NCN.
- That the time to survey would be less than 8 times the Sustrans survey, as more resource could be funded to work in parallel.
- The assumption is that surveying each section would be done once, rather than requiring multiple stages of surveying. There may need to be additional set-up for some suppliers to support this – for example where analysis is done by processing static images, the processing code may need to be upgraded to support all the required attributes.

The table below gives some estimates based on the assumptions above. It must be emphasised that these are only a general order of magnitude – getting more accurate estimates would require further work in Alpha..

With those caveats stated, the table breaks down the data acquisition into two stages:

- Getting to the point of knowing where the infrastructure exists at all
- Enhancing that data with all the LTN1/20 attributes

It has been split out that way particularly because the 3rd party and Algorithmic methods are both dependent, to varying degrees, on knowing the infrastructure exists in the first place.

Task	Open Steet Map	3rd party survey (eg OS or other)	LA survey	Algorithmic methods	Improved reporting during developments
Initial consolidated view (ie existence of infrastructure)	Already there.	Medium duration Medium cost	Heavily dependent on resourcing. Local knowledge should make this relatively 'easy' and possible to do in all LAs in parallel, but resource constraints counter this. Short duration Medium cost	Not really suitable for this initial discovery, as the methods rely on minimum initial information.	Very slow – depends on works being done in the relevant areas to 'discover' AT infrastructure. Very low cost Very long duration
Enhance data to full LTN1/20 level	Challenge is to get enough visitors to less popular areas to take the data to critical mass to ensure quality. This could be achieved by funding visitors, either directly or through challenges, competitions etc. With enough incentive, this could be completed	Very much dependent on how fast the data is to be achieved. Medium duration High cost	Very resource heavy to do this 'manually' by visiting and measuring, although potentially all LAs in parallel again. Medium duration Medium to high cost	Heavily dependent on usable source data (eg photographs, aerial photos, maps etc). However, work already exists which demonstrates some of this capability. Assume funding additional academic work by grants, that algorithmic analysis is	Probably included as part of above.

Task	Open Steet Map	3rd party survey (eg OS or other)	LA survey	Algorithmic methods	Improved reporting during developments
	very quickly due to the number of participants which could be enlisted. Short to medium duration Low cost			possible. Medium duration Low to medium cost	

Reviewing the options, it is clear that the fastest and cheapest way to get to at least a base level of confidence in the infrastructure is to use, and potentially incentivise participation in, OpenStreetMap, which already has large portions of the required data. However, if this data were to be adopted, concerns in some Local Authorities towards 'unofficial' data would have to be addressed.

Beyond that point, other methods could be used to enhance the information to the required level and/or target areas where OSM participation is weaker (i.e. less visited areas and routes). This would have the advantages of

- Targeting resource where it is most effective, without reinventing the wheel, and thereby constraining cost.
- Depending on commercial considerations, potentially improving OSM for community benefit, consistent with service standards and open-source ethos.

3.5.4 Key technical challenges

For most of the solution options, the technical challenges are relatively small and quite similar between options. They are enumerated separately here to address the specific question on this area in the DOS brief.

The key challenges for these break down into the following areas, which are then scored in the table below:

- Knowledge required of existing systems – the degree to which the existing GIS systems (if any) need to be understood in order to implement the option
- Integration and replication – the degree of complexity in implementing replication / synchronisation from LA systems, once the platform is in place
- Hosting Complexity
- Data Migration – the relative degree of migration of data from existing systems to the new platform, *not including* the integration and replication aspect which is covered separately
- Transition from current process – how complex the technical changes are to move from current state to future state; this reflects how much the local systems must change and how different the technology that the local users will need to use

	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared ArcGIS Online
Knowledge required of existing systems	Low	Low	Moderate	Moderate	Very High
Integration and replication	Moderate	Moderate	Moderate	Moderate	Low[xxiii]

Hosting Complexity	Low	Low	Moderate	Moderate	Low
Data Migration	Low	Low	Low	Moderate	Very High
Transition from current process	Low	Low	Low - Moderate	Moderate	Very High

3.5.5 Barriers to a solution

There are several factors which need to be taken into consideration when assessing the potential technology platform, and particularly the data collection, as they are potential barriers to successful implementation of the proposed solution.

3.5.5.1 Business / process change

Firstly, the overall solution is not solely an IT solution, there will also need to be degree of business change. The impact of this will vary depending on the options chosen, which is highlighted in the options table under the "degree of business change" row.

The way in which data completion is implemented will also have significant impact on the degree of business change required – the more emphasis that is given to acquiring data through the local authorities, and the more changes to the local IT environment (by requiring process changes to support data provision to the centre), the larger the business change effort that will be needed.

3.5.5.2 Minimise load on Local Authorities

An important consideration is to avoid creating onerous or expensive processes which Local Authorities will need to undertake.

This is important for two reasons:

- firstly, that some local authorities may have financial and staffing constraints and might not be able to take on additional work without impact to other responsibilities
- secondly, that any system which significantly increases workload without a commensurate benefit, is unlikely to gain support for implementation.

3.5.5.3 Demonstrable benefit

The mitigation for additional load, or other business change, is to demonstrate the benefits accrued from the delivery of the new platform. This is particularly important if additional data collection responsibilities are pushed into the existing data providers – they must be able to see corresponding benefits derived from the data collection if they are to buy-in to the revised process.

3.5.5.4 Cost and time of data completion

There are potentially significant costs and an extended duration in getting the data content to the point where the desired benefits of the service can be manifested. See section 3.5.2.4 for more detail on this point.

3.6 Technology Architecture

The technology architecture describes how the proposed solution options fit to architecture standards, physical constraints and any other architectural requirements.

3.6.1 Architectural standards

The potential solution needs to conform to the DfT Architectural Standards. These are wide-ranging and all will be considered in the eventual design, but the most key elements to support decisions now, are excerpted below. The sub-headings represent sections within the Architectural Standards document.

In general, each item is graded by conformance with the relevant architectural standard. 'Yes' is shown where these standards are met by the option, and 'No' where the relevant standards would not be met fully.

3.6.1.1 Guiding principles

Item	Description	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared
Build digitally integrated services	Build products and services that are flexible, reusable and integrate with new and existing services.	Yes	Yes	Yes	Yes	Yes
Data is an asset, shareable and accessible	Data is an asset that has value to the Department and should be managed accordingly. DfT staff must have access to the data necessary to perform their duties. Data should be shared across DfT family.	Yes	Yes	Yes	Yes	Yes
Maintain data confidentially, integrity and availability	Data will be maintained and managed with confidentiality, Integrity and availability risks in mind.	Yes	Yes	Yes	Yes	Yes
Use cloud first	Utilise cloud services to maximise the benefits of scalability and flexibility with 'as-a-service' technology Solutions.	Yes	Yes	Yes	Yes	Yes
Invest in enterprise-wide solutions	Investing in Commercial-off-the-shelf (COTS) enterprise solutions before developing our own solution can reduce complexity and cost.	Yes	No	No	No	Yes
Re-use before investment	Re-use existing products or services before investing in new solutions, if existing services meet the user need.	No	No	Yes	Yes	Yes

3.6.1.2 Principles and Standards

Item	Description	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared
Use public cloud first	Cloud Service providers are constantly innovating at a pace greater than could ever be achieved with 'on premise' delivery platforms.	Yes	Yes	Yes	Yes	Yes
Use appropriate platform	• Digital Services developed internally (within central DfT) must be hosted on Google Cloud Platform (GCP). • Externally developed and managed digital services, that require DfT staff logging on must be hosted on GCP. • Externally developed and managed digital services that are truly "hands off", (not managed, supported or accessed) for DfT should be hosted on GCP.- a. If this is not poss ble Microsoft Azure or Amazon Web Services are appropriate, subject to Architecture approval.	No This is a Software as a Service platform with no user control over deployment	No	Yes	Yes	No This is a Software as a Service platform with no user control over deployment

3.6.1.3 Applications

Item	Description	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared
Reuse existing application investments	Where an application has already been purchased and is being utilised, efforts must be made to utilise those existing applications investments.	Partly	Partly	Yes	Yes	Partly – consolidates existing ArcGIS licenses into one instance
Purchase rather than build applications	Commodity applications are often cost effective, offer greater supportability and upgradability.	Partly – Commercial-off-the-shelf and some bespoke	No	No	No	• Partly – Commercial-off-the-shelf and some bespoke
Build bespoke application when no commodity application is available.	Where no existing applications or commodity applications fulfil the business need then bespoke development of an application should be considered.	Bespoke minimised	Bespoke minimised	Bespoke development possible	Bespoke development required	Bespoke minimised

3.6.1.4 Development

Item	Description	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared
Build cloud native	To support the DfT Digital Technology Strategy, new services must be cloud ready.	Yes	Yes	Yes	Yes	Yes
Use open-source technologies[xxiv]	The use of open-source technologies prevents vendor lock-in.	No	Yes	Yes	Yes	No
Build services that work together	Integrated services are efficient, cost effective and promote mastering of data.	Yes	Yes	Yes	Yes	Yes

3.6.1.5 Data

Item	Description	Commercial / ArgGIS	Open-source	Leverage Tools	Integrate and re-platform tools	Replace LA systems with single shared
Collect data accurately	Regardless of the collection method; the quality, accuracy and the integrity of data should be maintained.	Yes	Yes	Yes	Yes	Yes
Store data suitably	Data should be stored and managed centrally. Business data should be stored within Platform (GCP).	No If enforced, this would mean replicating data to a GCP store	Yes	Yes	Yes	No If enforced, this would mean replicating data to a GCP store
Integrate data appropriately	• Design solutions that provide access to data via API's using standard consumable formats and transmission standards. • Data should be Open by default (subject to data security). • Where DfT data is hosted externally or being processed externally, a repository of the data must remain within DfT cloud environment	No If enforced, this would mean replicating data to a Google Cloud Platform store	Yes	Yes	Yes	No If enforced, this would mean replicating data to a Google Cloud Platform store

3.6.2 Outline Architecture

The definitive deployment architecture, will be elaborated during the Alpha and Beta phases, but the solution options can be described at a high-level

3.6.2.1 Commercial/ArcGIS or Open-source

The component relationships are fairly simple, regardless of which of these two options are chosen – a GIS tool, a database, an ETL tool and a series of local systems communicating with the shared platform

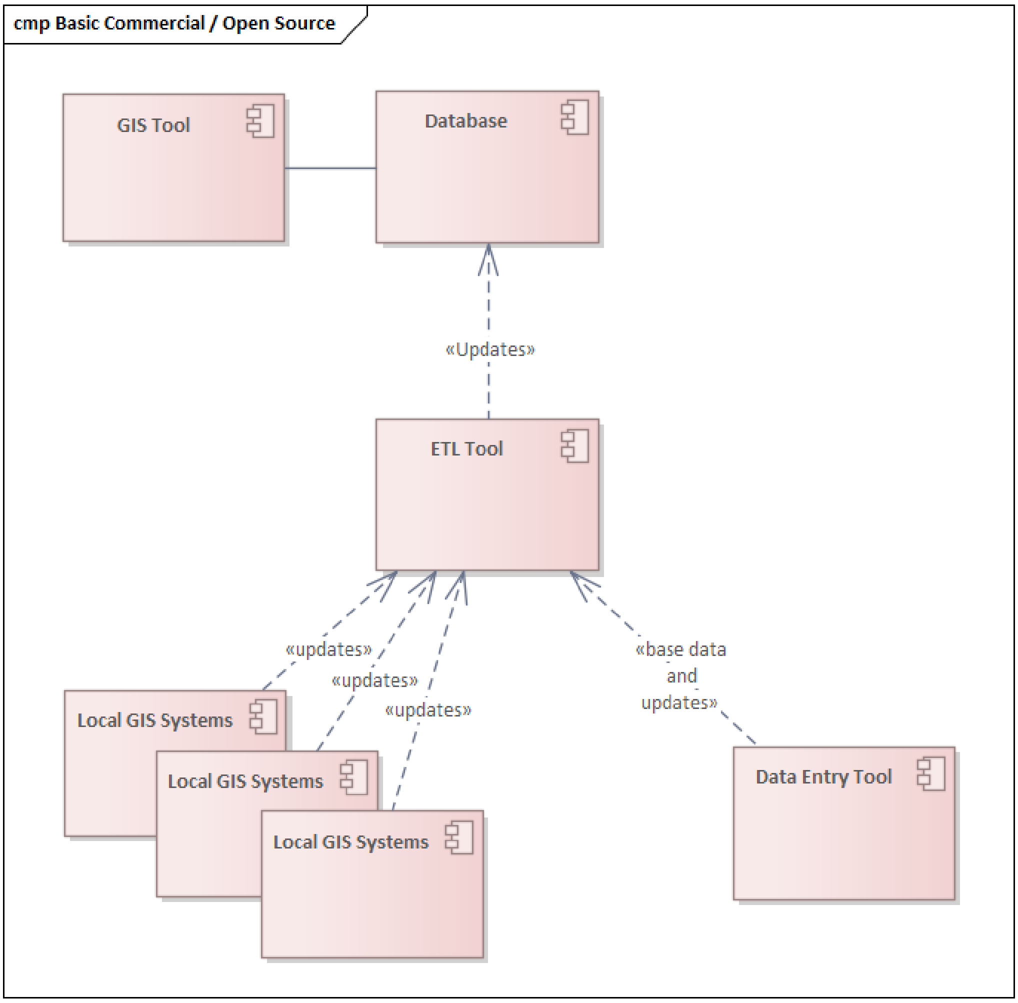

Figure 7 - Deployment pattern - Commercial or Open-source

3.6.2.2 Leverage Existing Tools

The precise make-up of this option will depend on further examination of the various tools already deployed, PCT etc are included as a well-known examples but are not necessarily ones which could be integrated like this.

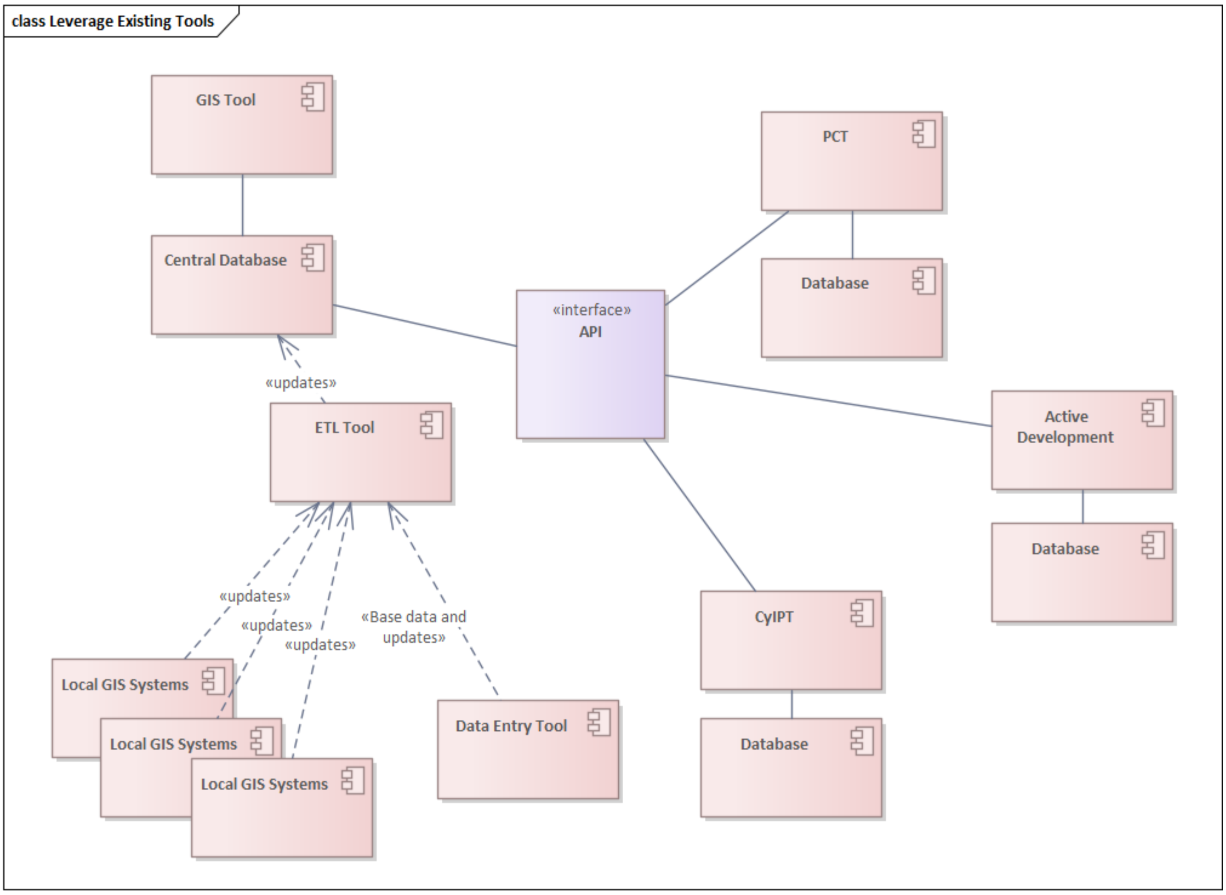

Figure 8 - Deployment pattern - Leverage existing tools

In this environment, any changes from local systems, and potentially from the other tools, would flow through the ETL tool to update all of the databases, most likely through interfaces/APIs to the existing tools.

3.6.2.3 Integrate and re-platform tools

As in the previous example, PCT et al are included as a well-known examples, but are not necessarily ones which could be integrated like this.

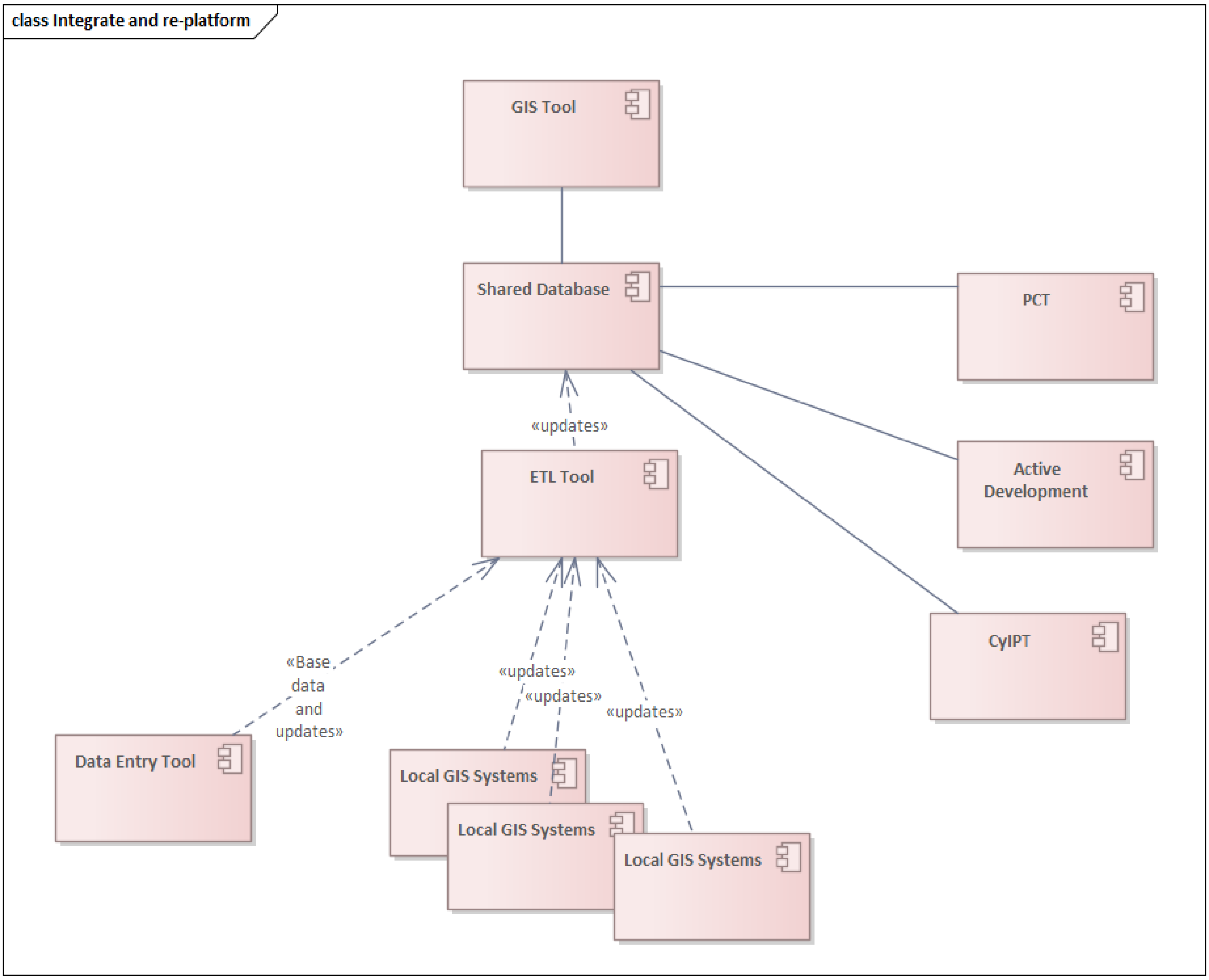

Figure 9 - Deployment pattern - integrate & re-platform

In this architecture, the various pre-existing tools would be updated to use a shared database, and all their existing data (where it was not duplicating data already held in the core database) would be migrated into the core.

Updates to the core database would be available to all applications at the same time.

3.6.2.4 Replace LA systems with single ArcGIS

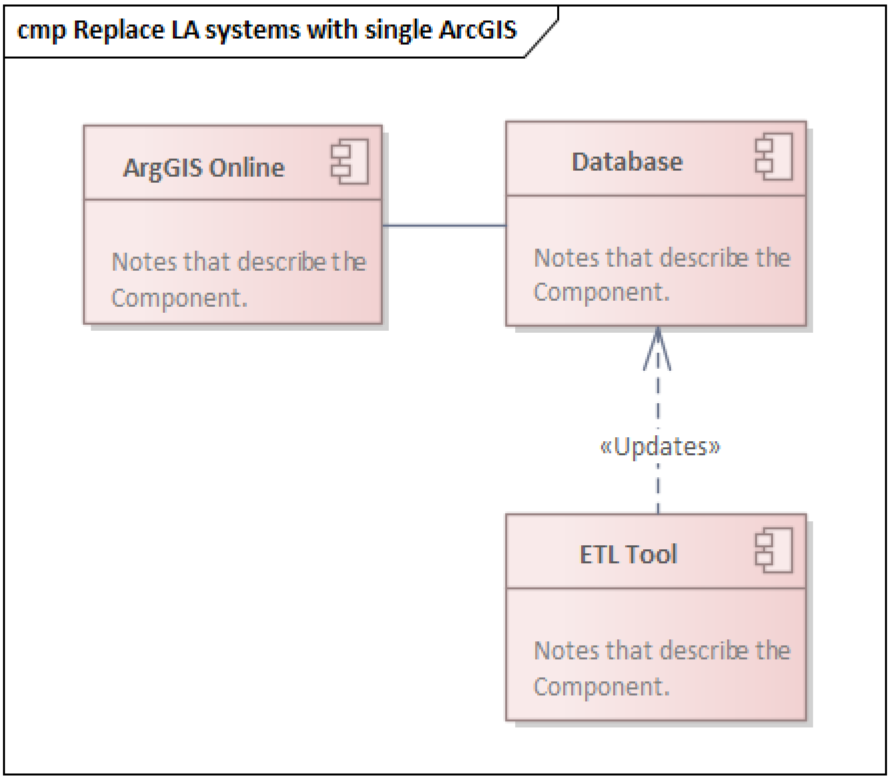

Figure 10 - Deployment pattern - Replace with single ArcGIS

This one is deceptively simple – the ETL tool remains to update the database with changes coming in from any other systems, and there is a single GIS & Database instance. The complexity is all about the journey to get to this configuration.

The ETL tool is still required for processing any map and data updates from 3rd party suppliers

3.7 Opportunities and Solutions

The following tables show the functional and non-functional requirements, which were identified as part of the user journeys and indicate which group(s) clearly expressed an interest in this element.

Columns without ticks do not necessarily imply that the group were not interested in a particular aspect, but they have not explicitly expressed a preference for it in the interviews.

These are divided into functional requirements (what the users want the system to do) and non-functional (things the system must do to be usable and supportable). These lists are not exhaustive lists of the requirements but represent those which have so far been elicited from user interviews.

3.7.1 Functional Requirements

Expressed Opportunity	LAs	DFT AT Policy	DfT Statisticians	Dft Analysts
1. Data to support bid management & review	Yes	Yes		
2. Condition of surface		Yes		
3. Understand maintenance needs and process	Yes	Yes		
4. Build sustainable workflow in LAs to collect & maintain data	Yes	Yes		
5. Identify barriers to access	Yes	Yes		
6. Share data consistently	Yes	Yes	Yes	Yes
7. Standardise reporting process to allow comparison between projects		Yes	Yes	Yes
8. Different levels of access to maps layers depending on how socialised they have been (mainly around aspirational plans)	Yes	Yes		
9. Map the use of different funding sources		Yes	Yes	Yes
10. Defined update cycle[xxv]		Yes	Yes	
11. Updates immediately reflected in maps, including time-based road closures / road works etc	Yes			
12. Access to safety information / making cyclists feel safer	Yes	Yes		
13. Ability to correct / validate data / mark data as validated	Yes	Yes	Yes	Yes
14. Ability to contact data providers to query / correct data	Yes		Yes	Yes
15. Segmentation of data (particularly benefits) for various metrics (IMD, health etc)		Yes	Yes	Yes

Expressed Opportunity	LAs	DFT AT Policy	DfT Statisticians	Dft Analysts
16. Represent topography of routes, especially gradients	Yes			
17. Capture local knowledge to remove dependency of staff members	Yes			
18. Network view of surrounding areas, including public transport nodes, to understand local links	Yes	Yes	Yes	
19. Querying data at various geographic divisions (county, postcode, etc)	Yes	Yes	Yes	Yes
20. Align data sets to specific policies	Yes			
21. Access to road widths	Yes		Yes	Yes
22. Model changes for public consultations	Yes			
23. Identify whether routes are segregated / allow journey planner to avoid non-segregated routes	Yes			Yes
24. Count cycling / walking users segmented by demographics		Yes	Yes	Yes
25. Traffic volume, speed, mix of vehicle types	Yes		Yes	
26. Get LAs started and thinking about their LA journey, even if the path quality is poor		Yes		
27. Convince local politicians of the necessity of AT		Yes		
28. Identify how funding makes better places to live (behavioural change data + comparison between LAs)		Yes		
29. Agreed on definition of cycling and walking network / Agree what needs to be mapped / Compare quality of infrastructure	Yes			Yes
30. Promote AT equally in urban and rural areas		Yes		
31. Interactive maps with selectable overlays (for internal and public use)	Yes		Yes	
32. Develop clear guidance, agreement and consultation on Rights of Way and Bridleways	Yes			
33. Label safety and continuity of paths, which are not LTN120 compliant		Yes		
34. Identify impact of bollard and neighbourhood calming measures		Yes		

Expressed Opportunity	LAs	DFT AT Policy	DfT Statisticians	Dft Analysts
35. Identify and map all infrastructure aspects of AT, even if they are not directly measurable		Yes		
36. Support for exploratory analysis			Yes	
37. Design inclusive paths	Yes			

3.7.2 Non-Functional Requirements

	LAs	Active Travel Policy	DfT Statisticians	Dft Analysts
38. Date to be exportable to Excel and other formats	Yes	Yes	Yes	Yes
39. See data sets instantly without formatting	Yes			
40. Data queries must be quick			Yes	
41. Functionality must be grouped in logical ways for different user groups	Yes		Yes	

3.7.3 Potential Features

This table lists some opportunities for potential features for the proposed solution, relates them to the numbered potential requirements above, and indicates whether they are:

- aligned to the MVP, to address the key target of identifying what infrastructure exists
- a potential enhancement to come later
- probably beyond the scope of an application

Where a potential feature is ticked in multiple columns, it indicates a partial implementation at one stage with further development to follow

Potential Feature	Aligned Opportunities	MVP	Enhancement	Out of scope?
Bid & benefits upload and analysis	1		Yes	
Cycleway metrics – surface, segregation etc	2	Yes		
Identification of barriers	5		Yes	
Expected maintenance requirements / cost	3		Yes	
Data collection and maintenance	4, 17	Yes		

Potential Feature	Aligned Opportunities	MVP	Enhancement	Out of scope?
Shared access to data / access control by role	6, 8	Yes		
Standardised reporting	7		Yes	
Funding source analysis	9		Yes	
Control of update timing / frozen views of point-in-time state / continual updating	10, 11	Yes		
Traffic & other safety information	12, 25, 34		Yes	
Segmentation data (IMD, health, age etc)	15, 24		Yes	
Mapping of funding sources to schemes	20, 27, 28		Yes	
Contact data supplier & supply correction	14	Yes	Yes	
Topography maps / route gradients	16	Yes	Yes	
Road widths	21		Yes	
Visibility of network extent and gaps	29, 32, 36	Yes		
Visibility of benefits / behaviour change	22, 26, 27, 35		Yes	
Selectable map layers	31	Yes		
Guidance on use of certain land features	33		Supporting information	Yes
Comparison of infrastructure	29	Yes		
Data analysis by multiple dimensions	15, 19, 24, 37			

4 Alpha Considerations

The function of the Alpha phase is to de-risk the project development by investigating and trialling options, doing more detailed analysis into targeted areas than is possible in the Discovery phase and to refine the project costing from Rough Order of Magnitude to a more detailed level.

4.1 GIS proof of concept

This activity would be to create a simple proof of concept using ‘manually’ loaded subsets of data. This would show that the aims of the development were possible to achieve from the tools and data which are expected to be available.

This is not about the creation of the final GIS implementation, it is intended to test the concept of implementing specific datasets, transferring and using the data. This is an example of the ‘fail fast’ approach of Agile – to determine quickly if there are clear blockers to the success criteria, before committing to the more extensive Beta phase.

4.2 Replication / Synchronisation mechanism

If the chosen system options include a requirement to replicate local authority data, to a shared repository, the precise mechanism for achieving this will need to be chosen and tested in alpha. There is already some ‘prior art’ in this area which may be able to be leveraged –local authorities already provide updates both to Sustrans (for National Cycle Network) and Geoplace (for National Street Gazetteer).

4.3 Trial and agree support tooling

As well as the core technology of the GIS tool, the cloud environment will include a number of options for database, authentication, CI/CD[xxvi] pipeline, data loading etc. Some of these will be able to be defined as best options, others will need to be trialled to ensure suitability or the best way of implementing them.

Ideally, all of these issues will be defined and implemented by the end of Alpha, to provide a firm basis for the developments in Beta, though some of them may need additional development in Beta.

4.4 Detailed review of technology of existing tools to see if they can be reused

Two of the options for delivering the desired functionality are to leverage the existing investment in active travel tools. On a superficial level, it is clear that a number of these tools are using similar data sets and could potentially be combined or augmented to deliver the required functionality.

However, there needs to be a detailed review of the technology used – programming language(s), database(s), GIS tool(s) – to see what common ground exists and whether these could be brought onto a common platform with the proposed Active Travel infrastructure tool, and whether it makes sense to do so.

4.5 Explore viability non-survey ways of establishing quality of infrastructure

There have been a number of comments during the discovery around ways of potentially deriving (primarily) qualitative measures of infrastructure by methods other than physically visiting and surveying – these include algorithmic processing of maps or photographs, and other methods of inferring the location and/or characteristics of the infrastructure, such as traffic analysis.

The review in alpha would be to examine all these available approaches and determine which one(s) of them provide useful and reliable data which could be incorporated into the data gathering

4.6 Detailed review of data collection options and selection

Similarly, there are a number of options for physically reviewing the infrastructure, whether by using local authorities, 3rd party surveyors or cloud sourcing the data.

There needs to be a definitive position on which one(s) of these approaches are viable to be used on cost and quality grounds.

4.7 Scope Data Migration Requirements

The last two solution options include data migration from their existing platforms. If one of these options is selected, the data migration requirements and scope will need to be determined in Alpha. For the last option, of replacing all the local authority GIS systems with a single ArcGIS instance, the migration is likely to be a very significant factor in the project duration.

4.8 Deepen Understanding of stakeholder processes identified in Discovery

In discovery we have identified the main processes each stakeholder works with. The user research in Alpha needs to revisit these processes with the stakeholders and confirm those. It is important to do this at the start so the user journeys correctly can be built on correctly. At present the user journeys and the service blueprint are a high-level overview. More details, if there are any, should be added in Alpha.

4.9 Develop low-fidelity prototypes

Following the deeper understanding of user journeys and the development of the service blueprint, first low fidelity prototypes should be developed and tested and discussed with the stakeholders. The prototype will be based on the knowledge learnt from the user journeys and service blueprint.

4.10 Iterate low-fidelity prototypes and define and adjust user needs

The low-fidelity prototypes will be developed and iterated through constant conversation with the stakeholders. This process will develop a deeper understanding and empathy of the user needs.

4.1 Initial Backlog of requirements

As part of Discovery, a set of needs and user journeys have been identified. During Alpha, these will need to be expanded into full agreed backlog to define the initial scope of what needs to be built in Beta. As with any iterative agile process, this will evolve during development, but it is important to have the initial backlog in place to have an agreed baseline scope of the work to be undertaken.

5 Meeting the Service Standard

5.1 Understand users and their needs

Understanding the users and their needs has been a key element of the discovery process. The discovery team have engaged with as many varied stakeholders as possible, both by interviews, blog posts and by running a survey.

The range of stakeholders interviewed included central DfT policy and analysts, several local or combined authorities and other involved parties.

5.2 Solve a whole problem for users

The whole of the Discovery phase has been about responding to the problem definition, refining it and then setting out the routes by which the problem can be solved. This discovery report embodies the potential solution to the problems.

The technology will be further tested and refined in Alpha and then developed during Beta, against an agreed backlog of requirements, to ensure that the solution fully covers the expressed needs.

5.3 Provide a joined-up experience across all channels

Providing a joined-up experience across all channels is an aspect that is not relevant to Discovery but will be a key factor in the application design, particularly in the Beta stage.

5.4 Make the service simple to use

Making the service simple to use is primarily an issue for Beta, but part of the discovery engagement includes identifying user process needs, current pain points and looking for simple ways of connecting them.

This is covered in considerable detail by the user journeys.

5.5 Make sure everyone can use the service

Making sure everyone can use the service is a consideration primarily for Beta, where the design will take into account the GDS guidelines for accessibility.

5.6 Have a multi-disciplinary team

Having a multi-disciplinary team is also primarily a consideration for Beta, but this has also guided the make-up of the team in Discovery, which was an integrated joint team of DfT and CGI, covering multiple skills including data & solution architecture, user research and GIS technology.

5.7 Use agile ways of working

Using agile ways of working is again, primarily a concern for Alpha onwards, but the discovery process has also been run as an Agile process, using elements of both Scrum and Kanban.

5.8 Iterate and improve frequently

As with the previous item, iterating and improving frequently is mostly of concern during Alpha and later, but the Discovery process has also used regular iteration with sprint reviews and retrospectives at the end of each sprint, and a planning session for the next sprint, all of which ensure complete transparency of progress.

5.9 Create a secure service which protects users' privacy

Privacy will mainly be a design and implementation concern during Beta, but Discovery has also identified some common requirements for levels of access for specific sets of data.

Beyond the requirements for user-authentication, there is not expected to be any personally identifiable information kept within the system.

5.10 Define what success looks like and publish performance data

The overall success of the platform, and the targets for performance levels are largely concerns of Beta, however some success measures have already been defined during the Discovery process, and these are outlined below.

Criteria	Notes
Access to data by key stakeholders (DfT/ATE, LAs)	Are there restrictions on data access?
Completeness (of variables)	How much missing data will this solution contain?
Ease of maintaining data	How easy is it to continue to update this data going forward?
Geographical coverage	How complete is the coverage of England?
Integration with other systems/ datasets	Can the system interact with existing systems/ data be overlayed or extracted easily

5.11 Choose the right tools and technology

Choosing the right tools and technology is a key concern of Discovery. The options for the technology platform are set out in section 3.5.3.1. A preferred solution will be chosen and validated in Alpha phase (see section 4 for Alpha considerations), in order to best deliver the desired outcomes of the project, as well as aligning to DfT's Architecture Standards.

5.12 Make new source code open

Making new source code open is not an issue during Discovery as no source code is being developed or specified but will be a guiding factor during any bespoke developments during Alpha and Beta stages.

5.13 Use and contribute to open standards, common components, and patterns

Using and contributing to open standards, common components and patterns will be a guiding factor during any bespoke developments during Alpha and Beta stages.

5.14 Operate a reliable service

The service design will be fundamental, mostly during Beta. In Discovery the main way that this criterion is expressed is through making technological choices which can support a reliable service being delivered.

To that end, all of the options identified are either fully expected to be able to support a reliable service or will be tested in Alpha to ensure that they can do so.

6 Glossary

Abbreviation or Term	Definition
API	Application Programming Interface – a defined interface for software applications to communicate with each other
ATE	Active Travel England
CA	Combined Authority
DfT	Department for Transport
DOS	Digital Outcomes and Specialists marketplace
LA	Local Authority
Segregated	In active travel infrastructure, cycle or walking routes which are physically separated from vehicular traffic, eg by wands, kerbs or other physical barriers
TOGAF	The Open Group Architecture Framework

7 Endnotes

[i] From http://www.opengroup.org/public/arch/p1/togaf_faq.htm

[ii] “cycling” is used as an umbrella term to cover all human-powered wheeled vehicles, typically using specifically delineated physical infrastructure (“cycling routes”) – but should be understood to include hand-cycles, wheelchairs, and variants of bikes, such as cargo bikes or cycle-towed trailers.

[iii] of the 21 LAs surveyed, they shared data with 8 different organisations (a mix of private companies, their residents/the public, local government and charities). The most common data recipient was the scheme funder, (80% of LAs surveyed shared their data with the organisation they were receiving money from)

[iv] See section 3.7 – there are conflicting requirements expressed for immediacy of update versus stable cycle for reporting

[v] Assumption is that this is likely to be funded up-front with no/minimal ongoing costs. See also Alpha Considerations 4.5

[vi] https://www.pct.bike/

[vii] https://www.cyipt.bike/

[viii] https://actdev.cyipt.bike/

[ix] Note that these tools are being provided as an example of tools which use similar datasets, they have not at this stage been reviewed for suitability for inclusion in this option

[x] Will depend on the exact products selected and the mix of writers to readers. ArcGIS costing is a complicated matrix based on specific features and particularly the mix of how many users are content creators (“writers”) and how many are only consumers (“readers”).

[xi] Will depend on the existing tools included within scope and the ability to extend

[xii] Though the products are without charge, it would be advisable to include some level of support, where available

[xiii] To be determined during Alpha – see 4.4

[xiv] Also dependent on the specific products chosen and their capabilities, but the assumption is that bespoke functionality will be minimised where possible.

[xv] Depending on the technology of the existing tools and the ability to be extended

[xvi] Re-platforming onto shared database may need re-coding parts of the existing applications

[xvii] Would reduce the current need for support in each LA

[xviii] Should also improve/safeguard currency of other tools

[xix] Assumptions: significant effort required to deconstruct existing tools for options 3 or 4 – this may drastically overstate the amount of effort required. There is a specific Alpha task to examine this in more detail, if one of these options is to be chosen. The replacement option assumes significant amounts of data migration and training – depending on the number of systems in use, and the amount of training required for users of non-ArcGIS systems this may be considerably over or under-estimated.

[xx] Assumption here is that some features are able to be provided by the existing tools once they share the data

[xxi] Assumptions: ArcGIS Online hosted on its own private cloud, with hosting costs included in the online cost. Other options are hosted serverless on Google Cloud Platform in a data-heavy but fairly processor-light environment. Server cost is 10% of production load in Alpha and 25% in Beta.

[xxii] Hosting and license costs are so insignificant in comparison to the development and maintenance cost that the overall cost is effectively just (A)

[xxiii] Even with a single system, the acceptance of changes will involve some procedural controls

[xxiv] Note that this contradicts the “purchase rather than build” principle in Applications

[xxv] This relates to the desire to have a reporting cycle where the data is fixed at points to allow consistency – e.g. the state at the end of a particular month across various dimensions

[xxvi] Continuous Integration/Continuous Deployment – a set of tools for automating version control, testing and deployment of changes to the application