



Department for Transport (DfT)

Discovery/Alpha into adding accessibility data to the national database of public transport stops.

The challenge

Context



NaPTAN stores where public transport stops are located
A national database of bus stops, train stations, ferry terminals, etc. allowing transport operators to plan multi-operator and multi-modal journeys.



It's an XML file
NaPTAN has existed since the late 90's. It's a big XML file where most local authorities (LA's) submit all their public transport stops, in a schema version (2.1) from the early 00's.



Data 'producers' and 'consumers'
Our data 'producers' are local authorities (LA's), who currently upload this NaPTAN data via a spreadsheet upload. Our data 'consumers' download it and use it in their services, e.g. journey planning services like Citymapper.

The Ask



Add accessibility data to NaPTAN
Discovery and Alpha (in line with GDS standards) into incorporating accessibility data for every public transport stop into NaPTAN.



Understand the data
Learn what data local authorities (LA's) have and can provide to NaPTAN. Learn what data consumers would need to show accessible journeys to passengers



Develop a proof of concept in 10 weeks
Show that we can serve accessibility data alongside NaPTAN data to data consumers. We had only 10 weeks to deliver this.

What we built + how research influenced the solution



Ingest data
Consume spreadsheets



Through UR we learned:
- The data's almost always going to come as a .csv file.
- LA's don't have the resources to transform the data for us, so we have to accept the spreadsheets as they are.
- They will often not be directly linked to NaPTAN stop identifier (ATCO Code). It may take some work to link.



Process data
Put it into BigQuery



Data is processed from spreadsheets into a unified schema. That schema was iterated on using feedback from UR, for example we learned:
- It's useful to include: kerb height+type, the presence of tactile paving.
- It's important to be able to verify the information e.g. a field to show when it was last updated.



Serve data
Give it to consumers



UR told us:
- We should have an option for consumers to bulk download the data (REST API), as well as an option to download data for individual stops to see the data they need (GraphQL).

An example of 2 queries made in the GraphQL interface to our data, to get the accessibility data between 2 dates. The second one only shows stops that are wheelchair accessible.

```

Run >
1 query {accessibilityData (capturedFrom: "2023-11-09", capturedUntil: "2023-11-10 23:59:59") {
2   atcoCode,
3   isWheelchairAccessible,
4   hasAssistanceService,
5   commonName,
6   dateCaptured
7 }}
```

```

1 {
2   "data": {
3     "accessibilityData": [
4       {
5         "atcoCode": "300000181N2",
6         "isWheelchairAccessible": true,
7         "hasAssistanceService": false,
8         "commonName": "Newland Centre",
9         "dateCaptured": "2023-11-09T00:00:00Z"
10      },
11      {
12        "atcoCode": "300000181OC",
13        "isWheelchairAccessible": true,
14        "hasAssistanceService": false,
15        "commonName": "Odeon Cinema",
16        "dateCaptured": "2023-11-09T00:00:00Z"
17      }
18    ]
19  }
20}
```

```

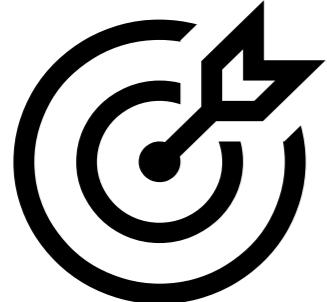
Run >
1 query {accessibilityData (isWheelchairAccessible: true, capturedFrom: "2023-11-09", capturedUntil: "2023-11-10 23:59:59") {
2   atcoCode,
3   isWheelchairAccessible,
4   hasAssistanceService,
5   commonName,
6   dateCaptured
7 }}
```

```

1 {
2   "data": {
3     "accessibilityData": [
4       {
5         "atcoCode": "300000181N2",
6         "isWheelchairAccessible": true,
7         "hasAssistanceService": false,
8         "commonName": "Newland Centre",
9         "dateCaptured": "2023-11-09T00:00:00Z"
10      },
11      {
12        "atcoCode": "300000181OC",
13        "isWheelchairAccessible": true,
14        "hasAssistanceService": false,
15        "commonName": "Odeon Cinema",
16        "dateCaptured": "2023-11-09T00:00:00Z"
17      }
18    ]
19  }
20}
```

My approach

The focus



I focused on data producers and data consumers

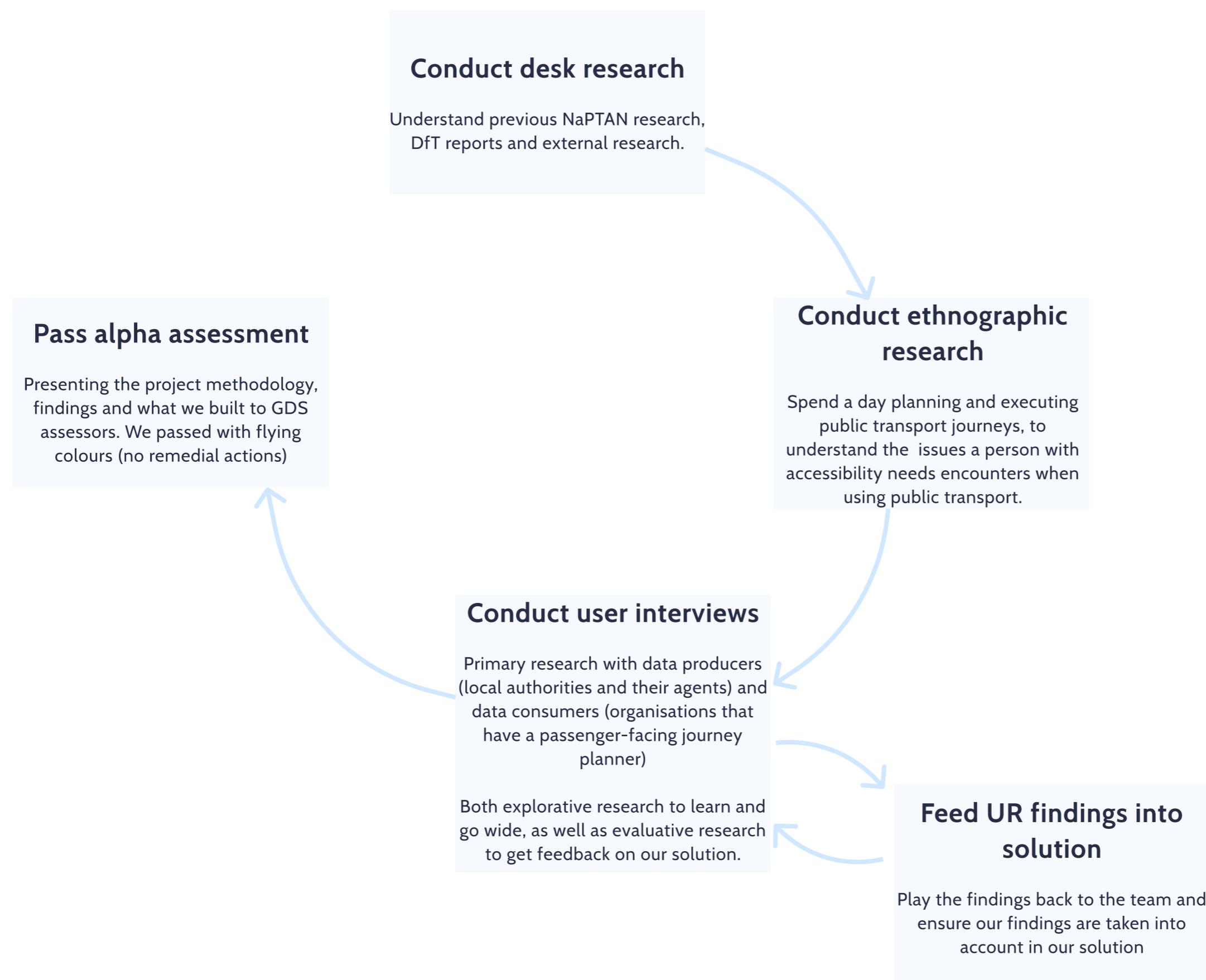
NaPTAN is essentially the platform that provides the data. It doesn't have control over the experience for passengers. Our primary users are data producers and consumers.

Travel planning apps have the interface to passengers. To create and test an end-to-end solution for passengers requires close collaboration with them, which I recommended as a focus for the Private Beta phase, after this discovery/alpha.

For this service to work, it needs:

1. To be able to ingest information from many different local authorities across the UK (data producers)
2. To be consumed and valuable to journey planners that provide the interface to passengers (data consumers)
3. To be valuable to the end users: passengers with accessibility needs, and provide them with agency and autonomy in making public transport journeys. but they're not the primary users of the service.

The process



Iterating on the schema with UR findings

Schema V1

```
AtcoCode: string (PK),  
StopCommonName: string  
IsWheelchairAccessible: boolean | null,  
HasAssistanceService: boolean | null,  
DateCaptured: Date -> Date when we last updated the raw dataset (base table)
```

This is the first schema the developers on the team came up with. It's rudimentary and the key elements are 2 boolean variables: is it wheelchair accessible? Is there an assistance service?

After iterating with feedback from user research, this is V2



Schema V2

 UR told us the consumers want the underlying data rather than the inference like boolean IsWheelchairAccessible y/n in V1. E.g. the kerb type is important for determining whether a bus can deploy a ramp, so provide the detail.

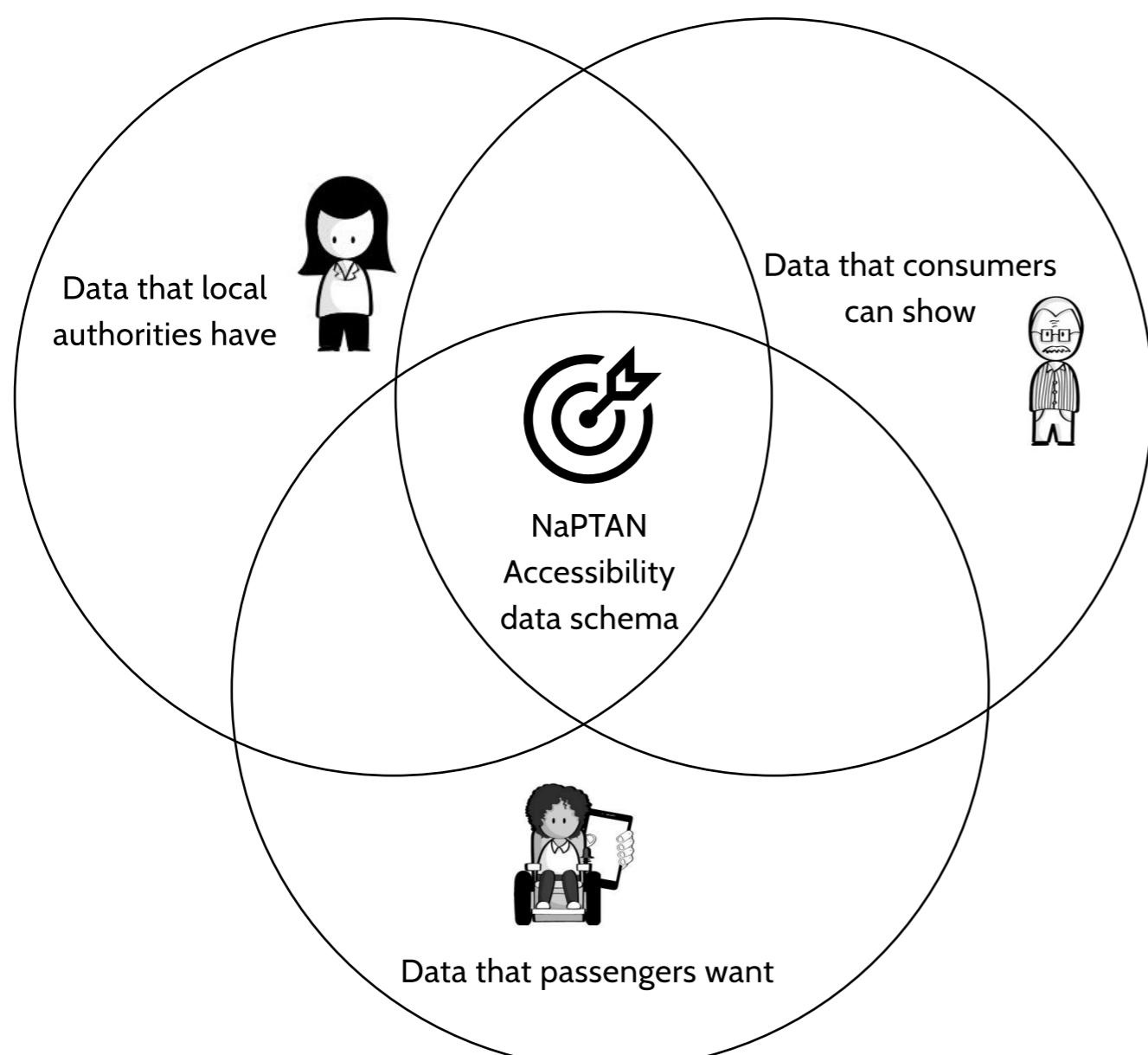


```
AtcoCode: string (PK)  
AccessibilityInferences: Object  
Kerb: string | null  
ShelterType: string | null  
ElectronicDisplay: boolean | null  
TimeTable: boolean | null  
DesignatedWheelchairSpace: boolean | null  
SafetyConcerns: string | boolean | null  
TactilePaving: boolean | null  
Notes: string | null  
Advisories: string | null  
DateCaptured: Date  
→ LastModified: Date
```

 UR told us that to verify the information, data consumers would find it important to understand when the data was last modified.



This process of iterating continued during the entire project, ensuring data consumers found the data helpful and data producers actually had that data and could supply it:



Ethnographic research



I spent a day with a powered wheelchair user and another user researcher.

The task for this user: plan and execute an accessible journey to Greenwich Observatory with public transport.

The route involved multiple modes: a bus, a London Underground ride and a ferry (Uber Boat).

The purpose was to understand first-hand the experience of a powered wheelchair user in planning and executing a journey using public transport.

We observed many issues that we had read about previously, and this was an incredibly valuable experience in developing empathy and first-hand understanding for the end users of service we are enabling.

We observed many aspects of the journey that could have been improved by information provided e.g. in a journey planning app, some examples:

- Is there assistance at the station/platforms involved in the journey?
- Is the bus stop involved accessible for my wheelchair? How steep are the ramps I'll be taking?
- Are there accessible toilets along the way?
- Is the wheelchair space occupied or free?

Some of the things observed were outside NaPTAN's remit, and others were within its remit.

Collaborating with disability groups

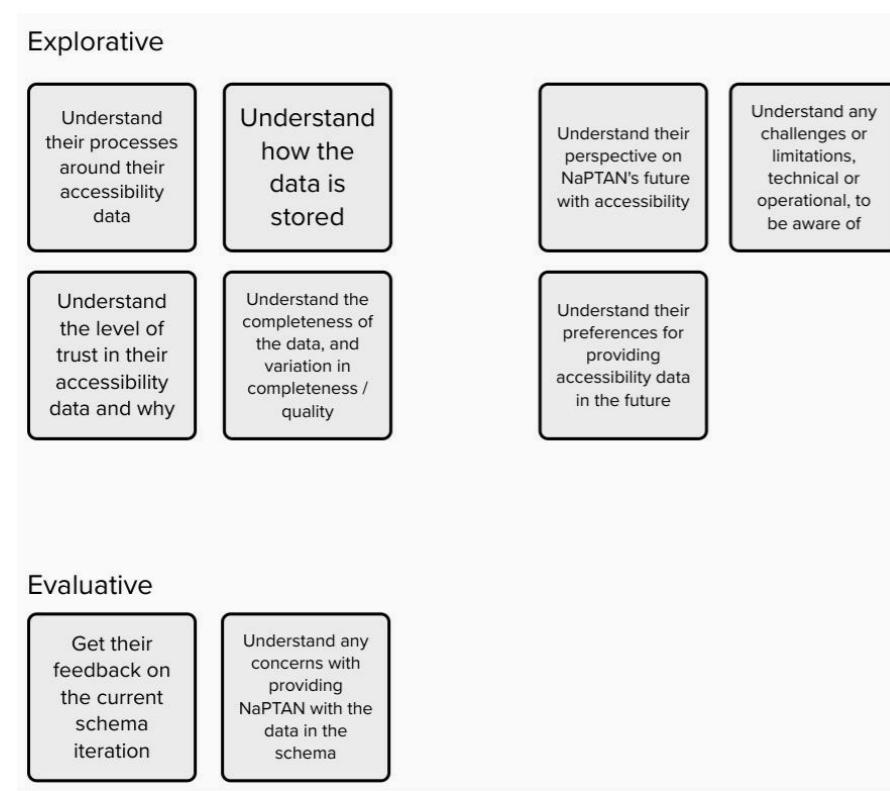


We had sessions with RNIB, Guide Dogs UK and DPTAC, who all represent disabled people. We checked our approach with them and the solution so far, and their feedback impacted our solution.

One example is RNIB and Guide Dogs UK both telling us that tactile paving would be a helpful data point in planning journeys for visually impaired people, so that was added to the schema after checking it's available in the LA data sets.

User interviews

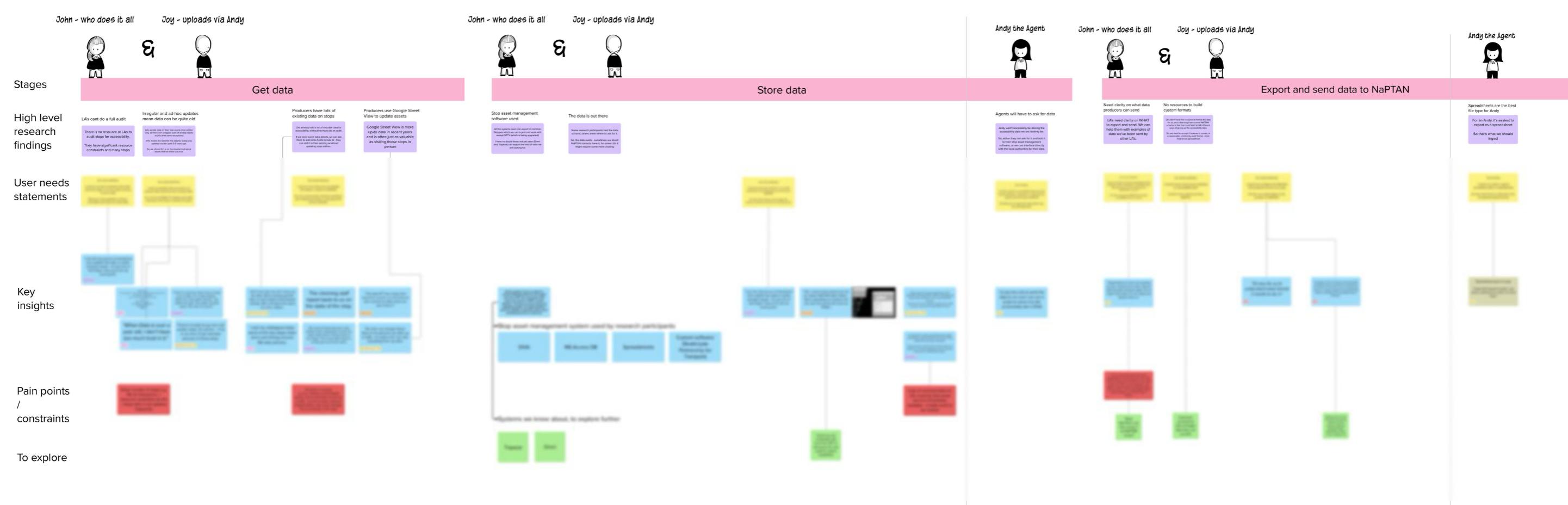
Research objectives for the producers:



Analysis of interview notes broken down by research objective

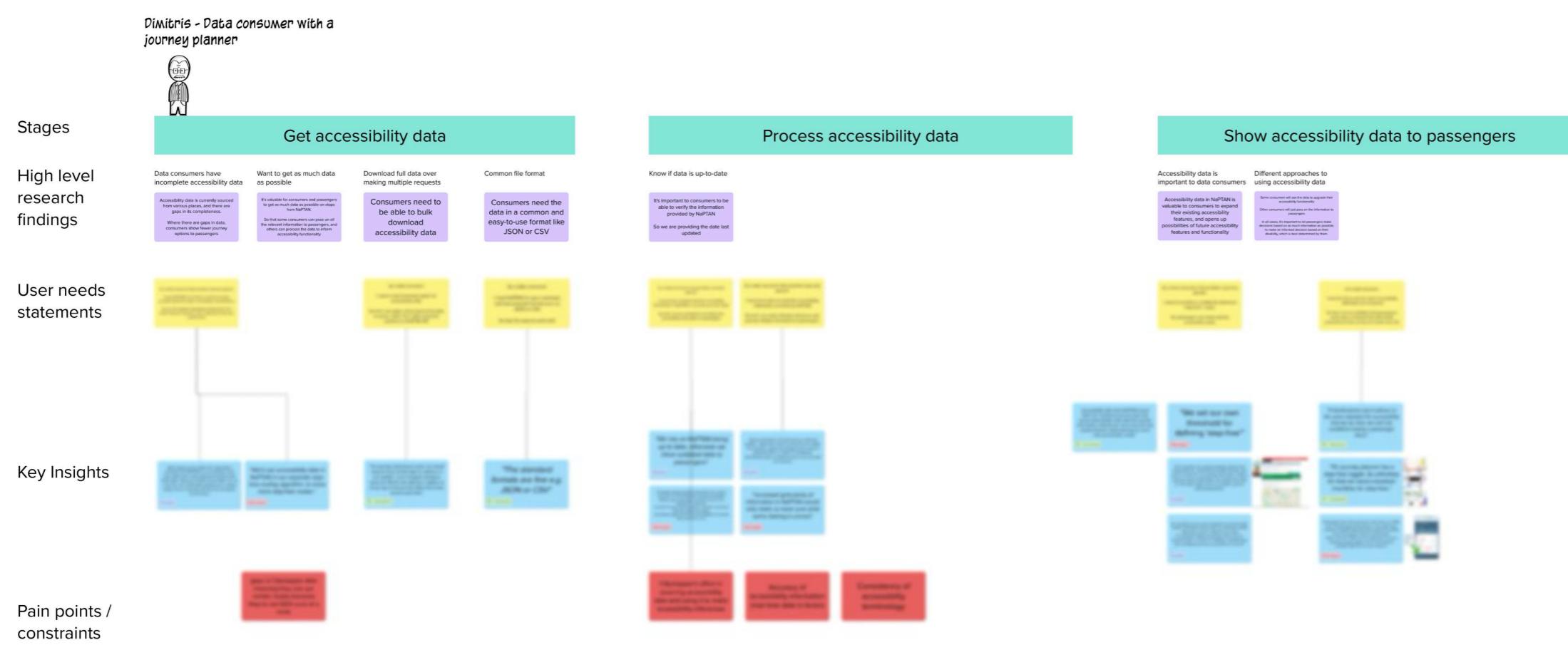


Summarised into this high-level overview of our user research findings



This summary artefact captures the research findings at a high level across the stages they go through: 'get data', 'store data' and 'export and send data to NaPTAN'.

I did the same for data consumers:



Feedback



The research plan you crafted has a good level of thoroughness. I had the opportunity to participate in few of the user research sessions, and I found the flow and structure you implemented to extract valuable information hitting the point.

During the playback sessions with the team, you distilled the complex information into its essence which has helped the team understand with ease. The user need statements were concise and effectively communicated the core requirements. Your role as a User Researcher in the Alpha assessment was significant, and your presentation was indeed impressive.

Lead Business Analyst



The User Research you have done for the NaPTAN Accessibility work has been brilliant. I have watched back several of the recordings of the sessions and they were well planned and delivered. You communicate very well with people, asking the right questions at the right times. The overall result is an efficient and effective conversation full of useful detail and insights. You also did a brilliant job of the presentations in the Inception of this SoW, in the UR playback sessions to the team and in the Alpha assessment presentation.

We have worked together on a few UI design refinements during this SoW and on that I found your approach to be logical and pragmatic. I really appreciate that you took the extra time for some ad-hoc sync ups with me on mapping out the accessibility data domain concepts and also cross-checking aspects of the data formats and platforms used by the data ecosystem.

Senior Developer