MadeTech

# MHCLG Case Level Data Collection

## Alpha Deliverable Pack

1. The problem to solve
2. Alpha approach & activities
3. Summary of findings
4. Recommended next steps

# The team

A blended team with members from MHCLG & Made Tech, with a mix of capabilities.

MadeTech

Ministry of Housing, Communities & Local Government

# A little about Made Tech...

Our mission is to improve society through technology.

Partnering with central government, local government and healthcare - GDS, MoJ, DVLA and NHS to name a few.

UK offices in London, Manchester, Bristol and Swansea.

We've been operating since 2012.

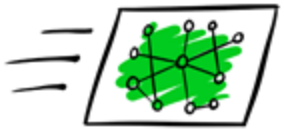

MadeTech

# Our missions

To help the public sector build world-class technology teams, that practice the sort of techniques and ways of working you would see in an internet-era business.

## Modernise

legacy technology and working practices



## Accelerate

digital and technology delivery

## Drive

smarter decisions with data and automation

## Enable

sustainable technology and delivery skills

1. **The problem to solve**
2. Alpha approach & activities
3. Summary of findings
4. Recommended next steps

# The problem to solve

Service context

CORE

Project vision

Actors and users

# MHCLG context

The Ministry of Housing, Communities and Local Government's goal is to be an **innovative**, **efficient** and **data driven** organisation in order to:

- Deliver better services to citizens
- Shape policy to benefit our communities and influence national debate
- Enable Local Authorities to create great places to live and work

The ministry is responsible for more than 30[1] statistical collections and 80[2] statistical releases every year.

This broad landscape of disparate data collections, data sources and data providers makes it challenging to obtain accurate and timely data.

1 https://www.gov.uk/government/organisations/ministry-of-housing-communities-and-local-government/about/statistics#statistical-collections
2 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896276/MHCLG_statistical_publications_2020-21.pdf

# Current challenges with data collection

Expensive legacy data collection systems aren't serving providers or analysts well

MHCLG want to build capability in machine learning and predictive analysis

The current technology stack is limiting automation

Local authorities are overwhelmed with data returns

Internal processes contain lots of nugatory work

**Good data + modern technology allows you to serve citizens better**

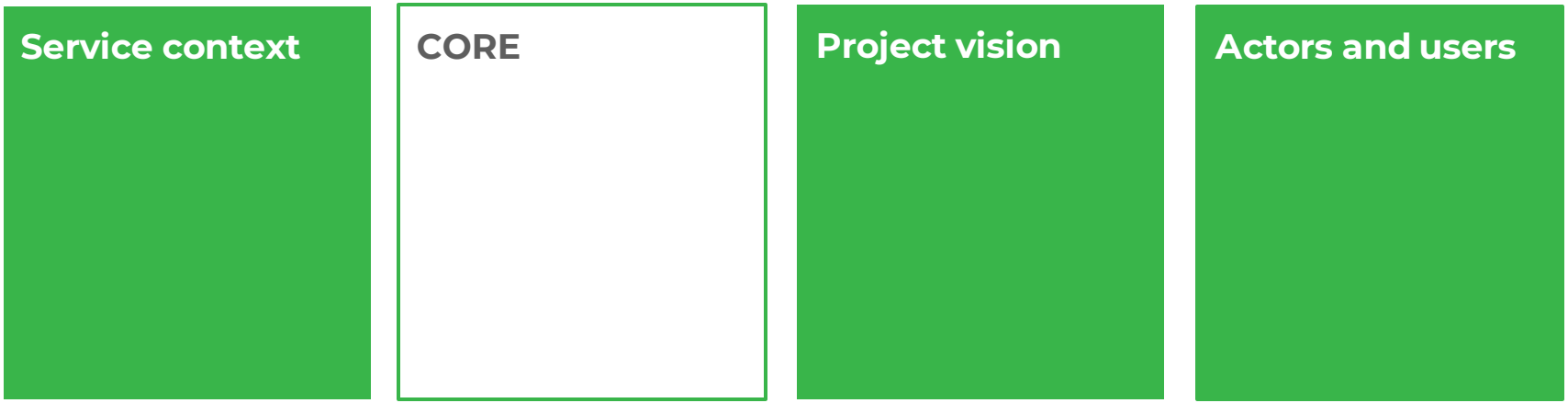
MadeTech

# MHCLG goals and aspirations

- ➔ Get richer and more accurate data in real time
- ➔ Modernise data collection methods, ensuring they're reusable across multiple datasets
- ➔ Serve digitally mature organisations better by allowing them to integrate with our systems / database
- ➔ Serve less digitally mature organisations better by providing a better user facing service to validate / submit data
- ➔ Lay the foundations for predictive analysis to inform policy interventions sooner

# As the pace of change increases and citizen needs evolve MHCLG must be enabled to respond quickly

# The problem to solve



# Pilot data set

MHCLG has multiple data collection systems on a shared infrastructure that all do broadly the same thing. One of these case level submissions is CORE, a **Co**ntinuous **Re**cording of Social Housing lettings and sales in England. We chose to use CORE as a pilot to allow us to deep dive into users needs, and get the service to market quicker.

CORE was selected because:

- 300,000 new social lettings reported per year, this figure has dropped by 17% in the past decade, whereas stock has increased by 3%.
- End of life technology
- 900 **active** organisations (~300 LAs ~ 600 PRPs) and 15,000 users

There are many use cases for the dataset, but it has limitations and some of it isn't considered reliable e.g homelessness data.

# What is CORE

https://core.communities.gov.uk/

Local Authorities and Housing associations are required to submit a log of each social housing sale or tenancy to CORE, including data about the property & the tenants or occupants.


Six different web forms for individual logs, or the option to bulk upload


Analysis portal to interrogate previous data sets for specific questions


Collection of related content including user guides, FAQs and alerts

MadeTech

# Problems with the existing CORE service

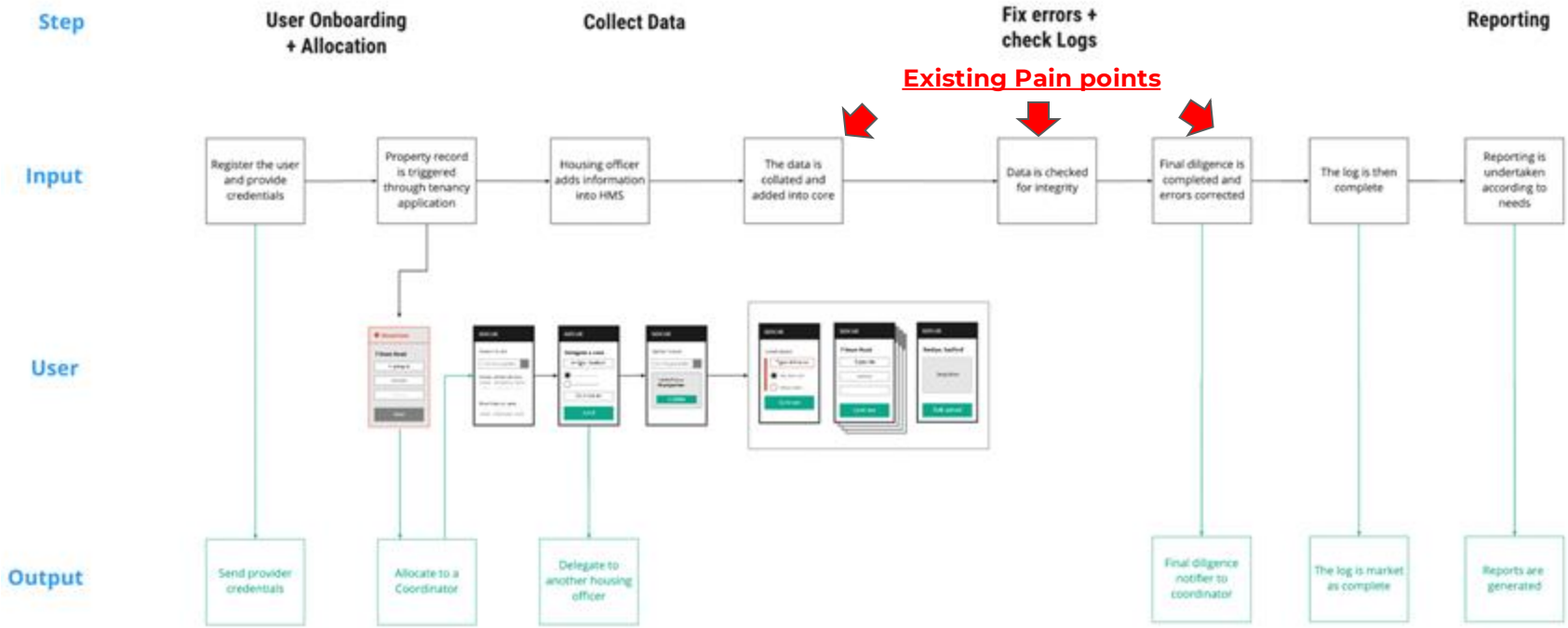
- A proportion of the social housing data MHCLG requests is collected as part of the social housing allocation process, but some is specific to CORE
- This data comes from multiple sources and takes a lot of time to collate
- Front line officers are spending up to 15-20 minutes per letting on manually re-keying data into the CORE website, which can cost medium-to-large organisations between ~£8,000 and £38,000 annually.
- Form and question design means some questions are simply skipped or guessed which leads to poor quality data coming in
- Housing managers spend a lot of time manually checking data for quality and completeness
- MHCLG's analysts spend up to 6 weeks chasing providers for missing data, and manually email users to query anomalies or quality concerns

# The current CORE team estimates it takes 3x longer to fix defects than it would in a modern tech stack

Technical overview document - Sharepoint

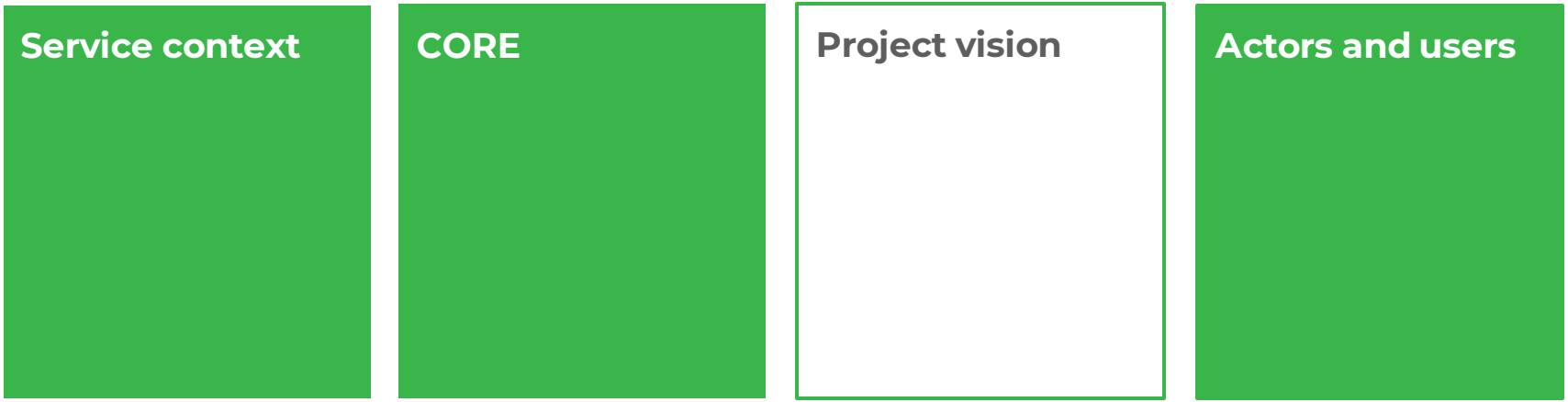
MadeTech

# Project context




MadeTech

# The problem to solve



# Alpha goal

The existing CORE website is due to be retired, and through a series of experiments, spikes and prototypes we've been testing assumptions and hypotheses for how we might:

- Save front line housing officers time to spend on their actual goals
- Get more accurate and comprehensive data into the dept, faster
- Offer value back to data providing org

*By answering these questions, we will start to draw conclusions & recommendations into how MHCLG should manage data collections and analysis in the future; particularly with regards to case level data collections.*

# Service vision

Make it simpler faster and more cost effective for organisations to share relevant, accurate and timely case level data on social housing

To help the government make the right decisions at the right time

## Vision

What is your motivation for creating the product?
Which positive change should it bring about?

**To help government make the right decisions, at the right time**

## Target Group

Which market or market segment does the product address?
Who are the target customers and users?

### Data Providers

**Agencies**

- Local Authorities
- Housing Associations

**Job Roles**

- Housing officers
- Lettings Manager
- BAs within Orgs
- Business support officers
- Housing performance manager

**Grouping**

### Data Consumers

- MHCLG Policy customers
- MHCLG Analysts

## Needs

Which problem does the product solve?
What benefit does it provide?

- Share data more easily
- Multiple submission methods → More choice → More automation
- Better usability → Faster submission
- Better usability → Less specialist training

- Better policy decisions
- Timely data submission
- Automated data QA
- Better quality data

## Product

What product is it?
What makes it stand out?
Is it feasible to develop the product?

- Multiple submission methods → Automation, Form Fills, Data Uploads → Users choose best for them
- Improve error display → Users can easily correct errors
- Easier to automate submission → Reduces human error
- User centred interaction design → Simplifies submission → Improves data quality

- Automated data analysis → Reduces time spent querying data

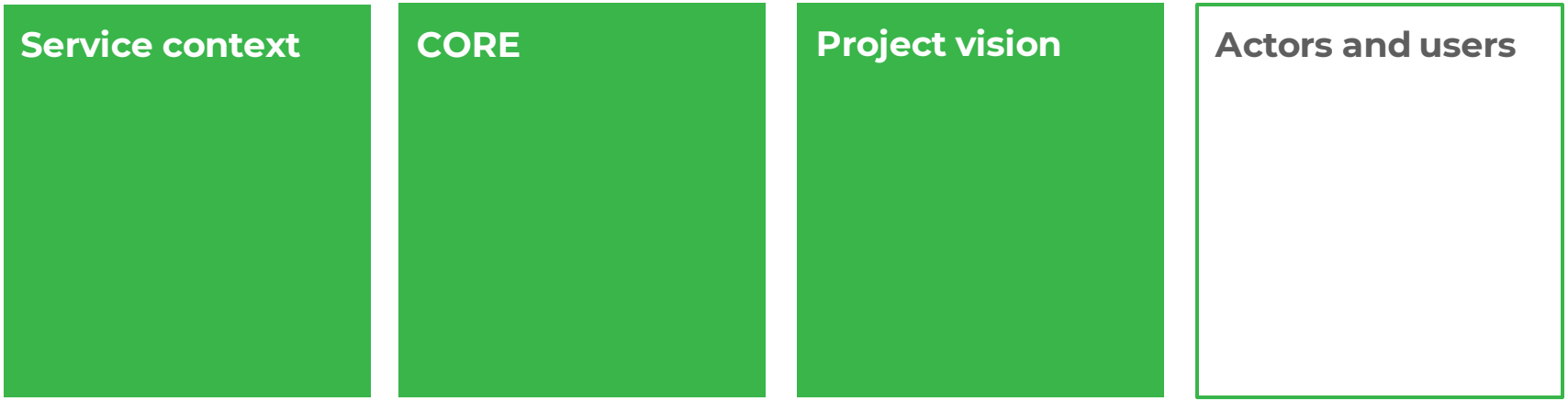
## Business Goals

How is the product going to benefit the company?
What are the business goals?

- Better data → Better insights → Better policy decisions
- Better data → Quantity
- Better data → Quality
- Better data → Timeliness
- Faster data ingestion → Set the standard for data ingestion

- More automation → Less manual tasks → More time for quality analysis and comparisons

# The problem to solve



# Key actors and users

There are three key groups of actors for CORE:

**Data providers**

- Local Authorities
- Housing Providers
- Tenants

**Data consumers**

- MHCLG analysts
- Policy makers
- Arms length bodies

**Support providers**

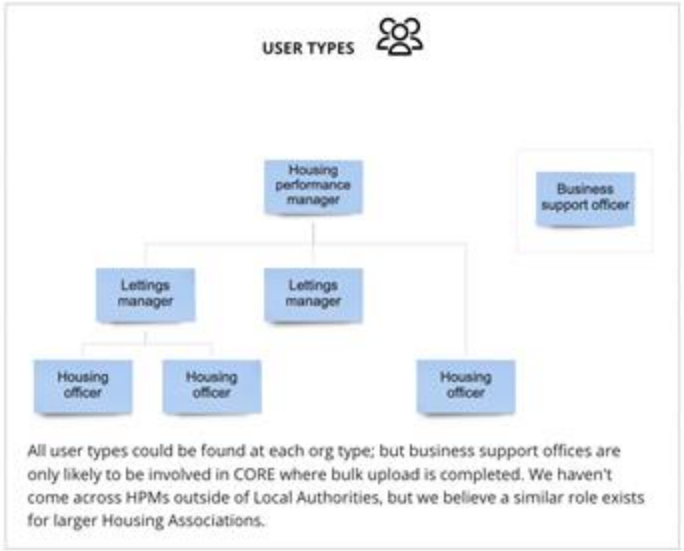
- Frontline helpdesk
- Training providers
- CORE product team

This Alpha was primarily focused on the needs and challenges facing Data Providers.
We also identified opportunities to improve the MHCLG analysts' processes that would improve the service for all users.

# CORE data providers

We identified two ways to differentiate data providers.

Identifying an individual user based on their role & responsibility to understand the wants, needs and context for that user.




Categorising organisations based on data storage & upload methods showed more commonalities than the size or function of the organisation.

# Service context for users

Users primary goals are to:

- Match the right people with the right property
- Get people in sustained tenancies
- Minimise void periods
- Meet their sign-up targets.

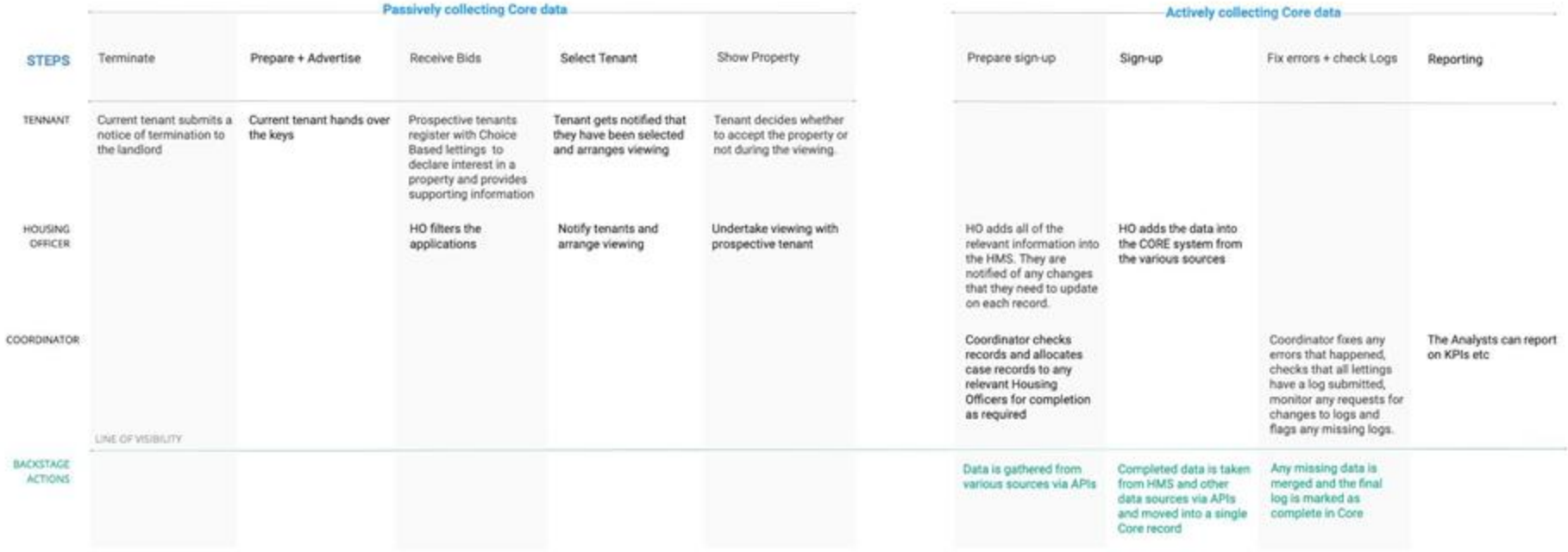
In amongst this, they're required by central government to log all new social housing sales and lettings with MHCLG and share relevant data about those sales and lettings.

This is data about:

- The sale or tenancy itself
- The property
- The household

# Service context

[View full service blueprint]




MadeTech

# User needs relating to data provision

- As a housing officer I need data submission to be effortless or automated so that I can spend more time on getting vulnerable people into sustained housing.
- As a housing manager/data coordinator I need to be able to share data with government quickly and easily without it disrupting our internal processes or the service we provide to tenants, so that we meet MHCLG requirements.
- As a housing manager/data coordinator I need to ensure data is quality assured so that we send MHCLG accurate data
- As a housing manager/data coordinator I need to ensure my team are working efficiently so that we can meet our organisation's housing targets
- As a housing performance manager I need to collect data so that it can inform business housing strategy.
- As a housing performance manager I need to benchmark our data collection and provision processes so that we can measure our housing strategy.

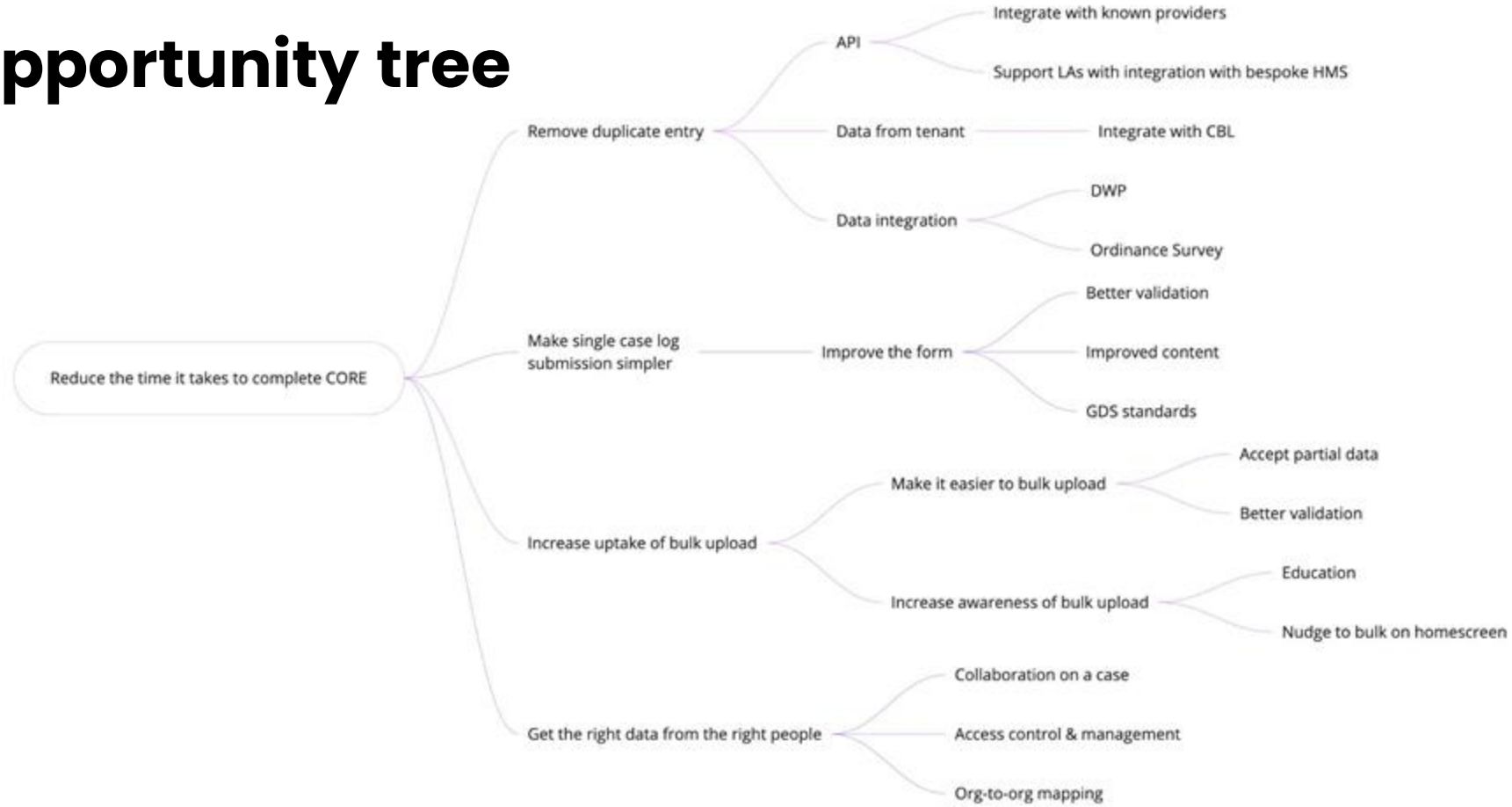
# User focused prioritisation

| Org Type | MHCLG Strategy |
| --- | --- |
| **Supplier Reliant CORE Module Users - 17%**<br>Bulk upload (+ some manual)<br>Proprietary tool | MHCLG should aim to automate the process for this group, to remove the need for a manual bulk upload. By adding an API and integrating with known HMSs we can get data quicker, and reduce the feedback loop on the validation errors - allowing them to be fixed within the HMS. This may also encourage a higher uptake on CORE modules, which will further reduce the burden of collating data in multiple systems, and increase the accuracy of the data overall. |
| **Supplier Reliant Form Fillers - 50-65%**<br>Single log<br>Proprietary tool | This group is the largest user group, and it's also the group the the most challenging process - manual form uploads, with multiple data sources for collation. The primary goal here should be to encourage automation, by making it a lot simpler to use a bulk upload functionality (or API). We will know this happens when more people tell us they use bulk upload, or have selected a CORE module in their HMS. The secondary goal should be to make improvements to the form, to reduce the burden on users. We will know this has happened when the manual entry speed decreases, and quality increases. |

# User focused prioritisation

| Org Type | MHCLG Strategy |
|---|---|
| **Not Supplier Reliant Form Filler - 5-15%**<br>Single log<br>Own solution | This group are unlikely to use a proprietary HMS, or build their own, due to the small number of logs. However this small number also means every form submission is likely to be even more burdensome. To reduce this burden MHCLG should improve the form to make it self explanatory with in form validation and guidance. Ensuring LAs can collaborate on a case to support their smaller HAs with getting the right data will improve the quality. Getting as much data as possible from other sources will also be particularly beneficial to this group. |
| **Uses bespoke solutions - 10%**<br>Bulk upload<br>Own HMS | The primary strategy for this group should be to improve the bulk upload process, to limit the number of manual interventions required and speed up the validation process. The API will be available to this group, but uptake would require each LA to integrate themselves. MHCLG could provide funding via the LGDCG to support LAs who share their custom HMS with other LAs, to encourage further reuse and allow LAs to move away from proprietary tools. |

# Opportunity tree



1. The problem to solve
2. **Alpha approach & activities**
3. Summary of findings
4. Recommended next steps

# Alpha approach & activities

| Approach | Hypotheses | Assumptions | Timeline |
| --- | --- | --- | --- |

# Our approach

- We worked as one team, fostering a culture of feedback and collective ownership
- We started using **design sprints** and then switched to **scrum methodology** based on team feedback
- We **tested** a range of **hypotheses** derived from user needs & pain points identified in discovery using prototypes
- We rapidly **iterated** prototypes to delve deeper into insights gleaned from user research
- All team members participated in **User Research** sessions

# Alpha activity themes

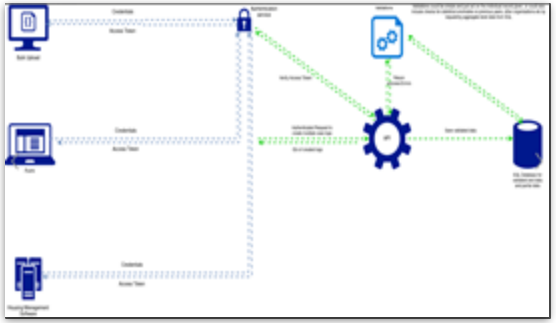
We have been exploring three main activity tracks throughout the alpha:

## Research and analysis

- Service blueprint generation
- Generative research
- Quantitative survey
- Provider interviews
- Metrics & data baselining
- Persona iteration


## Prototyping

- Service homepage
- Paper form
- Single case web form
- Bulk upload
- Data playback
- Derived variables



## Technical analysis

- Review of existing architecture
- Feasibility of the hypotheses
- Architectural research & decisions
- Data pipeline strategy

# Ways of working

- Working in the open, with regular retrospectives and open showcases
- Collaborative, with collective ownership of the project outcomes and priorities
- Experimental, focusing on rapid learning through ‘doing’
- User centred, with user needs at the centre of all the work done, and the entire team sharing responsibility for user research

**Key links**

Project plan

Showcases:

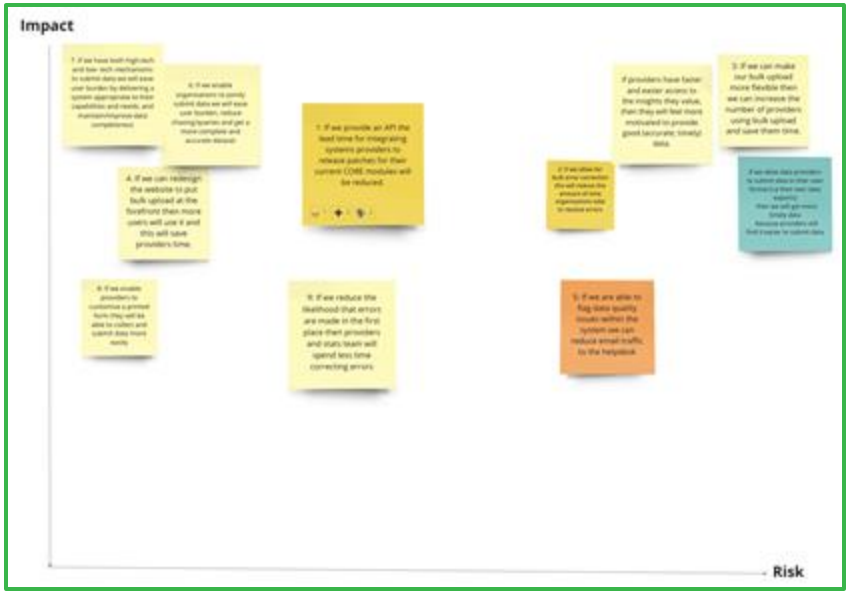
- Showcase 1
- Showcase 2
- Showcase 3
- Showcase 4
- Showcase 5
- Showcase 6
- Showcase 7
- Showcase 8
- Showcase 9

# Alpha approach & activities

| Approach | Hypotheses | Assumptions | Timeline |
| --- | --- | --- | --- |

# Hypothesis prioritisation

Based on the user needs, we identified several hypotheses for how we could reduce the burden on data providers; by making data submission quicker, more accurate and simpler. These hypotheses were prioritised by perceived impact & risk.




MadeTech

Hypothesis 1

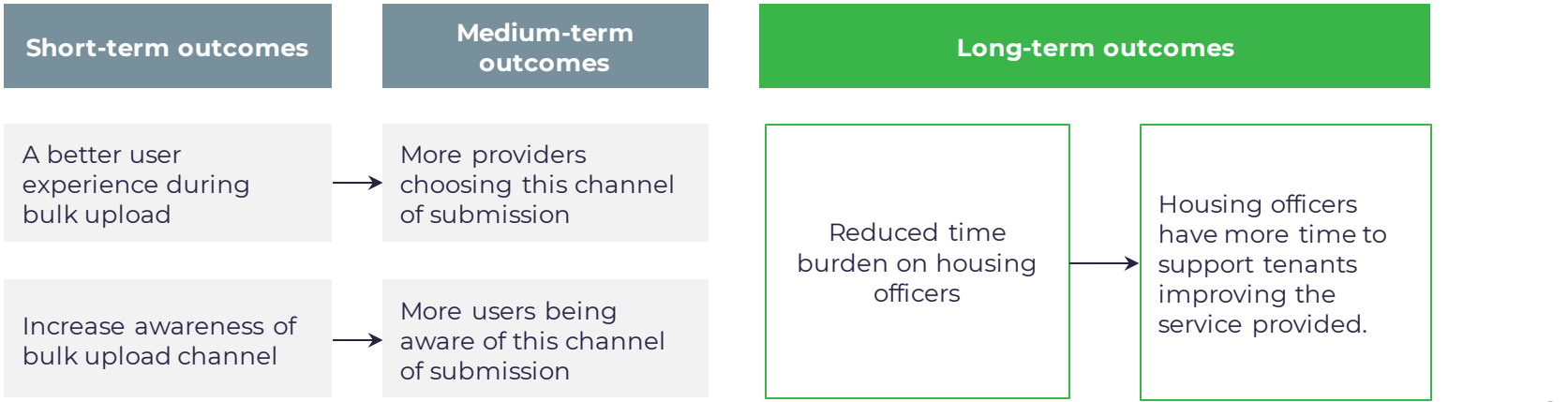
# Increasing the usage of bulk uploads

## Because we know that

The majority of providers use a very manual approach and 73% of cases submitted are completed using the online form via single file upload

## We believe that

Improving the bulk upload experience and nudging providers towards this channel will save a lot of time, effort and money.




MadeTech

Hypothesis 2

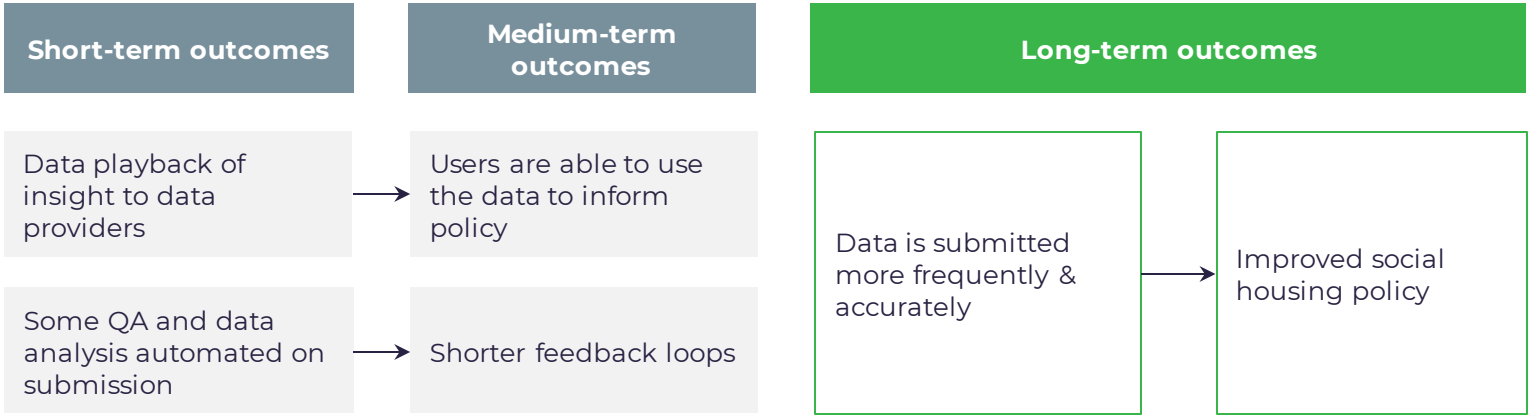
# Playing back data in the same system

## Because we think that

Users have told us they don't see any value in submitting CORE data, and aren't sure how or why the data is used. The delay between submitting data and hearing back about quality concerns is also a pain point.

## We believe that

If we playback data insights in the same system, users will be more motivated to submit accurate, timely data. If we can automate some of the data analysis then we can shorten the feedback loops.




MadeTech

Hypothesis 3

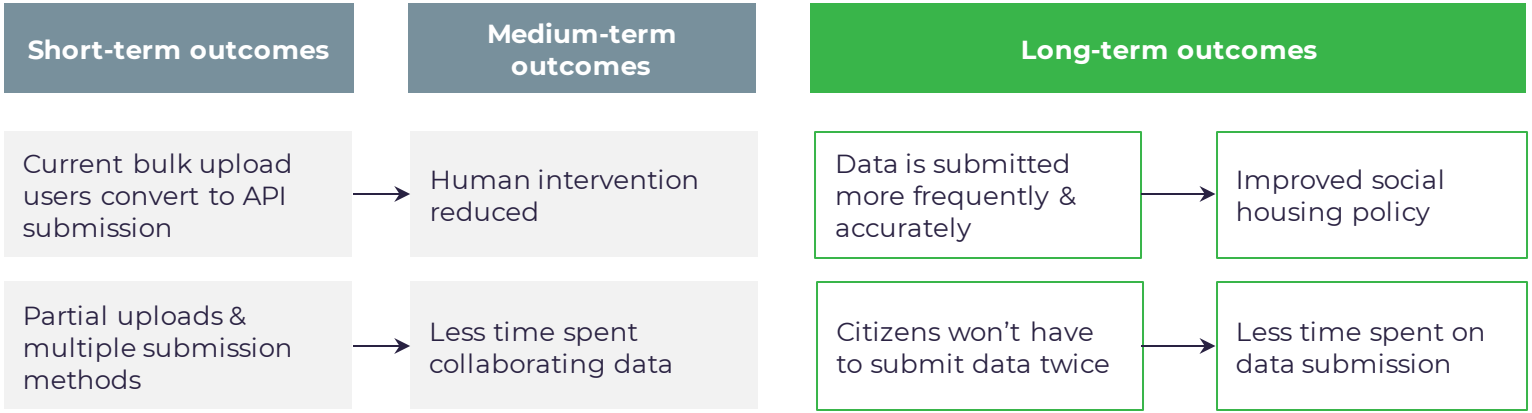
# API for data submission

## Because we think that

Users are double entering the data they have into CORE, and then spending significant amounts of time collating the remaining data points as needed.

## We believe that

Providing an API will allow automation of submissions, removing the need for human intervention. Allowing submission from multiple sources will reduce the need for data collation.




MadeTech

Hypothesis 4

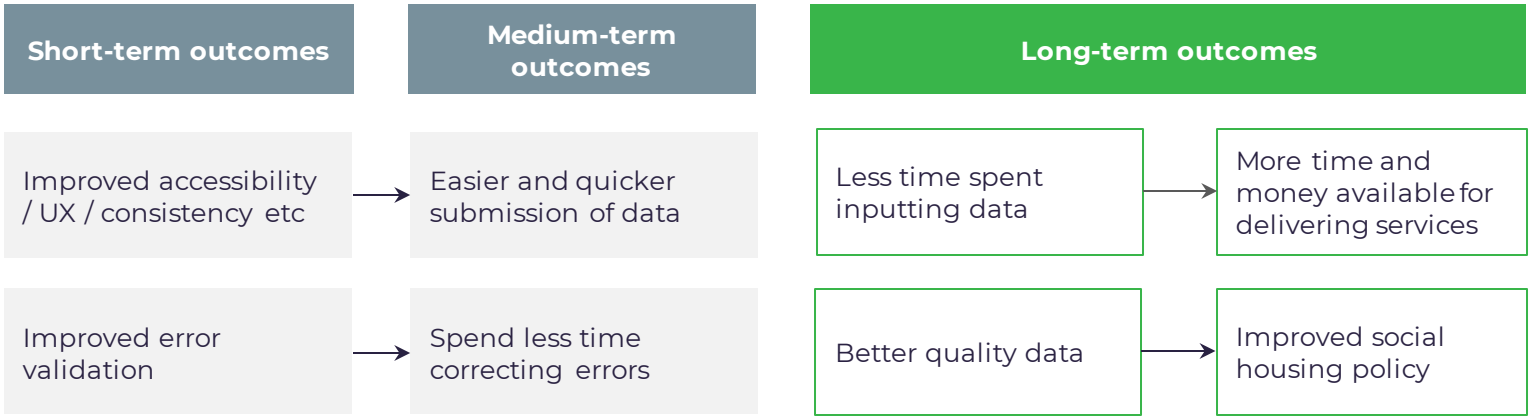
# Improving the single case submission form

## Because we think that

76% of CORE data is submitted using the single case form, which takes 15-20 minutes per case and is error prone. This is estimated to cost between £8,000 and £38,000 / year / organisation.

## We believe that

Redesigning the form to comply with GDS standards and meet the needs of users through better content, validation & contextual help, will make the form easier & quicker to complete.




MadeTech

# Alpha approach & activities

Approach

Hypotheses

Assumptions

Timeline

# Riskiest Assumptions

| Category | User Needs | Assumptions |
|---|---|---|
| Flexible uploads / Data from the source | As data provider, we need data submission to be fast and intuitive so that we can spend more time delivering services<br>As data providing org, we need a less manual way to submit data so that we can spend more time delivering services | • If we can make our bulk upload more flexible, then we can increase the number of providers using bulk upload and save them time.<br>• If we allow data providers to submit data in their own format (i.e their own data exports) then we will get more timely data because providers will find it easier to submit data<br>• If we reduce the likelihood that errors are made in the first place then providers and stats team will spend less time correcting errors |
| Data linking (external datasets) | As data providing org, we need data submission to be fast and intuitive so that we can spend more time delivering services<br>As data providing org, we need a less manual way to submit data so that we can spend more time delivering services<br>As a data provider we need our data to be accepted first time so that we don't have to do it again | • If we prefill fields then providers will be able to complete and submit form data faster<br>• If we link CORE data to data from other government departments, then we will be able to reduce the burden on data providers. |

# Riskiest Assumptions

| Category | User Needs | Assumptions |
| --- | --- | --- |
| Bulk upload nudges | As data providing org, we need a less manual way to submit data so that we can spend more time delivering services | • If we can redesign the website to put bulk upload at the forefront then more users will use it and this will save providers time. |
| Replaying insights back to the providers | As a data provider we need to see value of submitting data so that we can offset the cost<br>As a data provider we need an easy way to analyse / benchmark the data output so we can make decisions for our area | • If we make data available to providers to use then they'll see the value of providing it and will be more inclined to continue submitting it<br>• If we can can better educate users regarding what Core Data is used for, they will see more value in it and submit more accurate and timely data. We know this because many people state in UR that they don't know why they submit it. |
| Kiosk mode (Tenant fills in form) for tenant / data from the source | As a data provider we need our data to be accepted first time so that we don't have to do it again | • The more data tenants enter, the more complete and accurate data would be, especially if the forms were auto-translated into other languages and optimised for the use of assistive technology |

# Riskiest Assumptions

| Category | User Needs | Assumptions |
|---|---|---|
| Design better forms | As data providing org, we need data submission to be fast and intuitive so that we can spend more time delivering services<br>As a data provider we need our data to be accepted first time so that we don't have to do it again | • If we only use GOV UK form design elements to construct single log forms, users will find them easier to complete and the data submitted will be more accurate.<br>• If we clearly state how each data item will be used and what value it has for local organisations, they will invest more time in data collection and the data collected will be more timely and accurate. |
| Customised forms | As data providing org, we need data submission to be fast and intuitive so that we can spend more time delivering services | • If we enable providers to customise a (printed) form they will be able to collect and submit data more easily |
| Collaborate on a case | As data providing org, we need data submission to be fast and intuitive so that we can spend more time delivering services | • If we enable organisations to jointly submit data we will ease user burden, reduce chasing/queries and get a more complete and accurate dataset |

# Riskiest Assumptions

| Category | User Needs | Assumptions |
|---|---|---|
| Automated submissions | As data providing org, we need a less manual way to submit data so that we can spend more time delivering services | • If we provide an API the lead time for integrating systems providers to release patches for their current CORE modules will be reduced.<br>• If we provide an API then data providers will save time on their data submissions and MHCLG will receive more timely data |
| Error Handling | As a data provider we need our data to be accepted first time so that we don't have to do it again | • If we allow for bulk error correction this will reduce the amount of time organisations take to resolve errors<br>• If we are able to flag data quality issues within the system we can reduce email traffic to the helpdesk<br>• If we reduce the likelihood that errors are made in the first place then providers and stats team will spend less time correcting errors<br>• If we automate feedback that bulk uploaded data is invalid that will reduce the amount of time an organisation takes to correct it.<br>• If we bring more of the quality assurance calculations into the system, we can get more direct feedback from providers and save time |

For each hypothesis, we considered the assumptions that needed to be true for it to solve a user need, and tested these using prototypes

MadeTech

# Alpha approach & activities

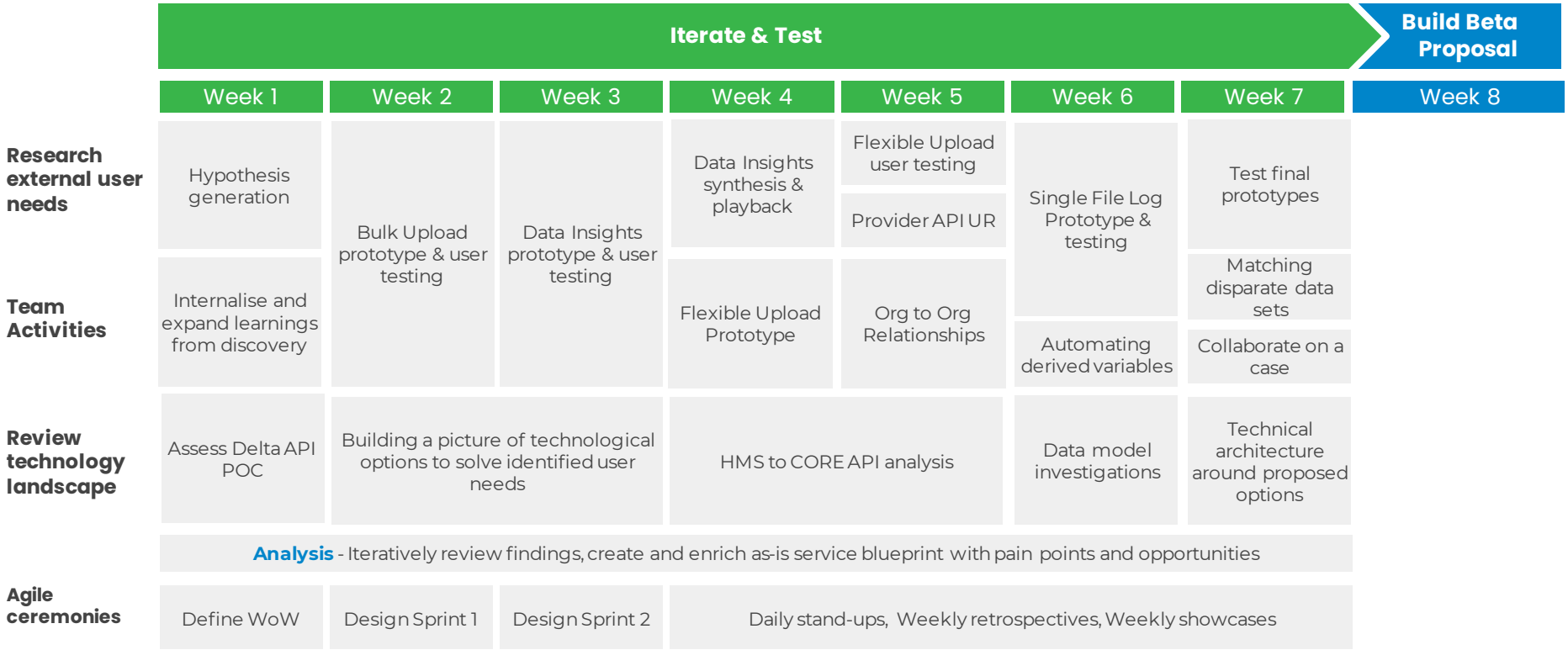
Approach

Hypotheses

Assumptions

Timeline

# Activity timeline




MadeTech

1. The problem to solve
2. Alpha approach & activities
3. **Summary of findings**
4. Recommended next steps

# Summary of findings

**Research and analysis**

**Prototype iteration**

**Technical exploration**

**Lessons learned**

# User centred research

Throughout the Alpha, we have continued to ensure user needs are at the forefront of all activities, continually prioritising based on user research and testing. All team members participated in user research, testing & analysis to build a shared understanding of user needs.

**User research activities included:**

- Persona and user type creation & iteration, used to segment and target research participants
- Quantitative user survey on CORE & HMS usage
- Contextual user interviews to build upon discovery research to understand current experience, pain points, barriers and behaviours
- Usability and scenario-based prototype testing with users
- Housing management system provider interviews
- Time studies to benchmark current experience of submitting individual logs
- Joint analysis sessions with all research observers and wider team

# Research participants

- 1 x user survey (205 users)
- 8 x rounds of user research
  - 4 prototypes tested
  - Policy user interviews
  - HMS Supplier interviews
  - Data provider interviews
  - API scoping
- 56 users spoken to
  - Data providers
    - 23 x housing associations
    - 10 x local authorities
  - 6 x policy makers
  - 4x Software providers

| | | Supplier reliant CORE module users: 17% | Supplier reliant form fillers: 50-65% | Not supplier reliant form filler: 5-15% | Uses bespoke solutions: 10% |
|---|---|---|---|---|---|
| **Research Interviews** Understanding our users and their processes | Data Providers | 2 LA<br>1 LM, 1 BSO, 1HPM | 2 HA, 1LA<br>1 LM, 1HPM, 2LM | 2 HA<br>2 LM | |
| **Hypothesis 1** Improve awareness and usability of bulk data upload | Iteration 1 | 1 HA<br>1 BSO | 3 HA, 1 LA<br>3 LM, 1 HPM, 1 BSO | | 1 HA<br>1 BSO |
| | Iteration 2 | | 1 ALMO, 2 HA<br>5 LM | 1 HA<br>2 LM | |
| **Hypothesis 2** Playing back data to providers | Iteration 1 | 1 HA, 1 LA<br>1 HPM, 1 BSO | 2 HA, 2 LA<br>1 HPM, 1 HO, 1 BSO, 1 LM | | 1 LA<br>1 LM |
| **Hypothesis 3** Integration via API | Software Suppliers | *4 x HMS Providers* | | | |
| | Data Providers | | 3 HA, 1 LA<br>4 HO | | 1 LA, 1 HA<br>2 LM |
| **Hypothesis 4** Single Form Submission | Iteration 1 | | 3 HA<br>5 HO | | |

MadeTech

# Alternative channels

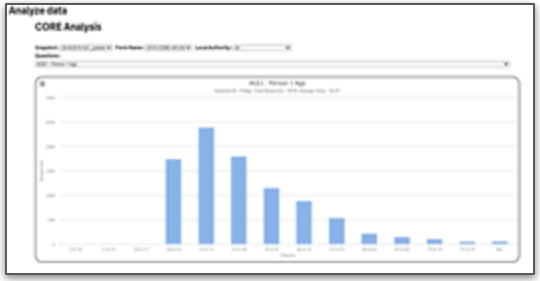
As well as submitting data via the CORE website, we also identified several alternative channels through which they may interact with the service.


**CORE Documentation**

- Service explanation
- FAQs
- Guides & manuals
- Service alerts


**User Assistance**

- Help desk
- Email support
- Call centre support




**Analyse CORE Data**

- Local authority level
- National level
- Raw data
- Reports

MadeTech

Research Insight #1

# Accessing data is a barrier to automation

There was a genuine appetite for automation (bulk upload, API) and it would solve a real problem.

However the users often can't, or don't know how to extract their data in bulk

This leads to:

- Less use of bulk upload functionality
- Inaccurate data & lower compliance
- Additional time spent collating data

The concept of partial uploads has implications on the way of working. Providers use CORE to monitor service delivery; so any change to this process would need careful user testing in Beta.

"The barrier is that we don't store all CORE questions in our Housing Management System"

"Collecting the information is probably the most difficult part"

The new service should:

- Provide digitally mature bulk upload users an API to connect to
- Move more users to bulk upload by being flexible about the data format & increasing bulk upload signposting
- Limit the need for manual collation by allowing partial uploads

Research Insight #2

# Users want to know the value of their data

When testing the data playback prototype, there was some interest in the data, but not enough to motivate them to change their processes. The lack of real-time & comparative functionality limits the usefulness of any data insights provided. Users want to compare themselves to other LAs, and to view data trends, real-time.

Users also expressed a lack of clarity around how and why MHCLG use their data. They want to understand how CORE data leads to improved social housing policies, that in turn provide better outcomes for citizens.

“I think that fact that we do spend lots of staff time completing the forms, it would be nice for us to benefit from that.

For years we have seen it as a data inputting exercise that we haven’t maximised the potential of”

# Memorable quote from Alpha

"We're collecting data for government, but not capturing it on our own system.

I want officers helping residents, I don't want officers helping the system. Officers need to just do their jobs and that will satisfy me as a manager, the council and government. Systems need to work for officers.

We need to save money as a council. We have community long covid. Councils house those with serious needs. I need to save money but also help those people. I need to be data driven, where do I allocate those resources? Help me target those resources, predict vulnerability... are there things we collect in a roundabout way that indicate that someone is struggling in life, that we can get an officer over there?"

The new service should enable real time data submission and faster feedback loops to provide  usable insights to LAs.

MHCLG should share the analyses & insights produced by ADD and shout about policy successes.

Research Insight #3

# Question design will improve the experience, but data quality needs to be tackled holistically

The prototyped single case submission form improved the users experience and reduced the mental load. Having the option to review answers at the end encouraged users to check their responses.

However, to make significant improvements to data quality, we need to look further upstream, to the collection & collation of data. There are some questions where we know accuracy is lower, for example HAs are often not provided with the “reasonable preference” data from LAs, which leads to this response being guessed or skipped.

“It's very awkward for a support worker to ask them these questions when they don't know them... so they don’t ask them”.

"The information you are collecting is probably not worth collecting in the state it is in""

**MHCLG should integrate with OGD data to reduce the number of data points requested, and increase accuracy.**

**For example the income & benefits question only has 40–50% completion, but we could collect this from DWP.**

Research Insight #4

# Users want CORE processes to improve

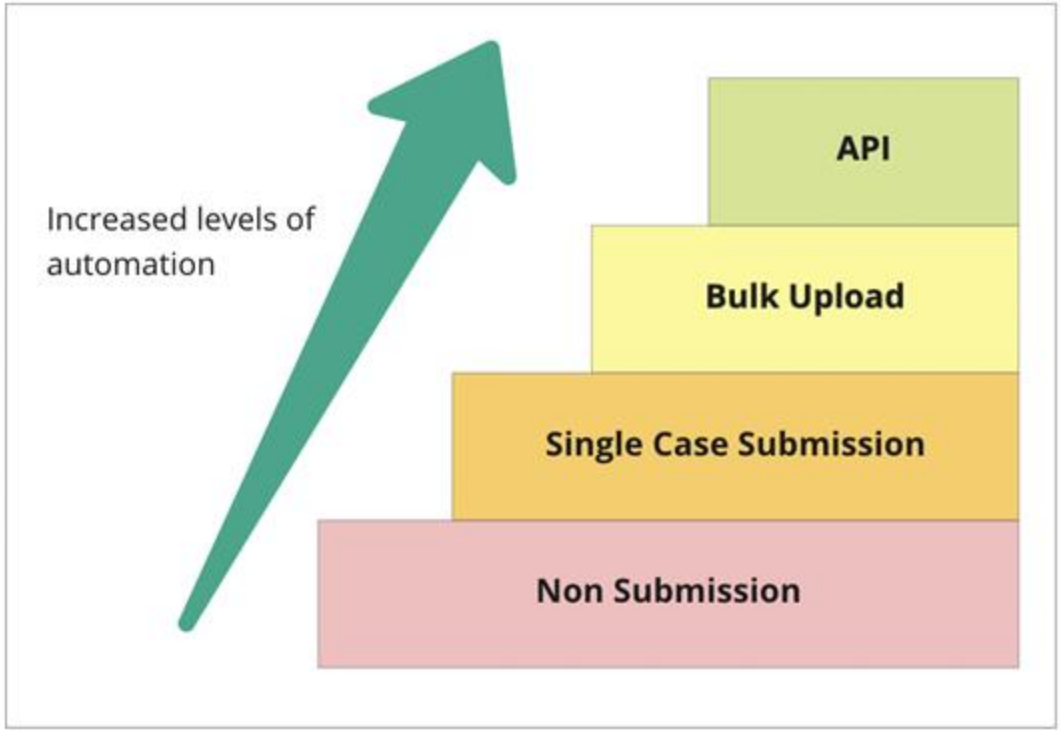
Throughout the alpha users showed genuine enthusiasm for any changes that would speed up and make more efficient the process of submitting CORE.

Some users had already attempted their own ‘workarounds’ to tackle their pain points, including designing a bespoke CORE form, a spreadsheet to collate data from different sources, or building their own validation processes to get around errors in CORE. However these solutions were only available to well-resourced organisations and were incomplete.

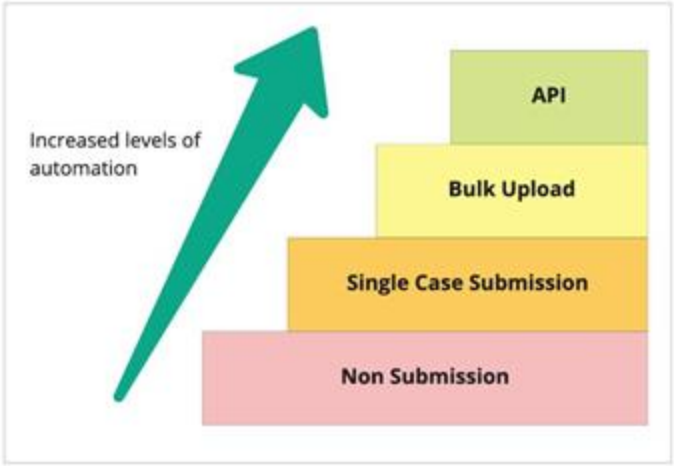
“From what I’ve seen...the wider team would be really interested in anything that saves time”

"Organisations would thank you for making it easier”

# Moving users towards increased automation



# Enabling increased levels of automation




- Submission API
- Bulk Upload
- Partial Bulk Uploads
- Collaborate on a Case
- Flexible Uploads
- Form Improvements
- Improved Question
- 3rd Party Data Linking
- Performance
- Aggregate Data
- Faster Feedback Loops

# Measuring the service

A successful service outcome should mean that a user has had their need met, whilst a key business or policy goal was achieved.

| User goal | KPI | Metrics |
|---|---|---|
| As a data provider we want to choose the best submission method for our organisation | Increased digital take up | Size of organisation's stock v their method of submission |
| As a data provider we want to reduce the time our housing officers spend completing forms | Time spent completing data submission | Cost per transaction |
| As a data provider we want our submission to be accepted first time and not be queried months later | More accurate data<br>Fewer help desk calls<br>Less analyst time chasing<br>Completion rate | Data quality metrics<br>No. help desk queries<br>No. of QA emails<br>Google analytics |

# Measuring the service

A successful service outcome should mean that a user has had their need met, whilst a key business or policy goal was achieved.

| User goal | KPI | Metrics |
| --- | --- | --- |
| As a housing officer I want to be able to complete a data return form quickly and confidently | Improved customer Satisfaction | GOV.UK Satisfaction survey |
| As MHCLG we need to annual changes to data requests to be accurate and efficient | Faster turnaround time for form changes | No. of sprints |

# Summary of findings

Research and analysis

Prototype iteration

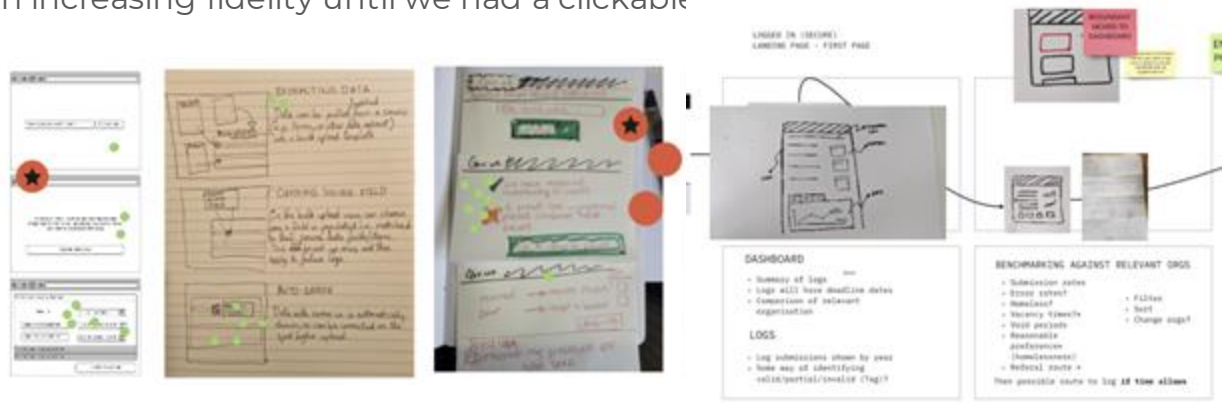
Technical exploration

Lessons learned

# Prototyping

Prototyping was driven by our hypotheses, and the assumptions that needed to be true in order for the hypothesis to be valid. We designed research activities that would give us the most amount of information, with the least effort. We used prototypes to test our different hypotheses and their inherent assumptions by addressing specific questions.

For each prototype; we co-designed paper solutions, and then iterated in increasing fidelity until we had a clickable




## Prototype details

Link: http://ec2-18-168-105-184.eu-west-2.compute.amazonaws.com/
un: mhclg pw: l3tm31n

Tech stack:
Express.js
Nunjucks
Python
AWS

Prototyping approach:
Hypothesis testing with throw away prototypes

Github repository:
https://github.com/communities-uk/mhclg-data-collection-alpha

Prototype iteration #1

# Data upload and mapping

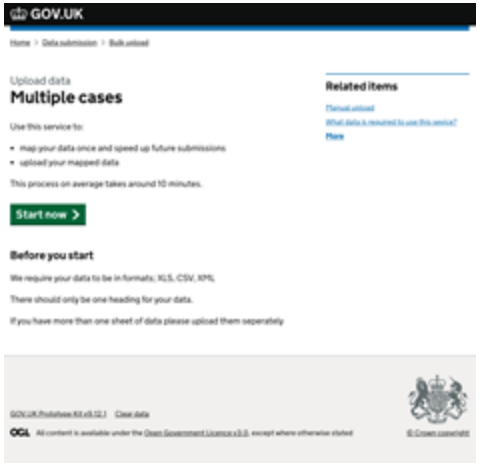
**Hypothesis**: We believe that improving the bulk upload experience and nudging providers towards this channel will save a lot of time, effort and money.

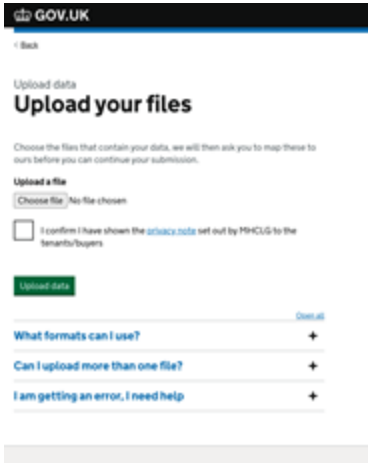
| What we did | What we learned | Actions |
|---|---|---|
| **User research:**<br>We tested first iteration of mapping fields with 5 users.<br><br>**Research goals:**<br>● We identified three key questions we wanted to answer with user research:<br>● Can we nudge users to bulk uploads?<br>● What would make bulk upload more efficient?<br>● What pain points not addressed by the prototype? | Mapping field values and ensuring there is a match in providers data and what MHCLG expects, was seen as valuable. This reflects a recurring pain point for providers around validation errors.<br><br>Users expressed a need for pre-populated fields to reduce manual data entry.<br><br>The current need to upload 4 separate CSV files was highlighted as a barrier.<br><br>Users highlighted that the annual question changes require them to update the submission methods each year. | Build a second prototype iteration which considers how might we allow users to map their data values to ours, and how can we improve error handling.<br><br>Allow organisations to set up templates, to reduce repeated entry of answers that remain the same.<br><br>Ensure bulk upload allows partial data, and 1 file instead of 4. Use question headers on the CSV template, not numbers. |

MadeTech

Prototype iteration #1

# Data upload and mapping





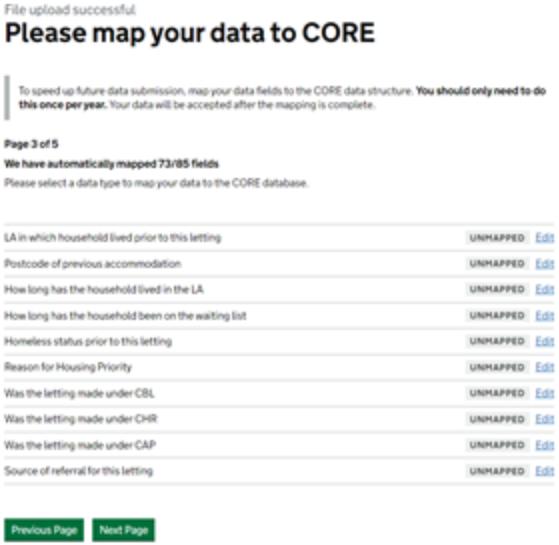

New bulk upload service pages, with improved explanation and in service guidance, following GDS best practices

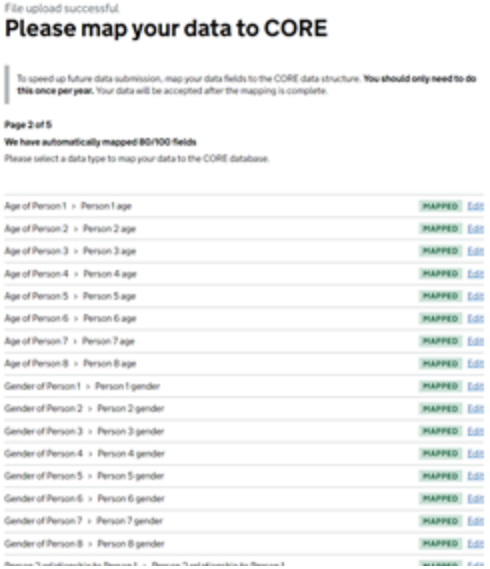
On upload, a response showing the columns automatically matched by the service, and an option for the user to match the remaining field - e.g. tenancy start date

MadeTech

Prototype iteration #1

# Data upload and mapping





Pages showing progress, clearly highlighting how many pages and fields are left.

An updated completion page, adhering to GDS guidelines

MadeTech

Prototype iteration #2

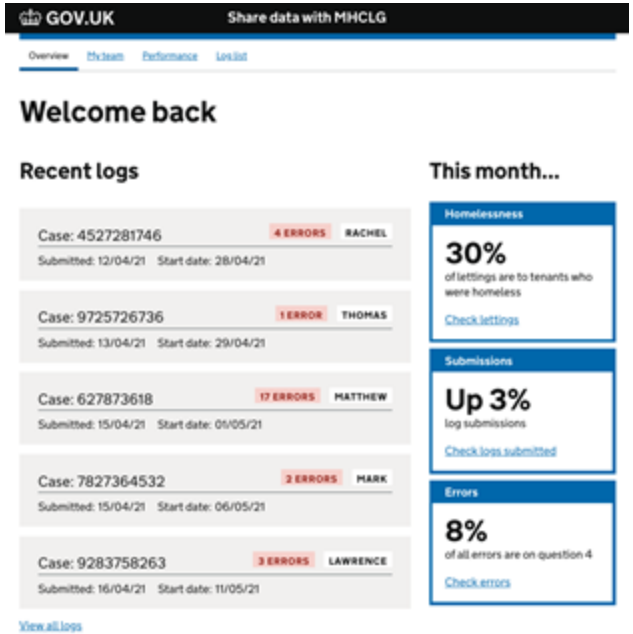
# Data insights & playback

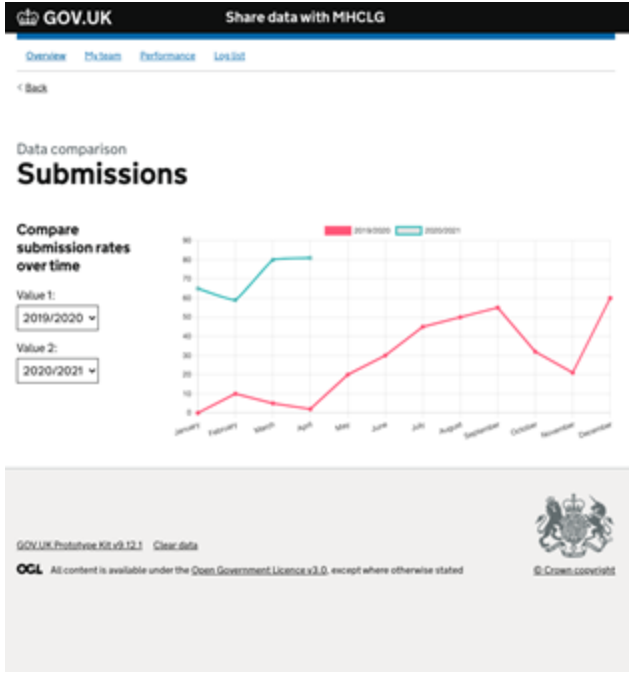
**Hypothesis**: If we playback data insights in the same system, users will be more motivated to submit accurate, timely data.

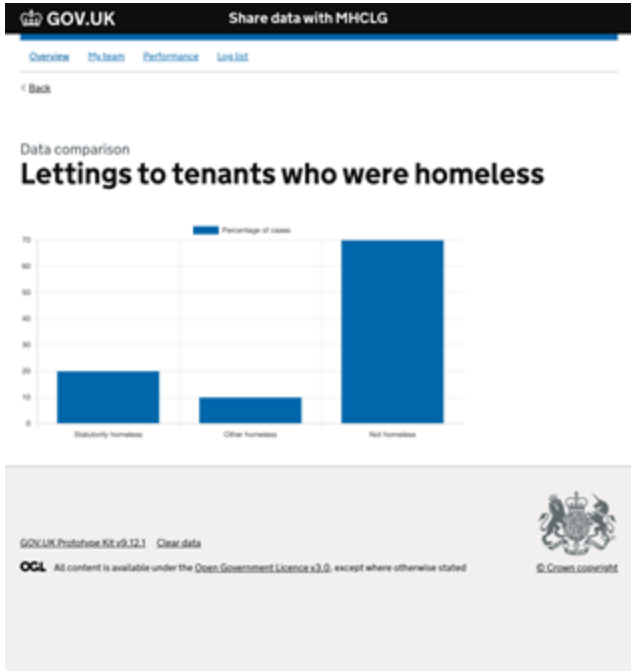
| What we did | What we learned | Actions |
|---|---|---|
| **User research:**<br>We tested the prototype with 7 users, from a variety of user groups.<br><br>**Research goals:**<br><br>● What do users imagine data insights / visualisation to be?<br>● How useful is this?<br>● Would they use this feature?<br>● Would users be more likely to submit timely, accurate data if CORE provided data these insights visualisation? | The data insights seemed to provide interest, but not a compelling motivation to adapt their processes, so this would not be incentive enough to improve data submission.<br><br>Data providers would be interested in real-time data analysis, where they could interact with the data to make comparisons and explore current trends.<br><br>Data coordinators would like an improved dashboard where they can monitor & support their organisations case submissions. | We deprioritised the hypothesis of a data playback during the Alpha, and instead pivoted to consider how we could shorten the feedback loops by:<br>● Automating data submission (e.g. via an API)<br>● Automating the data processing (e.g. quality assurance & derived variables)<br><br>While the research was inconclusive and did not support the hypothesis, we know that a paid tool HouseMark exists. Therefore we will continue to explore this hypothesis further in Beta, to understand whether a data playback is valuable to users. |

Prototype iteration #2

# Data insights & playback




A data coordinator dashboard, showing a summary of all organisations logs, with some data insight examples

Example comparison graphs, allowing users to compare their data to previous years, in a variety of graphical formats

MadeTech

Prototype iteration #3

# Mapping Cont. (Flatfile)

**Hypothesis**: We believe that improving the bulk upload experience and nudging providers towards this channel will save a lot of time, effort and money.

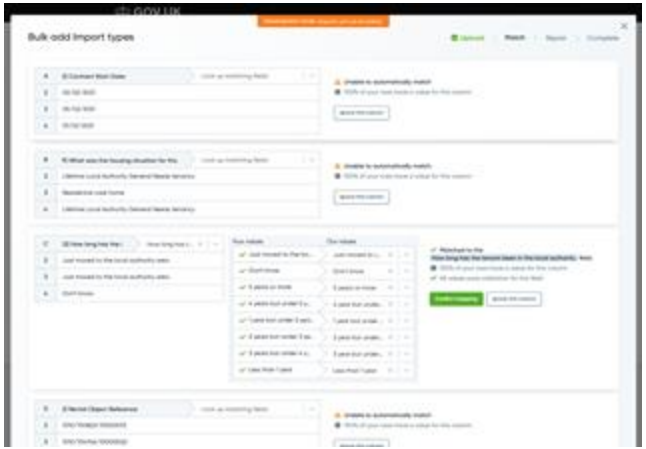
| What we did | What we learned | Actions |
|---|---|---|
| Continued exploration of mapping complex questions for bulk upload, using Flatfile to speed up development and insights gathering.<br><br>**User research:**<br>Tested with 4 users who do not currently use bulk upload<br><br>**Research questions:**<br><br>● Is a more flexible, bulk sharing option desirable to providers?<br>● Can they share information in this way? | There is genuine appetite for bulk upload (desirability) - It solves a real problem for users as a quicker mechanism for sharing data with MHCLG<br><br>But there are some viability and feasibility problems that need to be tested further (in beta) to make sure organisations are able to share information in this way. | The user research confirmed the value of a bulk upload functionality, but the following need to be tested further during Beta to make bulk upload accessible to more users:<br>● How can we make it easier for users to export data from their systems.<br>● How can we integrate with other data sets to make it easier for providers to collect good quality data.<br><br>The benefits of partial uploads were clearly demonstrated, but the implication on ways of working was a concern, as there were concerns it would be hard to manage. This needs to be tested further during beta. |

Prototype iteration #3

# Mapping Cont. (Flatfile)


We tested an improved service page, with more explanatory text.


Using Flatfile, we were able to test a tabular format for reviewing & updating data




This prototyped allowed automatic data mapping.


It made it easy to show all errors in one place, and quickly update as needed

Prototype iteration #4

# Single case form redesign

**Hypothesis**: Redesigning the form to comply with GDS standards and meet the needs of users through better content, validation & contextual help, will make the form easier & quicker to complete.

| What we did | What we learned | Actions |
|---|---|---|
| Redesigned the single case submission form, combining 6 separate forms into 1, using show/hide logic to amend the question flow as required. We used GOV.UK components and tested content improvements for the questions & contextual help.<br><br>**User research:**<br>Tested the prototype with 5 users, doing a usability comparison between the existing & to-be version. Iterated content between tests.<br><br>**See research goals on next page** | Users had a better experience using the prototype, which they valued<br><br>Unsure how much time it would save them. Most of the time is spent finding information in different places.<br><br>Users just don't have access to some of the information we ask for. This results in guesses, refusals and 'don't knows'<br><br>There is a mismatch in what we ask for and what data they have. | Build a better UI into any solution<br><br>Figure out the problem of disparate datasets<br><br>CBL research. More understanding of how we can get this information, or don't ask it.<br><br>Conduct content review with analyst and policy makers, AND users. Redesign how we can get accurate information to answer what MHCLG wants to know |

Prototype iteration #4

# Single case form redesign

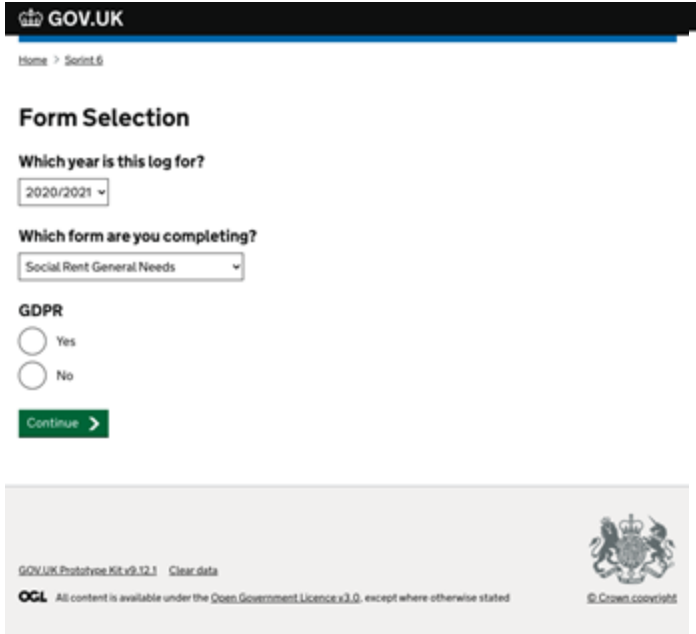
**Hypothesis**: Redesigning the form to comply with GDS standards and meet the needs of users through better content, validation & contextual help, will make the form easier & quicker to complete.

## Questions we wanted to answer

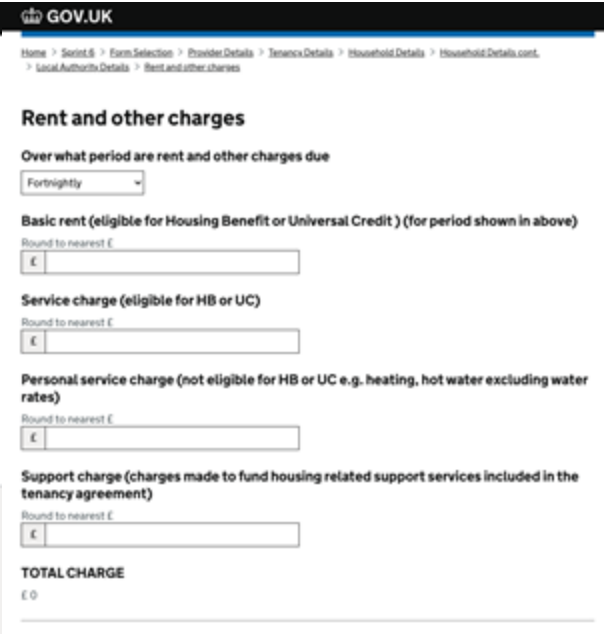
- Is the form quicker and easier to complete when organised into chunks with a re-designed question flow?
- Are certain GDS style guidelines appropriate for this type of user, an expert entering large amounts of data.
    - Less dropdowns
    - No question numbers
    - Way of asking questions (content)
    - Review of answers at end
    - Ability to go back and make changes
    - General UI
- Is there a better, more understandable way to word some of the lengthier questions?
- Does ‘de-cluttering’ by creating conditional questions make it easier to fill out the form?
- How can we better support users to not make errors?

Prototype iteration #4

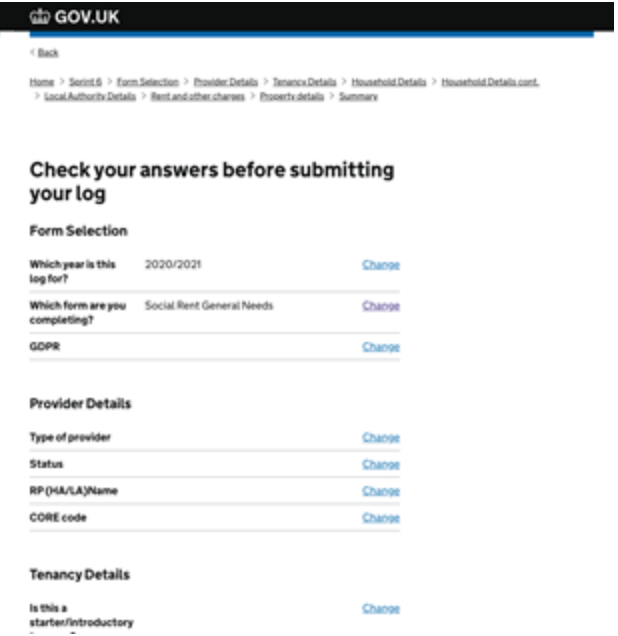
# Single case form redesign




We combined 6 forms into one, using question responses to show & hide subsequent questions




We provided in form contextual help to assist with completion, using GDS components.

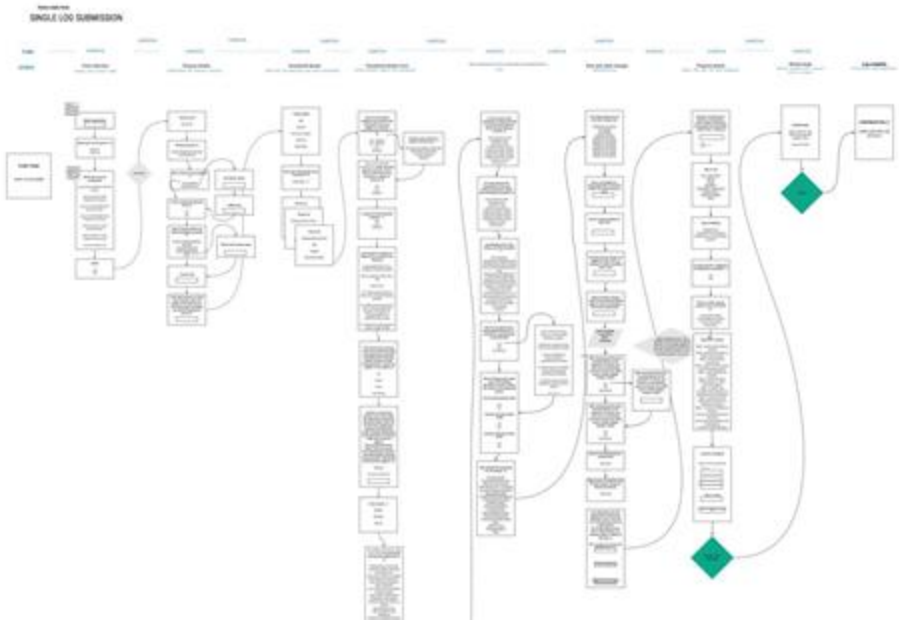



We allowed the respondent to review & change their answers at the end

MadeTech

Prototype iteration #4

# Single case form redesign - content design



| Questions to focus on | Known user (housing officer) pain points | Known user (analyst) pain points |
|---|---|---|
| Q28 Location of property (General needs) | List of all LAs 313 (cognitive overload) | Only collect postcode which isn't specific enough |
| Q14b Reasonable preference | Unable to answer | Accuracy and quality |
| Q13 Homelessness | Unknown why they struggle (How do they know what to put for a response) | Not accurate data collected/ Comparison with another data collection |
| Q11 Housing Situation | Unknown why they struggle (How do they know what to put for a response) | Not accurate data collected (Other option is used frequently) Comparison with another data collection |
| Scheme data not visible after code inputted | Trip validations (moved to area, same postcode) | Some errors (Wrong scheme selected) Until x date and using beyond that date. |
| Q6 In receipt of benefits | Struggle with what to choice and why without benefits details. | Accuracy (lots of don't know) |

We redesigned the question flow to group similar questions together & simplify submission. More detail can be viewed here.

We prioritised 6 questions that are known pain points, and iterated the content & contextual help for those. More detail can be viewed here.

MadeTech

# Prototype development approach

The prototype was bootstrapped using the GOV.UK Prototype Kit this allowed for easy use of common Government Design System components and quick development of prototypes.

Backend functionality was prototyped quickly using serverless technologies including AWS API Gateway and AWS Lambda with containerised functions.

The software was stored in an open repository in github and pushed up to AWS hosting using Terraform and Github actions.

# Prototype design principles

We used the GOV.UK Design System for all system prototypes except the second iteration of the flexible uploads prototype. Using the prototyping kit ensured WCAG compliance to maximise accessibility as:

- All colours used had a high contrast level
- All text was a good size
- The font was a clear typeface.

Each page was set out so that a user could zoom in up to 300% without the text spilling off the screen.  Elements were semantically ordered allowing the user to navigate most of the website using just a keyboard.

We assume it would be possible to navigate most of the service using speech recognition software, but this was not explicitly tested at this stage.

The Flatfile component was the only element that did not use the GDS prototyping kit.  Best effort was made to reuse GDS styling, but the component works in a modal and utilises a spreadsheet in the browser which means it would need redesigning in Beta to be accessible to all users.

# Summary of findings

Research and analysis

Prototype iteration

Technical exploration

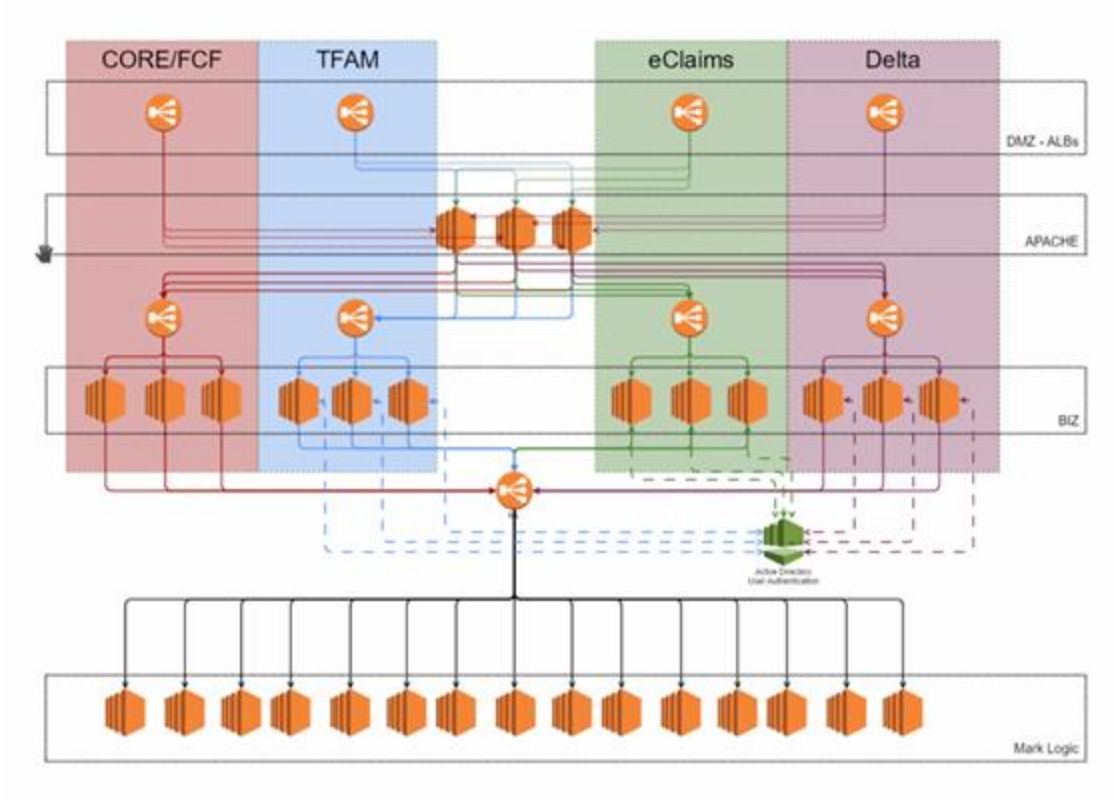
Lessons learned

# Technical exploration

This section covers the following areas:

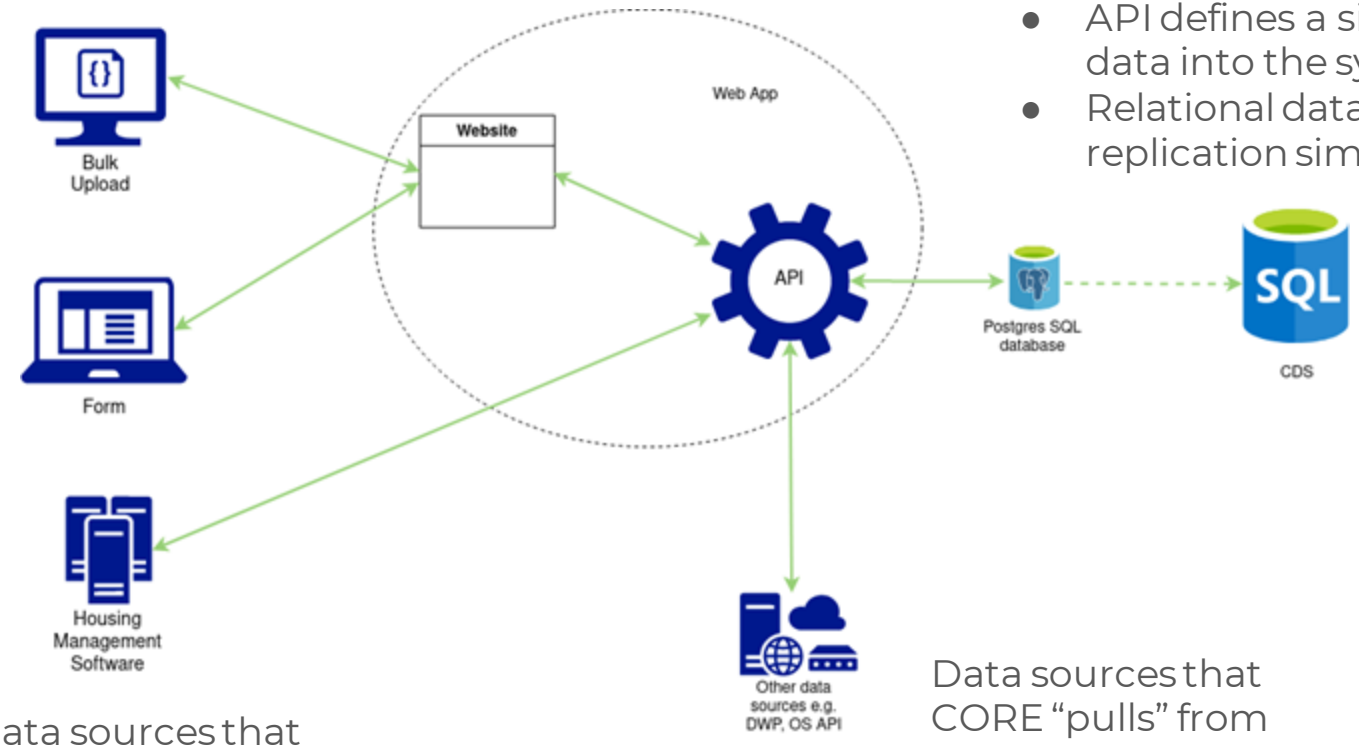
1. Architecture diagrams (current and proposed)
2. Technical spikes
3. Prototype approach
4. Security considerations
5. Alternative solutions
6. Data automation best practices

# Current technical landscape




MadeTech

# Proposed technical architecture




- API defines a single point of entry for data into the system
- Relational data store makes data replication simpler

Data sources that “push” data to CORE

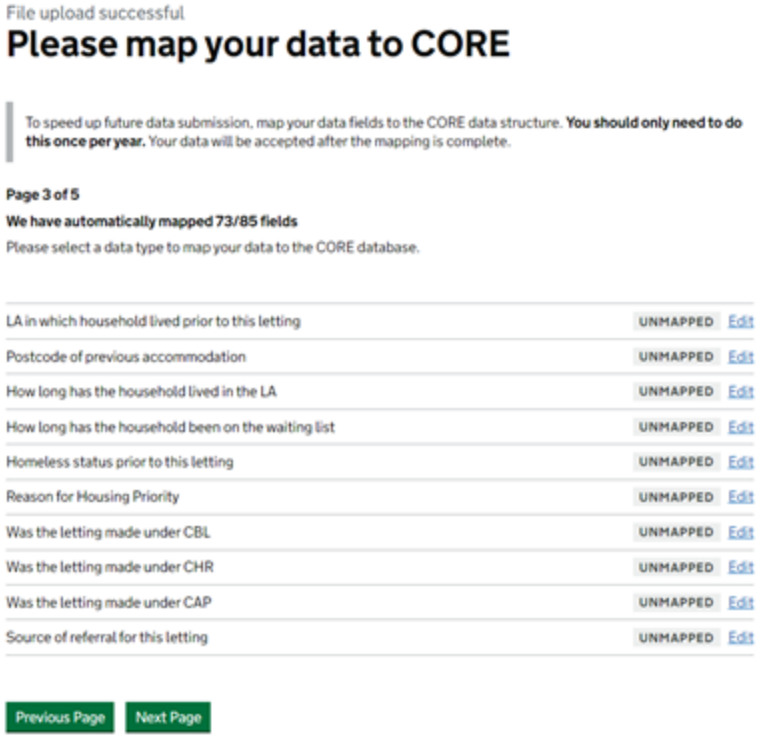
Data sources that CORE “pulls” from

MadeTech

# Technical risks and feasibility

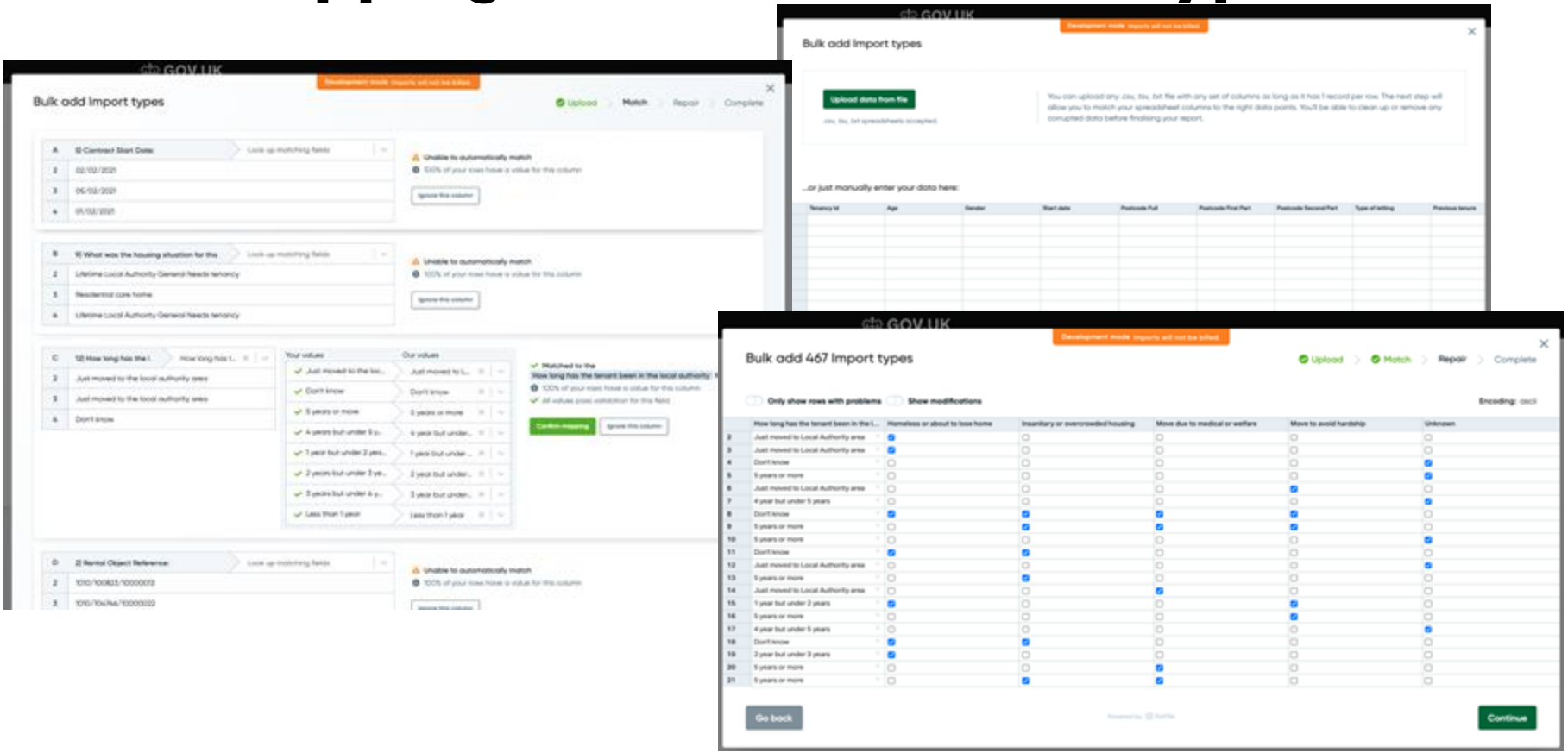
Two of the prototypes we built looked to explore the technical feasibility of data mapping in addition to the usefulness of such functionality to users.

**Prototype One - Data Mapping**




MadeTech

# Data Mapping with Flatfile - Prototype Three



Technical spike #1

# Parsing PDFs

| What we tried | What we learned | Actions |
| --- | --- | --- |
| Some organisations use electronic forms in PDF format to record data and then manually upload that data.<br><br>We tried to see how feasible it would be to parse the PDFs and recover the data automatically | This was not really feasible, since recovering data was reliant on the PDFs being created in a specific way which could not be guaranteed. In addition it was also not guaranteed that the data had field names attached. | No further action required - parsing PDFs reliably looks like a non-starter. |

Link to Github repository

Technical spike #2

# Data Mapping

| What we tried | What we learned | Actions |
| --- | --- | --- |
| This prototype looked at whether it may be possible to automatically map data from various users into CORE format. This comprised of a backend service which would accept CSV or XLS documents, parse the headers, and then attempt to match to CORE fields. | The prototype proved that we can do some automatic mapping with a manual override. It looked like there was complexity still to be solved in combining or splitting data such as dates or postcodes.<br><br>However, the users did not really understand the concept of “mapping” | Iterate on this to make what is happening easier to understand<br><br>Make it more obvious the benefits of doing the mapping here rather than in their own systems<br><br>Link to Github repository |

Technical spike #3

# Data Mapping with Flatfile

| What we tried | What we learned | Actions |
| --- | --- | --- |
| This was a second look at the mapping exercise in Prototype One, and looked at using the Flatfile tool to do the mapping from user data to CORE data. | We proved in this prototype that a more user friendly system could be made that could also deal with splitting or joining data such as dates or postcodes. | Still further content work required to assist in getting users to understand the concepts. Perhaps a video tutorial. |

Link to Github repository

# Security considerations

Security is baked into modern web application frameworks and as such was not something we tested in the prototype for a number of reasons: it is a solved problem if we use the frameworks and associated libraries correctly; the prototypes were built using tech that we would not necessarily use in a real application.

However, for a real application we should be certain to use framework features such as the following to avoid potential security concerns:

- ORM framework for manipulating databases (along with SQL Injection protection).
- Authentication frameworks/libraries to take advantage of well tested auth code.
- Platform level security and isolation, such as firewalls, load balancers and role based security, DoS protection etc
- Token based authentication for inter-application communication
- General data sanitization practices

# What else is available: complete solutions

A summary of any investigation into preexisting solutions, other public sector agencies & proprietary software options.

| Solution | Pros | Cons | Summary |
|---|---|---|---|
| Delta | Already in production and being used for other departmental data collection services | Generic service not designed for GDS service guidelines. Still requires development work to enable all the existing CORE functionality. Historically slow to change and inflexible. | Existing data collection system based on Orbeon Forms and MarkLogic |
| Existing CORE service | Already running in production | Concerns about code quality and service stability. Low confidence that additional features can be safely developed as-is | The existing service based on MarkLogic |

# What else is available: partial solutions

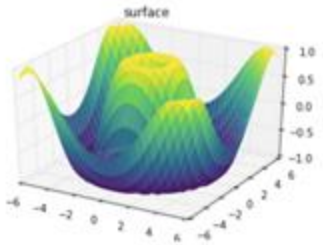
A summary of any investigation into preexisting solutions, other public sector agencies & proprietary software options.

| Solution Option | Pros | Cons | Summary | Output |
|---|---|---|---|---|
| Flatfile | Huge, highly functional and well backed software | No ownership (square peg, round hole)<br>Licencing fees<br>Accessibility issues | Flatfile allows you to create a customer data import experience with customised, inline validation. | View prototype |

# What does great data automation look like?

*MoJ Reproducible Analytics Platform (RAP):*

- Provides 300 analysts with best-in-class data science tools
- Real-time data modelling and analysis
- Makes it easier to tackle complex problems
- Reduces time consuming parts of reporting
- Processing of almost unlimited data sets at low cost becomes possible
- Analysis can be reproduced quickly and easily, with high quality assurance
- Visualisations can be made available quickly and easily, enabling compelling storytelling




Note that though the grid of values for a surface plot needs to be two-dimensional, it need not be rectilinear. Here is an example of creating a partial polar grid, which when used with the `surface3D` plot can give us a slice into the function we're visualizing:

In [10]:
```
r = np.linspace(0, 6, 20)
theta = np.linspace(-0.9 * np.pi, 0.8 * np.pi, 40)
r, theta = np.meshgrid(r, theta)

X = r * np.sin(theta)
Y = r * np.cos(theta)
Z = f(X, Y)

ax = plt.axes(projection='3d')
ax.plot_surface(X, Y, Z, rstride=1, cstride=1,
                cmap='viridis', edgecolor='none');
```

Sources:

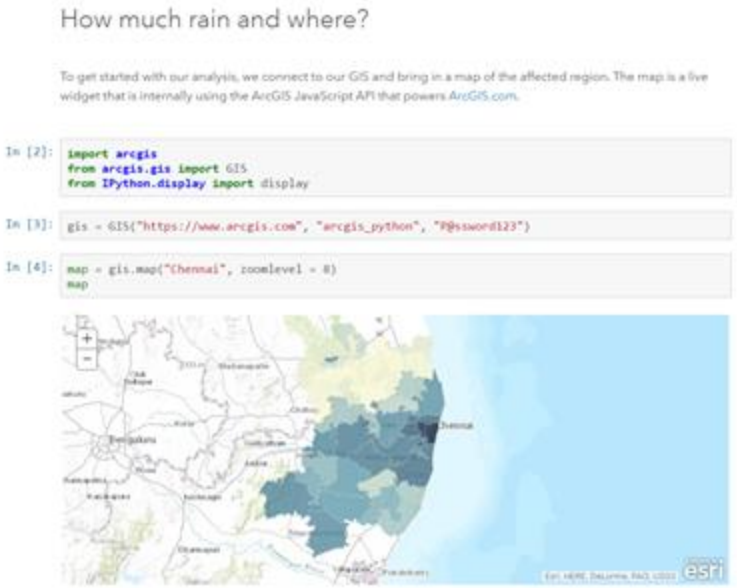
https://www.gov.uk/guidance/choose-tools-and-infrastructure-to-make-better-use-of-your-data
https://mojdigital.blog.gov.uk/2018/04/05/pushing-the-boundaries-of-data-science-with-the-moj-analytical-platform/

MadeTech

# General Best Practices

### Borrowing from Software Engineering

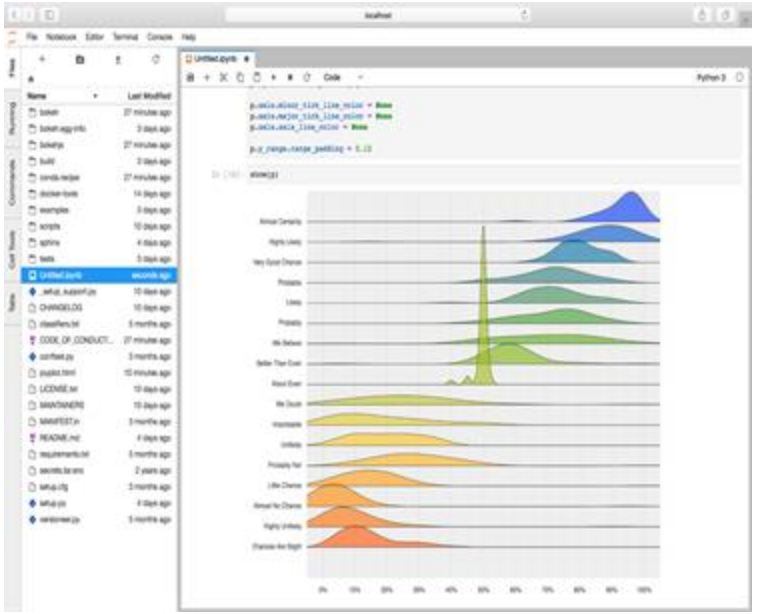
- Modern, open source tooling like Python, Pandas, Jupyter, R, Spark
- Common data formats like Parquet, SQL, CSV, JSON
- Data standards
- Reproducible, version controlled (Git) models, pipelines, code
- Agile development methods




MadeTech

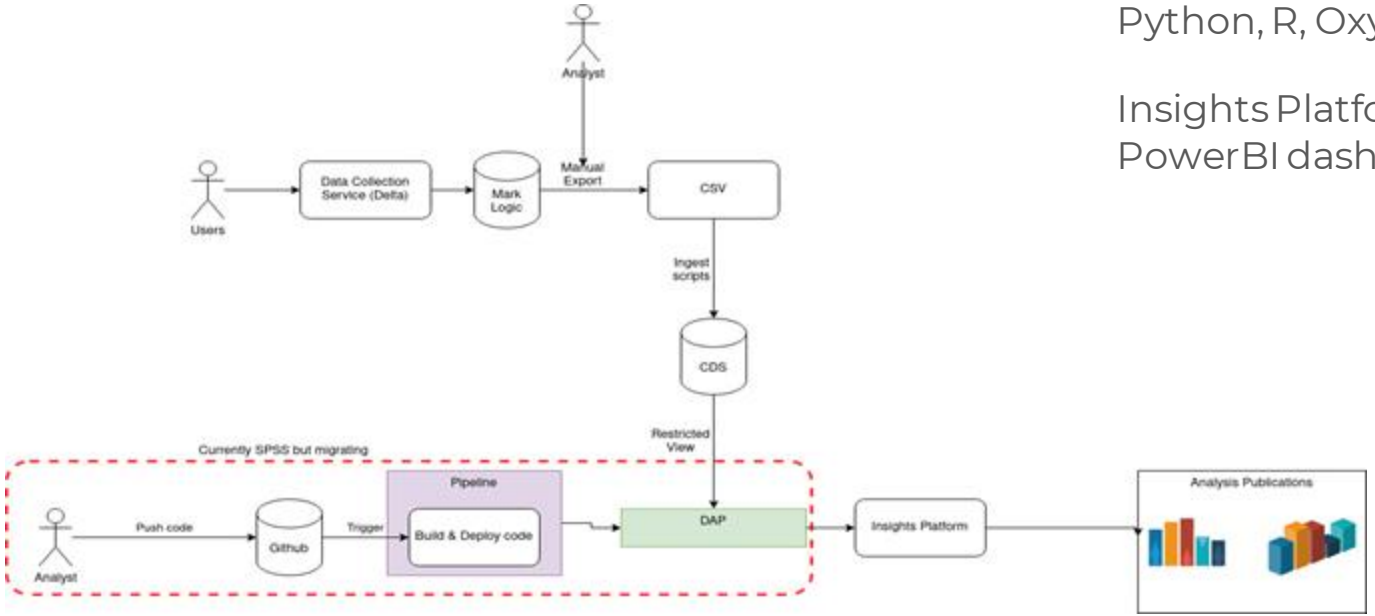
# Benefits of an advanced data automation approach

- Reduces vendor lock-in
- Cheaper & easier to scale to larger data sets & teams
- Faster/real-time reporting and visualisations
- Reproducible & auditable analysis
- Ability to make data more accessible to the wider public
- Opportunities to expand into advanced capabilities like AI and Machine Learning
- Automated testing to catch errors & regressions



MadeTech

# DAP, CDS & The Insights Platform

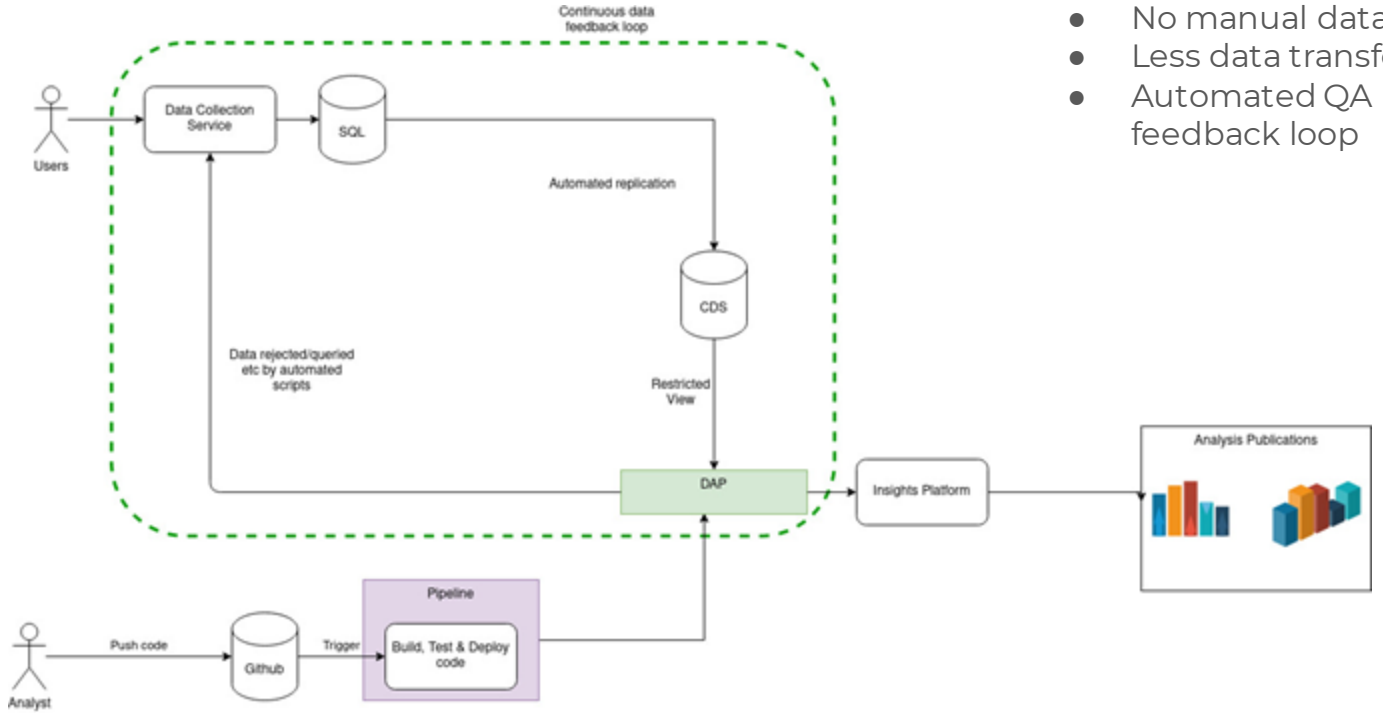



DAP provides a remote workspace with tools like Python, R, Oxygen etc

Insights Platform provides PowerBI dashboarding

MadeTech

# Faster feedback, better data, superior insights




- No manual data export/import
- Less data transformation & cleaning
- Automated QA creates a better feedback loop

MadeTech

# Summary of findings

**Research and analysis**

**Prototype iteration**

**Technical exploration**

**Lessons learned**

# Challenges from the Alpha

The team faced a lot of challenges throughout the Alpha, some of which have fed into our recommendations. Broadly, these can be categorised into four key areas:

## Vision

It was a challenge for the team to clearly define the service vision; particularly around whether is should be CORE centric or focus on a wider data strategy.

Understanding the definition of success with ADD & digital earlier may have helped prioritisation.

## Context

The service context is complex with many organisations & data providers who have multiple data collection & submission methods.

The Discovery uncovered some of these, but it was a challenge to internalise the findings at times.

## Research

The multitude of user groups & organisation types meant each research round had to be carefully designed to generate insights.

Recruitment research proved a challenge, as the research target groups were often niche.

## Team

The team was large, and had a few personnel changes which meant it was hard to maintain shared understanding.

At times it wasn't clear where decisions sat, and the Product Manager was not always empowered to enact their vision.

# Outstanding risks & assumptions

| ID | Description | Severity | Probability | Mitigation |
|---|---|---|---|---|
| 1 | Process changes for data providers may lead to delays, and increased payment requests for MHCLG | Major | Possible | • Test service & private beta with as many users as possible to ensure ease of use<br>• Invite users to access beta<br>• Keep previous service running until move to live |
| 2 | Running the old & the new service in parallel may lead to data discrepancies | Major | Possible | • Plan the roll out strategy as part of the Beta<br>• Test simultaneous service use with real users |
| 3 | MHCLG may not have a civil servant product team who can support the service in live | Minor | Probable | • Tender for DOS or GCloud contractors if required<br>• Select batteries included technology that avoids vendor lock in |
| 4 | The service becomes too CORE focused and limits the value delivered | Critical | Possible | • Perform a spike on H-CLIC to understand commonalities<br>• Perform an audit on case level data collection across MHCLG |
| 5 | The existing CORE service could reach end of life before the Beta is completed | Critical | Unlikely | • Existing incumbent supplier continue to support CORE service<br>• Beta to be launched before April '22 stats collection release |

| ID | Description | Severity | Probability | Mitigation |
|---|---|---|---|---|
| 6 | Scope is not achieved within agreed time frame or budget | Critical | Possible | • Prioritise features based on user needs<br>• Build 'just enough' for each feature for MVP<br>• Continually prioritise backlog to build upon an MVP |
| 7 | Budget runs out and the 'value unlocked' elements such as data playback, improved QA & derived variables are deprioritised | Critical | Possible | • Maintain sight that the vision is for better policy decisions; not better data collection.<br>• Build up a backlog of small improvements for data, and continually iterate<br>• Timebox initial QA & derived variables build |
| 8 | HMS Vendors may not integrate with API in time for 2022 stats collection | Minor | Probable | • Use contractors if required<br>• Select batteries included technology that avoids vendor lock in |
| 9 | Service may not solve a whole users need | Critical | Unlikely | • Continue to test with real users & perform research during Beta<br>• Release a private Beta |
| 10 | Internal MHCLG stakeholders disagree on service needs & goals, which lead to delays in decisions | Major | Possible | • Involve ADD & Digital stakeholders in the team for testing<br>• Ensure the Product Manager is empowered to make decisions<br>• Use the measurement framework & user needs to inform decisions |
| 11 | We do not improve workflow for users because they still need to collate data outside of CORE from multiple systems | Major | Probable | • Prototype & test the collaboration on a case service during Beta<br>• Audit existing data storage to understand the systems users collate from<br>• Allow partial uploads for Bulk & API data submissions |
| 12 | Data quality is not improved, or gets worse due to changes | Critical | Possible | • Compare & contrast submitted during private Beta & subsequent tests<br>• Codesign questions with policy, ADD & users |

# Outstanding risks & assumptions

| | Unlikely | Possible | Probable |
|---|---|---|---|
| **Minor** | Low<br>Accept the risk, routine management | Low<br>Accept the risk, routine management | Medium<br>Assign owner, review monthly<br><br>*Risk IDs: 3, 8* |
| **Major** | Low<br>Accept the risk, routine management | Medium<br>Assign owner, review monthly<br><br>*Risk IDs: 1, 2, 10* | High<br>Assign owner, review weekly<br><br>*Risk IDs: 11* |
| **Critical** | Medium<br>Assign owner, review monthly<br><br>*Risk IDs: 5, 9* | High<br>Assign owner, review weekly<br><br>*Risk IDs: 4, 6, 7, 12* | High<br>Assign owner, review weekly |

1. The problem to solve
2. Alpha approach & activities
3. Summary of findings
**4. Recommended next steps**

# Beta - case for change

The Alpha has demonstrated that it is possible to solve the users need to submit the right data at the right time, in the way that best suits them. Therefore we recommend proceeding to Beta, because:

- The current software is coming to end of life
- There are many accessibility issues and pain points for users. Redesigning the questions, UI & content will improve the experience for users, and may improve data quality.
- There are many improvements that can be made to bulk upload which will increase how many users can use it, over single case forms
- Increasing automation via bulk upload or an API will reduce keyboard time and human error
- Providing flexibility in data submission will reduce the need for data collation
- The learnings from this service can be applied to other data sets, particularly for other case level data collections.

# Options for Beta

There are three possible solutions MHCLG can proceed with in Beta

*Option 1*
**Bespoke case-level data collection service**

This would be a service dedicated to delivering on the hypotheses proven / tested during the Alpha and maintained by an MHCLG team

**Recommended Option**

*Option 2*
**Build CORE into Delta**

This option involves writing CORE in Delta, a system for data collection maintained by a third party vendor and used elsewhere in MHCLG. MHCLG could either compromise on their requirements, or wait for additional functionality.

*Option 3*
**Build CORE into Delta without using MarkLogic**

This is essentially the same as Option #2, but additionally replacing MarkLogic with a cheaper data storage solution. The replacement data storage is currently unknown.

Beta option #1

Recommended Approach

# Bespoke case-level data collection service

This would be a service dedicated to delivering on the hypotheses proven / tested during the Alpha and maintained by an MHCLG team

**Pros**

- Extensibility built into the new architecture
- Better tech stack, code quality, TDD & security
- Sustainable & quicker to develop
- New team; more innovation, working with ADD, less vendor lock in
- Improved benefits to data providers through playback & data management
- User centred; tested iteratively to meet a whole users need, better usability
- Lower ongoing cost
  At least £150k could be saved per year by moving to a modern solution

**Cons**

- Longer time to launch if compared to low featured CORE in Delta
- Slightly higher set up cost
- Change management to move collections into new solution *or*
- Cost of running two systems
- No UI driven changes

Beta option #1

Recommended Approach

# Bespoke case-level data collection service

This would be a service dedicated to delivering on the hypotheses proven / tested during the Alpha and maintained by an MHCLG team

**Risks**

- Increased change management within MHCLG & data providers
- More uncertainty in timescales than a known solution

**Recommendations**

- Dynamic “batteries included” programming language - Python/Django or Ruby/Rails
- Containerised (Docker) solution
- “Per-usage” cloud provider based
- Use “infrastructure-as-code” approach
- Use a relational SQL database such as Postgres
- Follow the GDS API standards (HTTP, RESTful)
- Architected in a modular way to optimise for being agile & adaptable

Beta option #2 & #3

# Delta

## Overview of the Delta service

- Generic data collection system
- Orbeon Forms + MarkLogic
- Form Builder UI
- Schematron based validations
- NoSQL (XML based) document storage that needs exporting to CSV to be ingested into the CDS
- Active Directory based user management

Delta is MHCLG's existing service for data collection. It's a central generic service using a proprietary data store and pro-licensed version of Orbeon forms. Most of the business logic currently sits inside the MarkLogic layer.

Beta option #2

# Build CORE into Delta

This option involves writing CORE in Delta, a system for data collection maintained by a third party vendor and used elsewhere in MHCLG

**Pros**

- If CORE compromise on their requirements, would be quicker
- Form Builder
- Known tool with in house experience

**Cons**

- Maintenance costs
- Third party vendor
- Further tied into MarkLogic license
- Slow iteration - unlikely to be able to build all functionality required
- Existing backlog & other DELTA users will further limit speed of development
- Unable to offer providers value or alleviate pain points, including accessibility issues.
- No user or organisational hierarchy
- Shared infrastructure creates single point of failure

Beta option #2

# Build CORE into Delta

This option involves writing CORE in Delta, a system for data collection maintained by a third party vendor and used elsewhere in MHCLG

**Risks**

- Development approach may not include full automated testing or monitoring
- Tightly coupled to other systems, which has led to downtime previously
- API functionality not yet implemented, external dependency is rate limiting
- May cost the same based on the sprint estimates, but with less functionality

**Recommendations**

- Adopt agile, test driven ways of working
- Increase test coverage of the system
- Add load test to ensure adding CORE will not cause performance problems
- Refactor the most tech debt heavy parts of the system. Look at the areas that have been hardest to change, most error prone or have seen the most code churn
- Consider extracting business logic from MarkLogic where possible to prevent further vendor lock in

Beta option #3

# Build CORE into Delta without using MarkLogic

This is essentially the same as Option #2, but additionally replacing MarkLogic with a cheaper data storage solution.

**Pros**
- UI powered form builder
- Known tool with in house experience
- Not tied into MarkLogic

**Cons**
- Initial expense & time of replacing MarkLogic likely to be more expensive than building a new solution
- Replacement carried out by the vendor
- This replacement will limit other development, meaning CORE is unlikely to have all the functionality required
- Existing backlog & other DELTA users will compete for priority
- Unable to offer providers value or alleviate pain points

Beta option #3

# Build CORE into Delta without using MarkLogic

This is essentially the same as Option #2, but additionally replacing MarkLogic with a cheaper data storage solution.

**Risks**

- The new data store may not be significantly cheaper
- Replacement currently unknown & may not be defined until Delta service design project completed
- API functionality not yet implemented, external dependency is rate limiting
- Likely to cost more & take longer than a new service

**Recommendations**

- Adopt agile, test driven ways of working
- Adapti data collection to allow for more efficient storage.
- Reduce the cost of data storage by using open source solutions (Postgres/MySQL for relational or MongoDB for NoSQL)
- Choose an SQL based database to enable easier more efficient replication to CDS
- Consider migrating other services away from MarkLogic one at a time until it can be removed entirely

# Recommendations

**User centred policy design - bring policy closer to delivery**

- Involve Policy SMEs in service development
- Data, digital & policy work to build better outcomes for users
- Collectively design data collection and adjust based on outcomes

For the CLDC Beta, we would recommend closer working with Policy SMEs

**Prioritise reducing burden to frontline staff, so they can deliver better services to citizens**

- Continue prioritising the reduction of user burden
- Focus on delivering increased value for all actors
- This will lead to increased compliance and better citizen facing services

https://mojdigital.blog.gov.uk/2019/02/27/what-has-user-centred-policy-design-achieved-in-the-moj/
https://dataingovernment.blog.gov.uk/2020/03/24/rappers-delight/

# Recommendations

**Improved engineering practices**

- TDD
- Unit, functional, load
- Automated pipelines for testing & deploying code
- Reproducible artefacts
- Documentation
- Monitoring & Alerting
- Optimise for agility & productivity

**Confidence & Agility**

*Having good automated testing at all levels, automated pipelines for reproducibly building and deploying your code and good monitoring for your infrastructure enables you to make changes quickly and with confidence. It ensures a reliable service. Choosing open source tooling and requiring thorough documentation ensures that there is no vendor lock in - anybody can pick it up and work on it.*

# Recommendations

**Share context & constraints, then empower teams to make decisions**

- Reinforce and align on shared goals between Digital, Data and Policy
- Put users (local government and the citizen) at the centre
- Empowered and aligned Service Owners and Product Managers
- Guided by user needs & department strategies.

**Embrace Agile & Lean approaches to service design & product development**

- Maximise the work not done
- Take an iterative approach
- Test with real users
- Work in the open

Incorporate learnings back into the development cycle, and expect that priorities might change.

https://mhclgdigital.blog.gov.uk/2019/09/16/defining-the-service-owner-role-for-mhclg/

# Recommendations

## Data Collection & Ingestion

- Single “gateway” and shape for data going into the system
- API submission where possible
- Minimise unnecessary data transformations by storing in the form it’s actually needed in
- Review feasibility of data being asked for


## Tidy Data

*Having a single point of entry to the system that all data has to pass through in a certain shape, makes it easy to singularly define validations and derivations that are easy to test, reason about and audit*

*Minimising manual data entry errors and reviewing the feasibility of what is being asked for (and finding alternative sources where feasibility is not high) will help ensure the data collected is high quality*

*Choosing the right storage model for the data helps to reduce the amount of “data cleaning” work needed to analyse it, resulting in less error prone processes.*

# Recommendations

## Open Data as a Principle

- Go beyond publishing raw data
- Publish the team's domain expertise
- Automated, real-time QA, reporting and analysis of data & insights

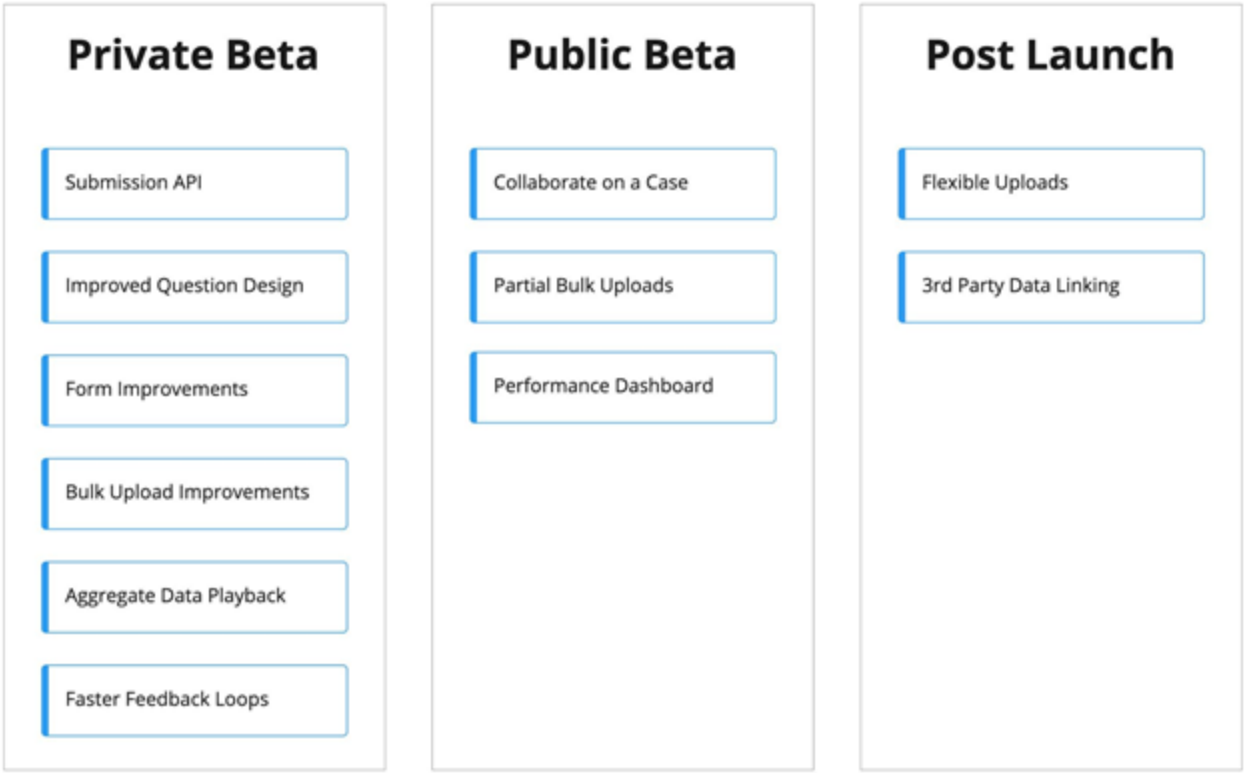
## Lay the foundation for leading on data science

*The value of the social housing data collection service lies in the domain expertise, analysis and insights produced from the data. The more we can codify, automate and publish that, the more they can push forward into new areas of analysis.*

*The more we can automate QA and action it in real-time, the better data we will have.*

MadeTech

# New Service Roadmap




MadeTech

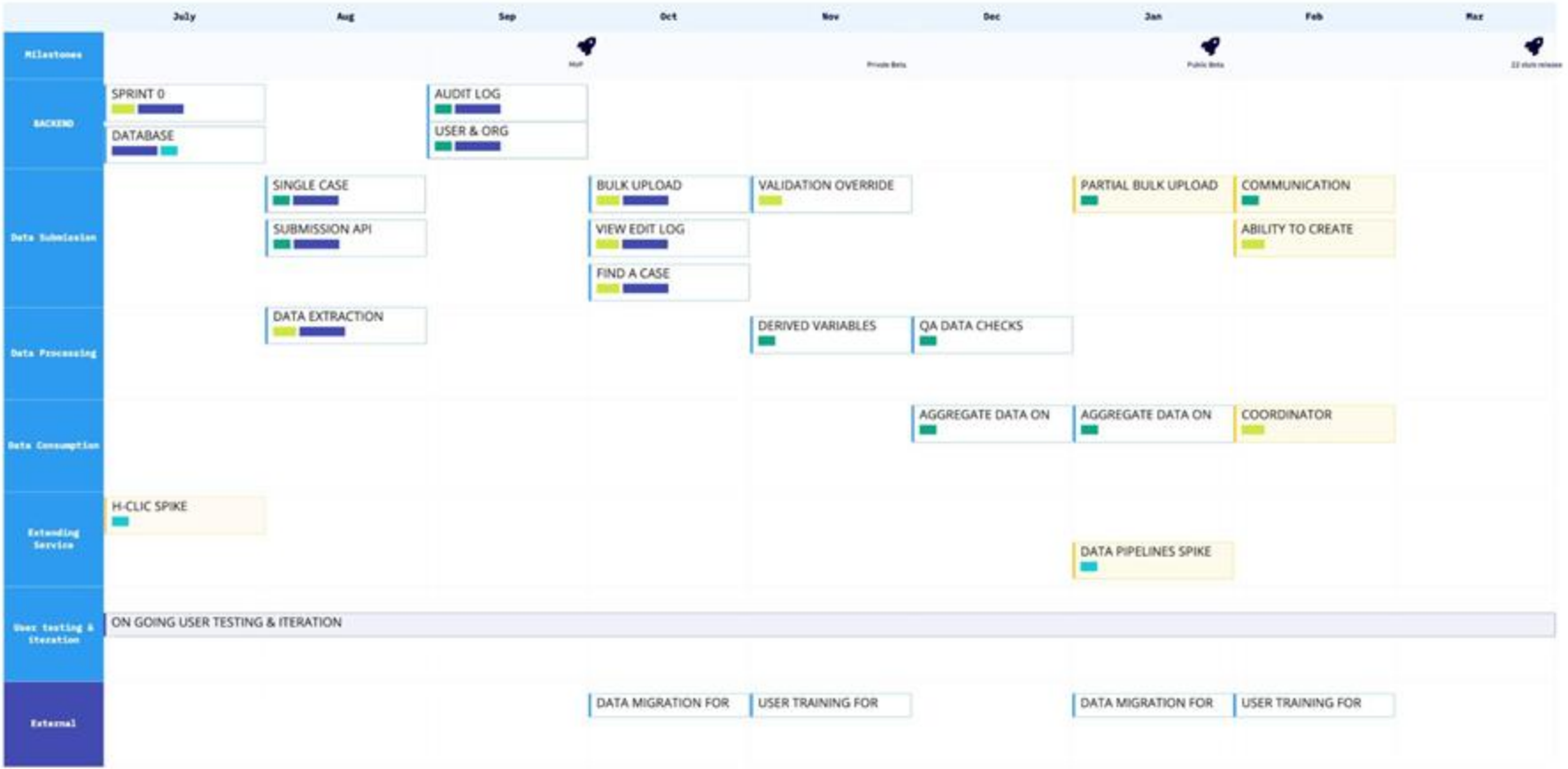
Key

- NOW
- NEXT
- LATER

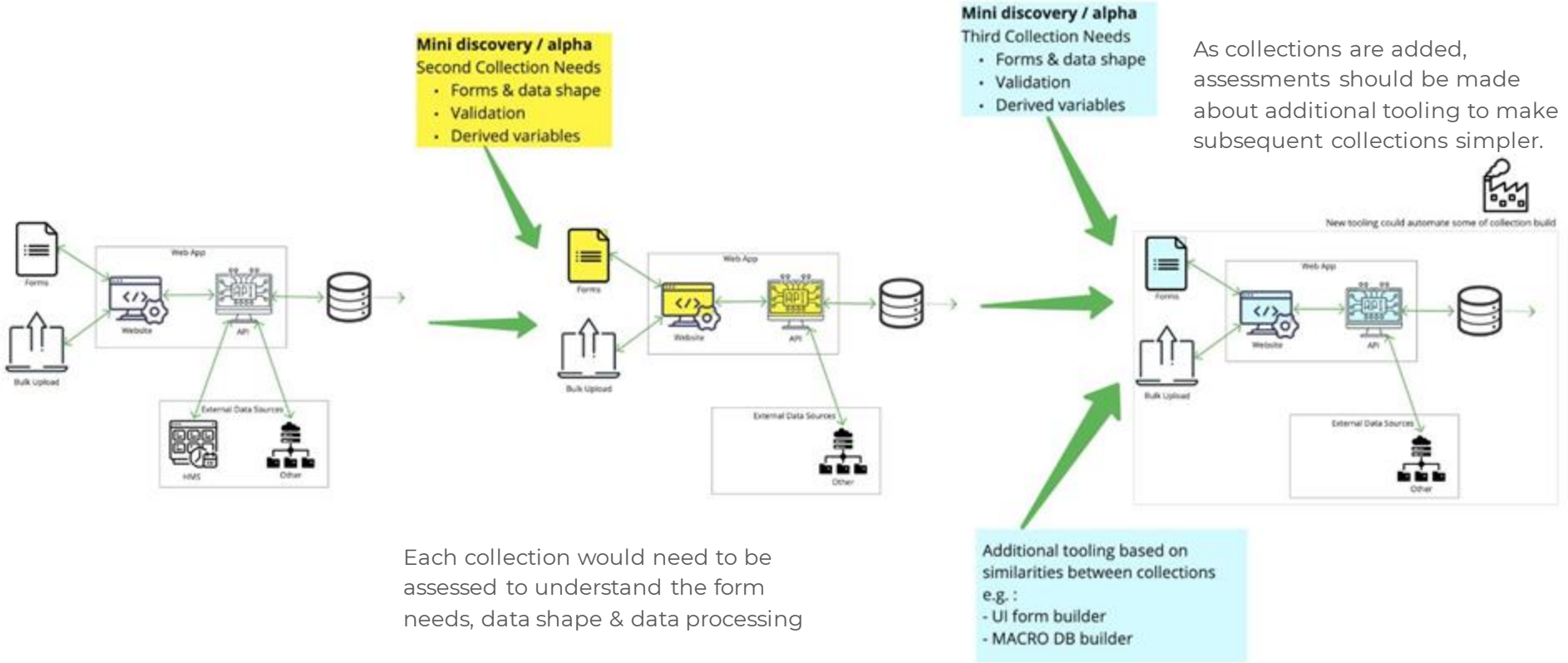
To Schedule

- 3RD PARTY DATA
- FLEXIBLE UPLOADS
- CBL INTEGRATION

[View detailed roadmap]



# Creating additional services



# Short discovery for each collection answering

Does the service have broadly similar requirements?

- User management
- Data pipeline
- Collection methods
- How often does the collection change?
- Is the collection repeatable?

What is the data shape for the new service?

- Questions & data shape
- Validations & derived variables

How can the data better inform policy?

- Work with a Policy SME to understand intent
- Identify data that the department already has
- Identify opportunities to link data sets internally & externally
- Design the questions holistically

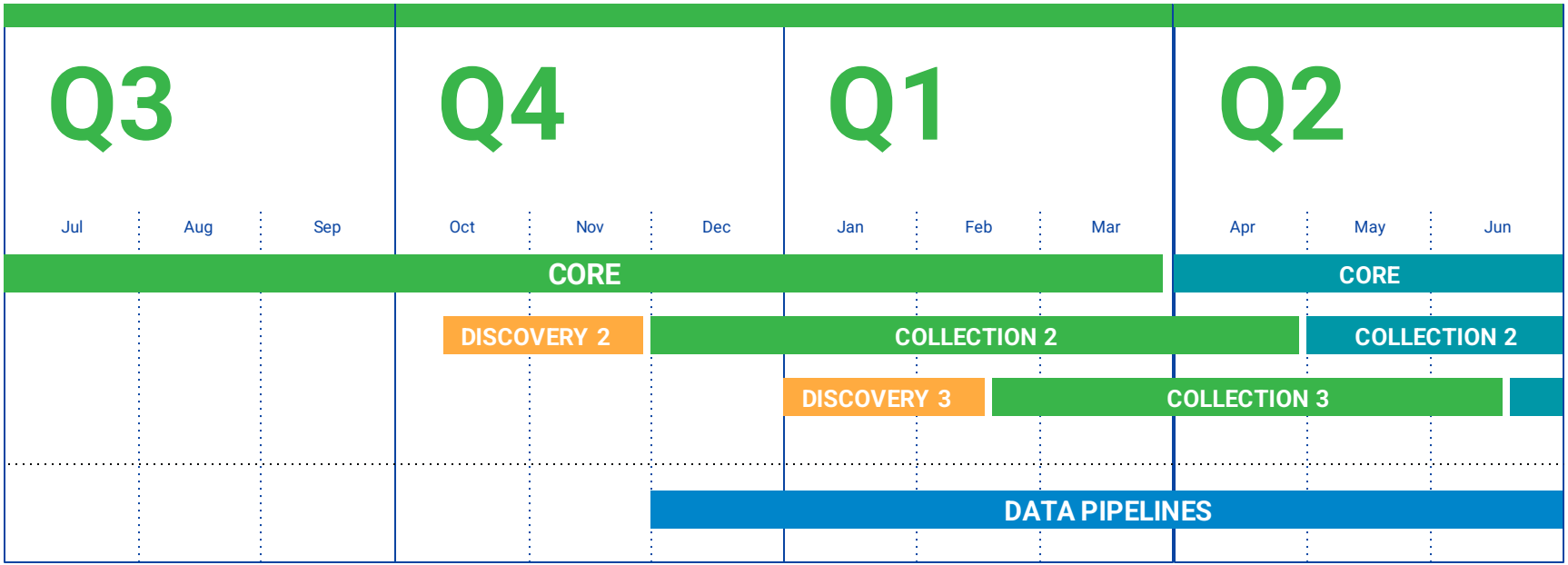

**Collection Migration Team***

*If building in parallel

MadeTech

# Unlocking the benefits of extensibility

MHCLG could invest in additional teams to run in parallel to assess other collections and then use the CORE service as a whitelabel template to shorten the development cycles




Key | Discovery / Alpha | Beta | CI / Live

MadeTech

# Indicative team costs for beta

The proposed costs are based on the complete team in place from 9th August 2021 through to 31st March 2022.

During this phase, we would recommend MHCLG hire an engineering product team that we can upskill & onboard to make way for a transition to Live.

**Beta Proposal Figures**

| Role | Discounted Rate | Days | Budget |
|---|---|---|---|
| Delivery Principal | £906.50 | 31 | £28,102 |
| Delivery Manager | £637.00 | 144 | £91,728 |
| Technical Lead | £760.00 | 144 | £109,440 |
| Senior Engineer x 2 | £637.00 | 288 | £183,456 |
| Front End Engineer | £560.00 | 144 | £80,640 |
| Design Lead | £906.50 | 31 | £28,102 |
| Interaction Designer | £637.00 | 144 | £91,728 |
| Content Designer | £546.00 | 144 | £78,624 |
| | | **Total** | **£691,819.00** |

# Thank you.

## Any questions, please get in touch

Madetech.com
hello@madetech.com
@madetech

# Appendix

Feature Impact

| Org Type | MHCLG Strategy | Functionality |
| --- | --- | --- |
| **Type 1 - 17%**<br>• Bulk upload (+ some manual)<br>• Proprietary tool | MHCLG should aim to automate the process for this group, to remove the need for a manual bulk upload. By adding an API and integrating with known HMSs we can get data quicker, and reduce the feedback loop on the validation errors - allowing them to be fixed within the HMS.<br><br>This may also encourage a higher uptake on CORE modules, which will further reduce the burden of collating data in multiple systems, and increase the accuracy of the data overall. | Submission API (with amendments possible)<br>Derived variable<br>3rd party data linking (potentially including own systems) |
| **Type 2 - 43-60%**<br>• Single log<br>• Proprietary tool | This group is the largest user group, and it's also the group the the most challenging process - manual form uploads, with multiple data sources for collation.<br><br>The primary goal here should be to encourage automation, by making it a lot simpler to use a bulk upload functionality (or API). We will know this happens when more people tell us they use bulk upload, or have selected a CORE module in their HMS.<br><br>The secondary goal should be to make improvements to the form, to reduce the burden on users. We will know this has happened when the manual entry speed decreases, and quality increases. | Flexible uploads<br>Bulk upload<br>Single case form submission & validation<br>Choice based lettings integration |
| **Type 3 - 20-30%**<br>• Single log<br>• Own solution | This group are unlikely to use a proprietary HMS, or build their own, due to the small number of logs. However this small number also means every form submission is likely to be even more burdensome.<br><br>To reduce this burden MHCLG should improve the form to make it self explanatory with in form validation and guidance.<br><br>Ensuring LAs can collaborate on a case to support their smaller HAs with getting the right data will improve the quality. Getting as much data as possible from other sources will alspo be particularly beneficial to this group. | Org to Org mapping<br>Choice based lettings integration |
| **Type 4 - 10%**<br>• Bulk upload<br>• Own HM | The primary strategy for this group should be to improve the bulk upload process, to limit the number of manual interventions required and speed up the validation process.<br><br>The API will be available to this group, but uptake would require each LA to integrate themselves. MHCLG could provide funding via the LGDCG to support LAs who share their custom HMS with other LAs, to encourage further reuse and allow LAs to move away from proprietary tools. | Search for a case<br>Flexible uploads |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **Bulk upload prototype**<br>Prototyped bulk upload changes including field mapping.<br><br>Tested with seven users | To test the assumption that if we can make bulk data uploads easier, then providers will spend less time filling out forms | - Users who already knew about bulk upload understood the data mapping concept, but others needed more explanation.<br>- This demonstrated a need for clearer guidance around Bulk Upload.<br>- Users struggle to get all the information needed together into one place in order to achieve a bulk upload.<br>- Users expected and wanted to map data values in addition to fields.<br>- Users thought 'mapping' meant it would pre-populate some fields that were always the same for every CORE submission. | User Research<br>Research playback<br>Prototype live demo<br>Username: mhclg<br>Password: l3tm31n |
| **Flexible upload prototype**<br>Evolved the 'Bulk Upload' into 'Flexible upload' using flatfile.<br><br>Tested with XX users from XX profiles | Testing of the bulk upload prototype demonstrated that users wanted more flexibility, and the ability to map specific field values from their source systems to CORE fields & values. | There is genuine appetite for bulk upload (desirability) - It solves a real problem for users as a quicker mechanism for sharing data with MHCLG<br><br>But there are some viability and feasibility problems that need to be tested further (in beta) to make sure organisations are able to share information in this way. | User Research<br>Research playback<br>Prototype live demo<br>Username: mhclg<br>Password: l3tm31n |
| **Data playback prototype**<br>A prototype that displayed key statistics and KPIS.<br><br>Tested with seven users from a variety of organisation and user types. | To test the assumption that if providers have faster and easier access to the insights they value, then they will feel more motivated to provide accurate, timely data. | - A lot of interest in KPIs around validation issues & errors. This highlights what a painful part of the upload process this is.<br>- Opinions about what insights would be useful varied hugely between organisations and job role. The delay between submission and stats release was highlighted as a concern for the usefulness of this data<br>- There is at least one commercial company (HouseMark) that charges for these sort of insights, so we know there is a demand for these insights in some form | User Research<br>Research playback<br>Prototype live demo<br>Username: mhclg<br>Password: l3tm31n |
| **Provider API User Research**<br>Facilitated group discussion with 5 main HMS software suppliers | To understand commercial viability and technical feasibility of connecting to an API | - Not all CORE modules are paid for services.<br>- Some HMS would integrate with an API for free<br>- The concept was seen as technically viable, and was supported | Research plan<br>Data playback |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **Supplier API User Research**<br>We ran research with some of our largest data providers to talk about API submission | We want to know: If an API was provided through software suppliers, to what extent would it be adopted and would it solve problems that the CORE module doesn't? | - Users felt an API would save them time, reduce errors and streamline the process.<br>- Users felt the IT capability of their organisation could be a challenge in integrating with an API<br>- Not all information is stored in one HMS, we need to consider how a solution would account for this.<br>- Users have put in a lot of effort to make the process work for them and value the solutions they have come up with – a change process will be needed. | Research plan<br>Research playback |
| **Collaboration on a case**<br>We ran a series of service design workshops to design a service that would allow multiple users to work on a single case - considering the technical feasibility, alerts and notifications & ideal journey and architecture. | *"Users struggle to get all the information needed together into one place in order to achieve a bulk upload."*<br><br>We know that a number of different users in an organisation have to collaborate on a case in order to submit a complete log submission | - Developing this functionality could result in a service that's closer to a case management system than a data submission service<br>- It could be technically feasible to bring disparate data together from different submissions by different actors<br>- Allowing users the ability to collaborate can reduce the burden of submissions and collection of data | Service blueprint<br>User story map |
| **Matching disparate data sets**<br>Ran a spike to understand how data could be integrated API Integration, Bulk Upload (e.g. CSV / XML), Manual individual form submission | *"If data comes from a number of different sources, how can we bring that together into one coherent case?"* | - There is no one identifier common to all supplier systems that can be used to tie together all disparate data sets<br>- Manual matching is currently possible through a variety of fields<br>- Automated matching could be possible by letting users identify identifier fields or potentially by looking for common fields | Technical feasibility |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **Collaborate on a case - technical feasibility**<br>We conceptualized how a collaboration service would work | Different parts of the data are provided by different systems and sometimes at different times or even by different people. Currently they're forced to collaborate offline and have a single person upload. We wanted to visualise collaboration built in. | - There are multiple possible workflows depending on whether users are expecting to be able to edit a case independently or whether a case is handed over "relay-style"<br>- Technically similar to accepting data at multiple times but UI/UX is important | Service blueprint |
| **Naming the service**<br>We spoke to users to understand how they describe the service, and used this to inform potential names. | We think a name that more accurately reflects the purpose of the service will help users understand the value of the data they're providing better | - Good service names are verbs and describe a task not a technology or department<br>- Good service names use the words users use<br>- e.g. Share CORE social housing data | Sharepoint presentation<br><br>This needs further testing in Beta. |
| **Contextual help**<br>Reviewed the help pages & content, and made recommendations for bringing inline to the form | Users told us that it's hard to navigate between guidance and the forms. Bringing help into the form is in line with GDS guidelines. | - There are lots of additional guidance<br>- Help & guidance was particularly useful for questions where something is asked for that is very CORE specific | Confluence<br><br>This needs further testing in Beta. |
| **Content review**<br>Reviewed the existing forms & guidance | To provide recommendations for areas that content design could simplify the submission process | - There were 123 downloads on the CORE site, going back to 2013<br>- Although the forms showed 20 questions, the number of multipart questions made it significantly higher<br>- Some of the dropdowns had hundreds of responses to look through | Content audit |
| **Paper form redesign**<br>Redesigned the paper form using best practices | We believe that the paper form has been the basis of which our questions have been formatted. We wanted to explore how a "redesign" of the paper form would impact on the current | - Additional information is slipping into questions (framing and hints/tips)<br>- Tenants can refuse to provide a lot of information<br>- There are ways of presenting a question to aid in usability<br>- Grouping is key to allow users to avoid context shifting and slow | Paper form |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **Data linking DWP**<br>Met with DWP, and then wrote a pitch covering data protection strategy | To reduce the burden on local authorities to provide the same data to government multiple times | - DWP already shares this sort of data with other services and are enthusiastic about this approach<br>- There's a process in place starting with a pitch, a legal feasibility check and then a technical feasibility check | Confluence write up |
| **Data linking ONS project**<br>Reviewed data linking opportunities in a project that's run by ONS known as the Integrated Data Platform | To link with other Government Departments, ONS and Data Labs | - This will provide MHCLG analysts the opportunity to use linked data & shared analysis tools to make policy decisions<br>- The project is in it's infancy and unlikely that the tools will be available for a few years | Confluence write up |
| **Data linking Ordnance Survey API**<br>Reviewed the existing Ordnance Survey API to understand whether it could be used for address matching | To investigate being able to use UPRN as a property identifier so that we can store property details and help pre-fill those form sections without requiring local authorities to collect UPRN themselves | - An API is available that lets you search UPRNs by postcode<br>- The API is free for public sector organisations<br>- Address matching is difficult and likely to be error prone so we will not be able to rely on having an accurate UPRN 100% of the time | Confluence write up |
| **Data linking Choice Based Lettings**<br>Reviewed the CBL market, and compared questions asked by 4 example forms | To collect data from source when the tenant applies for social housing, rather than relying on a housing officer to do the work | - CBL forms are not standardised and ask different questions in different formats with varying amounts of CORE data<br>- There are 6 main CBL providers, which implies relatively few HMS to integrate with in order to collect this data - potentially worthwhile | Confluence write up |
| **Single Log File Creation**<br>Content design for questions, iterated via user testing | Users told us that they don't always understand what the questions mean. ADD told us that certain questions have a high error rate. | - The way we ask questions does not line up to the way housing officers use data, which is confusing<br>- Some of the questions are guessed, because they don't receive that information | Research playback<br>Confluence page |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **API feasibility & recommendations**<br>Reviewed the functional requirements for an API, and potential options for solving | To validate the feasibility of the proposed solution, and provide some technical options to take into Beta | - An API would be feasible & would solve a users needs<br>- We could allow partial uploads for the API<br>- We could validate & authorise as required<br>- A limitation would be that changing the questions would require code changes to the API | Confluence page |
| **Authorisation & access control**<br>A review of the existing org structure, and the potential tech choices to support these | To validate the feasibility of the proposed solution, and provide some technical options to take into Beta | - The Org structure allows parent & child orgs, where parent orgs can submit data on behalf of the child<br>- Schemes allow orgs to prefill some data on the forms<br>- There are different Roles with different permissions | Confluence page |
| **Consideration of frameworks**<br>Background, considerations & options for frameworks | To decide which framework to use in Beta | - It's important that the service works for everyone, including those with slower internet, or JavaScript turned off<br>- There are certain elements that add complexity to implementing a progressive web app e.g. API & Offline Storage | Confluence page |
| **Data field automapping**<br>Review of options & use of field automapping | To validate the feasibility & build the bulk upload prototype in sprint 2 | The algorithms were run over the "Hand matched" contrived data set. The TF-IDF replicated the human column mapping with ~59% accuracy, while the Fuzzy Matching approach replicated it with ~77% accuracy. These scores are relatively heavily influenced by the amount of missing columns we had and how many columns are "variations on wording" vs a totally different way of asking. ~60% overall matching seems appropriate as an estimate. | Matching function code<br>Confluence page<br>TF-IDF algorithm<br>Fuzzy Matching algorithm |
| **Data pipeline**<br>Review of the way data is processed by CORE currently | To understand the existing pipelines in order to inform the proposed solution & replacement | - Data enters via forms or bulk upload<br>- 300,000 records a year, ~120 data points / record<br>- Various levels of validation<br>- Data ends up in CORE & CDS | Confluence page |

| What we did | Why we did it | What we learnt | Outputs |
|---|---|---|---|
| **Preliminary system architecture**<br>Compared options for architecture based on key considerations for maintainability | To validate the feasibility of the proposed solution, and provide some technical options to take into Beta | - Product team will be 2 developers and a product manager<br>- System should be easily maintainable and adaptable given that team Common tool kits, languages, infra et<br>- Form changes happen annually. Optimizing this process is key<br>- There is likely to be 2 forms (sales & lettings), 1 API, 1 Bulk upload<br>- Additional forms and APIs may need in the future | Preliminary architecture<br>System components |
| **UI/Low Code form changes**<br>Considered the options for question changes | Changing the questions is a pain point in the existing system, so solving this is key in a new solution | 3 potential options compared<br>- Off-the-shelf enterprise UI software. i.e. Delta/Orbeon forms (low code solution)<br>- Custom UI driven approach<br>- Code driven improvements | Confluence page |
| **Validations, QA, modelling and reporting**<br>Investigated how validations and QA work and explored opportunities to automate them. | Aside from the general aspiration that we should aim to automate things to reduce cost per transaction in our service, we have a number of hypotheses around improving the data coming into the department and reduce the burden on data providers. | - Differences between validation & QA<br>- Hypothetical solution for this | Confluence page |
| **Delta consideration**<br>Various rounds of investigation & review of the options for Delta | To help with final recommendations for new service, and to answer the question about whether to put CORE into delta | DAP - Data Analysis Platform<br>Delta capabilities & roadmap<br>Delta Vs New Service Pros and Cons<br>Pros and Cons of using DELTA for Core and HCIC<br>Two distinct services or not? | |
| **Conversation with HouseMark**<br>Spoke to HouseMark to understand the service they offer to LAs | HouseMark present CORE data back to LAs for a fee, which implies viability for data playback. We wanted to understand this further. | They provide a benchmarking product - the model includes Cost, comparisons for value for money, performance voids, arrears repairs and turnaround. A lot of the value comes from comparison metrics | Confluence |

# User personas

## Housing Officer

### Bio

Y has been a housing officer for 3 years. Her day-to-day life involves visiting new tenants in their homes to complete the sign-up process on site, and completes follow-up admin back in the office. Every week she signs up on average 5 new tenants. Due to the pandemic, she is much more desk-based now, completing more sign-ups from her office.

### Goals

To get tenants into their new homes as smoothly as possible and make sure these tenants are well-supported.

### Digital confidence

Much of her work has been paper-based until recently when she has switched to filling in PDFs. She is used to transferring data from paper onto a housing management system, but is not confident using excel and needs support from her manager when the system doesn't work how she expects it to.

### Tasks

- Interviews tenant to ask CORE related questions
- Fill in CORE form on paper or directly onto website
- Searches for extra CORE data in housing management system, application forms, emails and other placecs
- Often completes CORE returns in batches of about 6

### Quotes

> CORE. Obviously something we've been told to do

> We just don't have certain datapoints in the system. We're happy to submit everything we have. Question 8 - I have some data but if I don't have this or I don't have time then it'll be refused.

> Once you get in you manipulate the rent to make it go through. You may lower the rent value so that it passes and put the rest in the service charge.

### Motivations


### Frustrations

She has to go into multiple different systems and application forms to find all the information required by CORE which makes it a lengthy process

She doesn't see the value of doing CORE and perceives it as an burdensome task in an already lengthy process.

She worries about the sensitivities of asking some questions to the tenants, for example around pregnancy and income.

She doesn't know the answer to a few questions or doesn't understand what is being asked

# User personas

## Lettings Manager (Data Coordinator)

**Bio**

X became a lettings manager for her housing association after being a housing officer for 10 years. She manages a team of 4 housing officers and is responsible for all Supported Housing lettings in her organisation.

**Goals**

To make sure that her team of housing officers complete sign ups following the correct process

**Digital confidence**

She is confident using the housing management system and other systems required in lettings and often supports the housing officers in using them. However, she is not very confident in using technical systems beyond this.

**Tasks**

- Checks that all CORE submissions have been completed
- Prompts housing officers to complete missing submissions
- Makes sure housing officers are set up on CORE correctly
- Deals with validation errors and coordinates team to correct them
- Responds to communications from MHCLG
- Sometimes fills in parts of CORE once a lettings officer has completed questions with tenants

**Quotes**

> By way of the information itself we don't particularly benefit from it

> We already have this information, it's a bit of the process that seems very manual

**Motivations**


**Frustrations**

Validation issues are pain to resolve

Has to prompt housing officers to complete CORE logs at the end of each month

What CORE requires sometimes doesn't match the specific circumstances of their properties/tenancies

CORE is a time-consuming process

# User personas

## Business Support Officer

**Bio**

Z has been supporting the New Build Lettings Team for 3 years. They also provide support to other teams in the social housing department and ensures the day to day running of systems used by staff. They are an expert in the CORE system, regularly submitting data by bulk. They occassionally input CORE forms themselves after being passed completed paper forms by housing officers.

**Goals**

To ensure the smooth running of systems and processes which enable housing officers to do their jobs.

**Digital confidence**

Adept at navigating various systems and solving problems within them. Works closely with the IT team to make changes to Housing Management Systems.

**Tasks**

- Extracts CORE data from systems and submits bulk uploads
- Sometimes inputs all CORE forms onto website from paper forms filled out at
- Works with IT team to make changes to HMS annually once question changes have been released
- Helps navigate the CORE website, resets passwords and carries out training

**Quotes**

> It takes a long time to complete validations, even though I know the data is correct, because it's our data.

**Motivations**


**Frustrations**

It takes a long time and coordination to update systems to reflect annual question changes in CORE

Housing Officers are prone to forgetting user names and passwords but cannot reset them themselves

Validation issues are pain to resolve

# User personas

## Housing Performance Manager

**Bio**

W is responsible for the performance of the social housing team and using data to gather insights and influence strategy at an organisational level. W comes from a business analyst background and is used to working with data.

**Goals**

To ensure the housing team is efficiently getting the right tenants into social housing and that he has the right information to make decisions about social housing strategy and spot arising problems.

**Digital confidence**

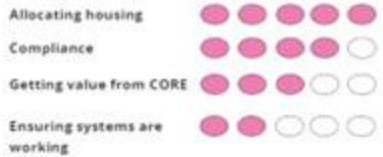
Adept at navigating various systems and solving problems within them. Has a background in data analysis.

**Tasks**

| | |
|---|---|
| Oversees the submission of data to CORE | Might be responsible for submitting bulk data to CORE |
| Deals with validation errors | Analyses housing data to gain insights and recommend changes to housing strategy |

**Quotes**

**Motivations**




**Frustrations**

Finds it difficult to use CORE data from their organisation and make meaningful decisions from it.

It takes a long time and coordination to update systems to reflect annual question changes in CORE

# User personas

## Housing Performance Manager

**Bio**

W is responsible for the performance of the social housing team and using data to gather insights and influence strategy at an organisational level. W comes from a business analyst background and is used to working with data.

**Goals**

To ensure the housing team is efficiently getting the right tenants into social housing and that he has the right information to make decisions about social housing strategy and spot arising problems.

**Digital confidence**

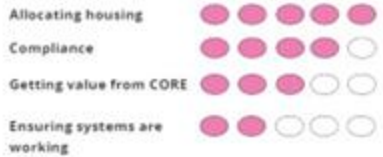
Adept at navigating various systems and solving problems within them. Has a background in data analysis.

**Tasks**

| | |
|---|---|
| Oversees the submission of data to CORE | Might be responsible for submitting bulk data to CORE |
| Deals with validation errors | Analyses housing data to gain insights and recommend changes to housing strategy |

**Quotes**

**Motivations**



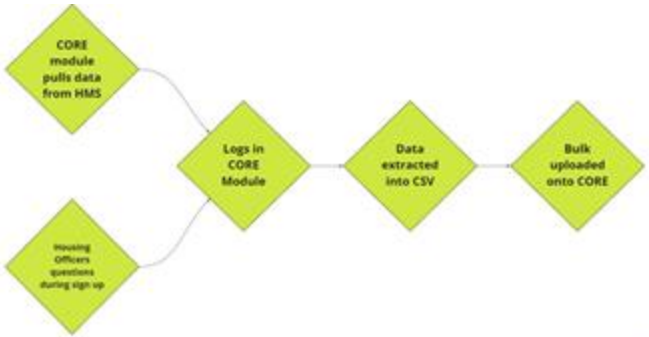

**Frustrations**

Finds it difficult to use CORE data from their organisation and make meaningful decisions from it.

It takes a long time and coordination to update systems to reflect annual question changes in CORE

# Org Type 1

## Supplier Reliant
## CORE Module Users




**Type 1**

Software Supplier | CORE module

Bulk upload ( Individual Form Submission )

17%

**Type of organisation:** Could be Local Authority or PRP
**Amount of Social Housing Stock:** Medium - Large
**Number of new lettings a year:** ~1800

**Digital Maturity level**
Reliant on a large software supplier.

Has staff to complete data collation and administration.

**Reliance on 3rd party software supplier**
A large supplier such as Northgate, Capita or Civica is used to run their Housing Management System. This system includes a 'CORE module' where CORE data is input and stored.

**How is data collected?**

- Data is automatically pulled into the CORE module from other parts of the Housing Management System.
- Housing Officers input extra questions straight into the CORE module (either during signup or afterwards)

**How is data collated?**

- Usually have a member of staff responsible for this e.g. a Data Systems Administrator or Performance Analyst.
- They click a button on the CORE module and it downloads relevant data.
- This is converted into a CSV file using the template the CORE website provides.

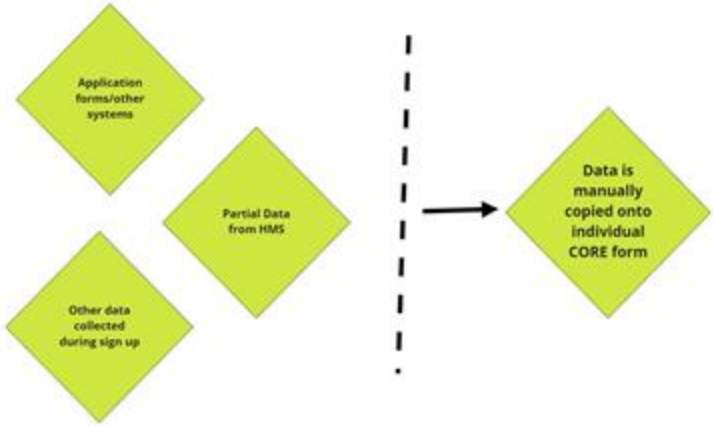
**How is data submitted?**
Mainly by bulk upload, but manual uploads are required when there is a lag in getting the CORE module updated.

**Pain points**

- Question changes by MHCLG means the CORE module needs to be updated by the supplier each year. This takes 3-4 months in which the CORE module cannot be used. This means they are always behind in their submissions.
- Dealing with a high number of validation errors (sometimes 5v0%) once the bulk upload has been submitted takes a huge amount of time.
- Collecting and submitting CORE is a heavy workload burden
- It's difficult to see the value of CORE, as they don't know what MHCLG uses the data for.

# Org Type 2

## Supplier Reliant Form Fillers




**Type 2**

Software Supplier

Individual Form Submission

42% - 60%

**Type of organisation:** Local authorities and other providers (small and large)
**Amount of Social Housing Stock:** All sizes
**Number of new lettings a year:** Any number

**Digital Maturity**
- Reliant on a software supplier to run their Housing Management System
- Uses spreadsheets and PDF forms for manual data entry.

**Reliance on 3rd party software supplier**
A supplier is used to run their Housing Management System. They may use a larger supplier such as Capita, or a smaller one. Information for housing applications and information relevant to their business processes (on the property, tenancy and tenant) are collected in their HMS. But there is no CORE-specific function in the system.

**How is data collected?**
- They may have developed their own PDF/paper CORE form making adjustments to the original to fit their own processes and terminology.
- Housing Officers and other staff input data directly onto these forms, either during the sign-up or afterwards, copying information from the HMS.
- OR Housing Officers may input data directly onto the CORE website.

**How is data collated?**
- An administrator uses the PDF forms and manually enters this information onto the CORE website for each individual log.
- Sometimes Housing Officers do collation themselves

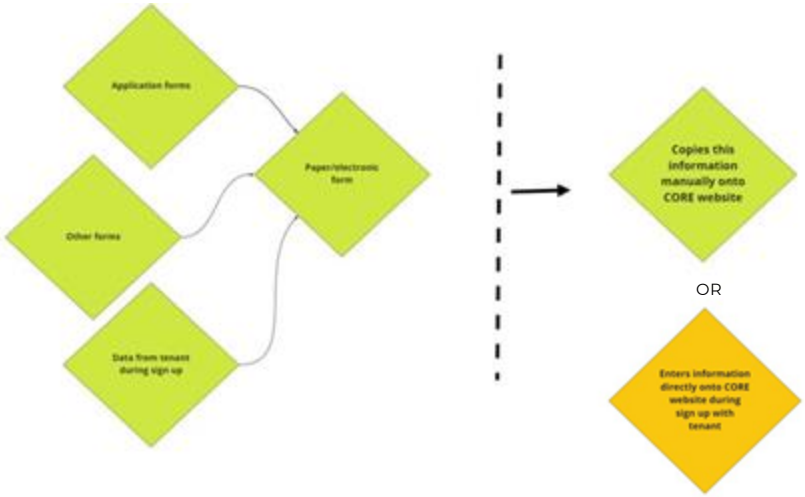
**How is data submitted?**
By individual form submission on the CORE website

**Pain points**
- Double entry of data into CORE which they have already entered into their HMS at other stages of the application process creates a heavy workload.
- They would like access to a 'CORE module' in their HMS to reduce this double entry but it is either too expensive or doesn't exist. v
- Some are interested in some type of bulk upload but don't have the relevant information on it or know how to do it well.
- Changes made to the CORE questions by MHCLG each year mean they have to amend their own bespoke form regularly.
- Poor usability of CORE website makes it difficult to answer questions quickly and accurately.

MadeTech

# Org Type 3

Not supplier-reliant – form fillers




## Type 3

No supplier

Individual Form Submission

20% - 30%

**Type of organisation:** Often small housing associations but could also include local authorities
**Amount of Social Housing Stock:** Small stock
**Number of new lettings a year:** Small number

**Digital Maturity level**
Not reliant on a software supplier. Uses own solutions to manage housing. Paper based or form based.

**Reliance on 3rd party supplier**
No reliance on a software supplier. May have previously used one but it didn't work for them.

**How is data collected?**
- Either on paper forms or input straight onto the CORE website during sign up.

**How is data collated?**
- Usually no collation stage involved.
- Sometimes housing officers will complete the paper form and pass this onto another officer/administrator to input onto the CORE website.

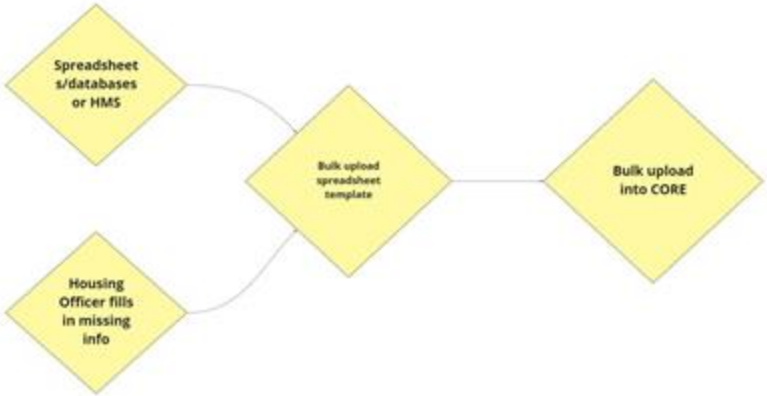
**How is data submitted?**
Mainly by manual log entry on the CORE website


**Pain points**
- Manually collecting the forms is time consuming.
- Little value is seen in collecting CORE information.
- Poor usability of CORE website makes it difficult to answer questions quickly and accurately.

MadeTech

# Uses bespoke solutions

Technology independent




## Type 4

Do their own workarounds

Own system | No supplier | Bulk upload | 10%

**Type of organisation:** All types
**Amount of Social Housing Stock:** Large or Small stock
**Number of new lettings a year:** Large or small number

**Digital Maturity level**
Not reliant on a software supplier. Have come up with their own solutions and work-arounds to manage housing, often databases and spreadsheets.

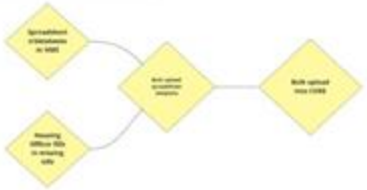
Some are very digitally mature, others rely on spreadsheets

**Reliance on 3rd party supplier**
No reliance on a software supplier. May have previously used one but it didn't work for them.

**How is data collected?**
- Collected in their own spreadsheets and databases or bespoke HMS
- Extra questions collected at sign up

**How is data collated?**
- Data automatically pulled from spreadsheet records or databases to put onto their own bespoke CORE form.

**How is data submitted?**
Bulk upload



**Pain points**
- Question changes by MHCLG means their system needs to be updated each year. This can take a lot of time and result in a backlog of CORE submissions
- Dealing with a high number of validation errors (sometimes 5v0%) once the bulk upload has been submitted takes a huge amount of time.
- Collecting and submitting CORE is a heavy workload burden
- It's difficult to see the value of CORE, as they don't know what MHCLG uses the data for.

MadeTech

# Org Type 1 user needs for Beta

| Persona | Pain Points | User Needs | Tasks |
|---|---|---|---|
| Housing performance manager | Finds it difficult to use CORE data from their organisation and make meaningful decisions from it.<br><br>It takes a long time and coordination to update systems to reflect annual question changes in CORE<br><br>UP TO 50% DATA RETURNED WITH VALIDATION ERRORS | I want to use the data collated for my own internal reporting so that I can use the data to keep track of how we are doing as a service<br><br>I want to benefit from collating and sharing my information with government so that I am motivated to spend the time doing it, and do it well (data quality) | Oversees the submission of data to CORE<br><br>Might be responsible for submitting bulk data to CORE<br><br>Analyses housing data to gain insights and recommend changes to housing strategy<br><br>Deals with validation errors |
| Business Support Officer | It takes a long time and coordination to update systems to reflect annual question changes in CORE<br><br>Housing Officers are prone to forgetting usernames and passwords but cannot reset them themselves<br><br>Validation issues are pain to resolve | I need to be able to access data collected in my systems during service delivery so that we reduce the amount of manual duplication into CORE and save time | Extracts CORE data from systems and submits bulk uploads<br><br>Sometimes inputs CORE forms onto website from paper forms filled out at signup if CORE module delayed<br><br>Works with IT team to make changes to HMS annually once question changes have been released |
| Housing officer | She worries about the sensitivities of asking some questions to the tenants, for example around pregnancy and income.<br><br>She doesn't know the answer to a few questions or doesn't understand what is being asked<br><br>She has to backfill CORE data when the new CORE module is updated each year which is time consuming | I need a clear explanation of what the service is and its benefits so that I understand the benefits of each sharing mechanism and know which is right for me<br><br>I need the process of sharing data to be as easy and quick as possible so that my service can use their time on service delivery, or making better use of the data we have collated to improve our service | Interviews tenant to ask CORE related questions<br><br>Enters data into HMS system & CORE module<br><br>Backfills CORE module when questions eventually changed |

# Org Type 2 user needs for Beta

| Persona | Pain Points | User Needs | Tasks |
|---|---|---|---|
| Housing performance manager | Finds it difficult to use CORE data from their organisation and make meaningful decisions from it.<br><br>Would like access to a 'CORE module' in their HMS to reduce the double entry their team is doing, but it is either too expensive or doesn't exist. | To ensure the housing team is efficiently getting the right tenants into social housing and that he has the right information to make decisions about social housing strategy and spot arising problems. | Oversees the submission of data to CORE<br><br>Analyses housing data to gain insights and recommend changes to housing strategy<br><br>Deals with validation errors |
| Lettings manager | Validation issues are pain to resolve<br><br>Has to prompt housing officers to complete CORE logs at the end of each month<br><br>What CORE requires sometimes doesn't match the specific circumstances of their properties/tenancies CORE is a time-consuming process | I need the process of collating CORE data to be as easy and quick as possible so that my service team can use their time on service delivery<br><br>I need to be able to access data collected in my systems during service delivery so that we reduce the amount of manual duplication into CORE and save time<br><br>I need to be able to keep track of incomplete, outstanding returns so that I can easily see what needs to be done, and when | Checks that all CORE submissions have been completed<br><br>Prompts housing officers to complete missing submissions<br><br>Makes sure housing officers are set up on CORE correctly<br><br>Deals with validation errors and coordinates team to correct them<br><br>Responds to communications from MHCLG<br><br>Sometimes fills in parts of CORE once a lettings officer has completed questions with tenants |

# Org Type 2 user needs for Beta

| Persona | Pain Points | User Needs | Tasks |
|---|---|---|---|
| Housing officer | She has to go into multiple different systems and application forms to find all the information required by CORE which makes it a lengthy process<br><br>She doesn't see the value of doing CORE and perceives it as an burdensome task in an already lengthy process.<br><br>She worries about the sensitivities of asking some questions to the tenants, for example around pregnancy and income.<br><br>She doesn't know the answer to a few questions or doesn't understand what is being asked | I need a way to feedback to MHCLG within the digital experience so that I have a joined-up way of explaining why data is what it is or query anything, without having to phone / email<br><br>I need to be able to view contextual error feedback (in the form/bulk upload/API) so that I know where errors are, and correct them quickly before submitting<br><br>I need the process of sharing data to be as easy and quick as possible so that my service can use their time on service delivery, or making better use of the data we have collated to improve our service | Interviews tenant to ask CORE related questions<br><br>Fill in CORE form on paper or directly onto website<br>Searches for extra CORE data in housing management system, application forms, emails and other places<br><br>Often completes CORE returns in batches of about 6 |
| Business Support Officer | Housing Officers are prone to forgetting usernames and passwords but cannot reset them themselves | I need to be able to access data collected in my systems during service delivery so that we reduce the amount of manual duplication into CORE and save time | Sometimes inputs all CORE forms onto website from paper forms filled out at signup<br><br>Helps navigate the CORE website, resets passwords and carries out training |

# Org Type 3 user needs for Beta

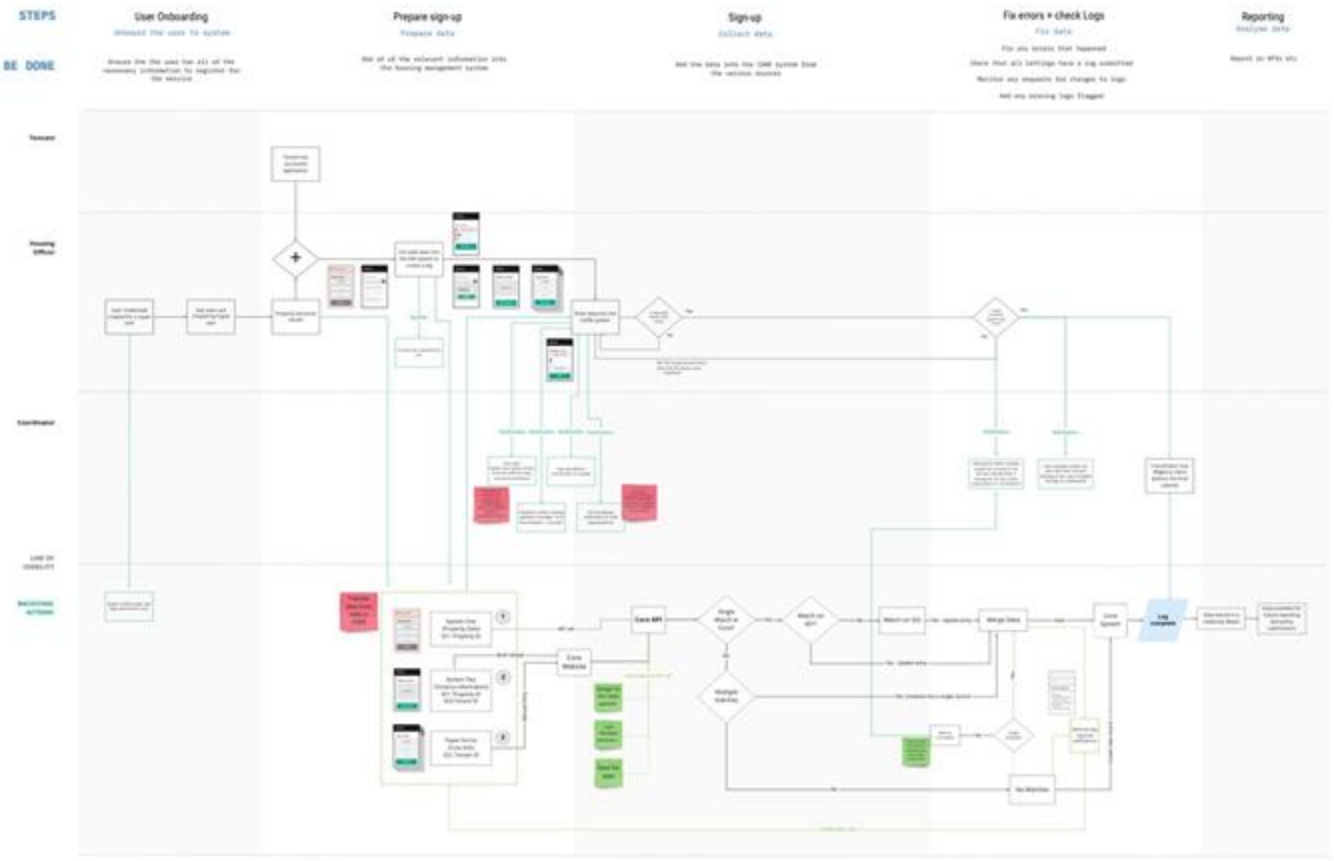
| Persona | Pain Points | User Needs | Tasks |
|---|---|---|---|
| Housing officer | She has keep record of CORE questions on paper forms, or spend additional time with tenants directly completing the website.<br><br>She doesn't see the value of doing CORE and perceives it as an burdensome task in an already lengthy process.<br><br>She worries about the sensitivities of asking some questions to the tenants, for example around pregnancy and income.<br><br>She doesn't know the answer to a few questions or doesn't understand what is being asked<br><br>Doesn't complete the form very often, so finds it laborious each time | I need a way to feedback to MHCLG within the digital experience so that I have a joined-up way of explaining why data is what it is or query anything, without having to phone / email<br><br>I need to be able to view contextual error feedback (in the form/bulk upload/API) so that I know where errors are, and correct them quickly before submitting<br><br>I need the process of sharing data to be as easy and quick as possible so that my service can use their time on service delivery, or making better use of the data we have collated to improve our service | Interviews tenant to ask CORE related questions Fill in CORE form on paper or directly onto website |

# Org Type 4 user needs for Beta

| Persona | Pain Points | User Needs | Tasks |
|---|---|---|---|
| Housing performance manager | Finds it difficult to use CORE data from their organisation and make meaningful decisions from it.<br><br>Question changes by MHCLG means their system needs to be updated each year. This can take a lot of time and result in a backlog of CORE submissions | I want to use the data collated for my own internal reporting so that I can use the data to keep track of how we are doing as a service<br><br>I want to benefit from collating and sharing my information with government so that I am motivated to spend the time doing it, and do it well (data quality) | Oversees the submission of data to CORE<br><br>Analyses housing data to gain insights and recommend changes to housing strategy<br><br>Deals with validation errors |
| Housing officer | She worries about the sensitivities of asking some questions to the tenants, for example around pregnancy and income.<br><br>She doesn't know the answer to a few questions or doesn't understand what is being asked<br><br>She has to backfill CORE data when the new CORE module is updated each year which is time consuming | I need a clear explanation of what the service is and its benefits so that I understand the benefits of each sharing mechanism and know which is right for me<br><br>I need the process of sharing data to be as easy and quick as possible so that my service can use their time on service delivery, or making better use of the data we have collated to improve our service | Interviews tenant to ask CORE related questions<br><br>Enters data into HMS system & CORE module<br><br>Backfills CORE module when questions eventually changed |
| Business Support Officer | Question changes by MHCLG means their system needs to be updated each year. This can take a lot of time and result in a backlog of CORE submissions | I need to be able to access data collected in my systems during service delivery so that we reduce the amount of manual duplication into CORE and save time | Extracts CORE data from systems and submits bulk uploads<br><br>Sometimes inputs CORE forms onto website from paper forms filled out at signup if CORE module delayed<br><br>Works with IT team to make changes to HMS annually once question changes have been released |

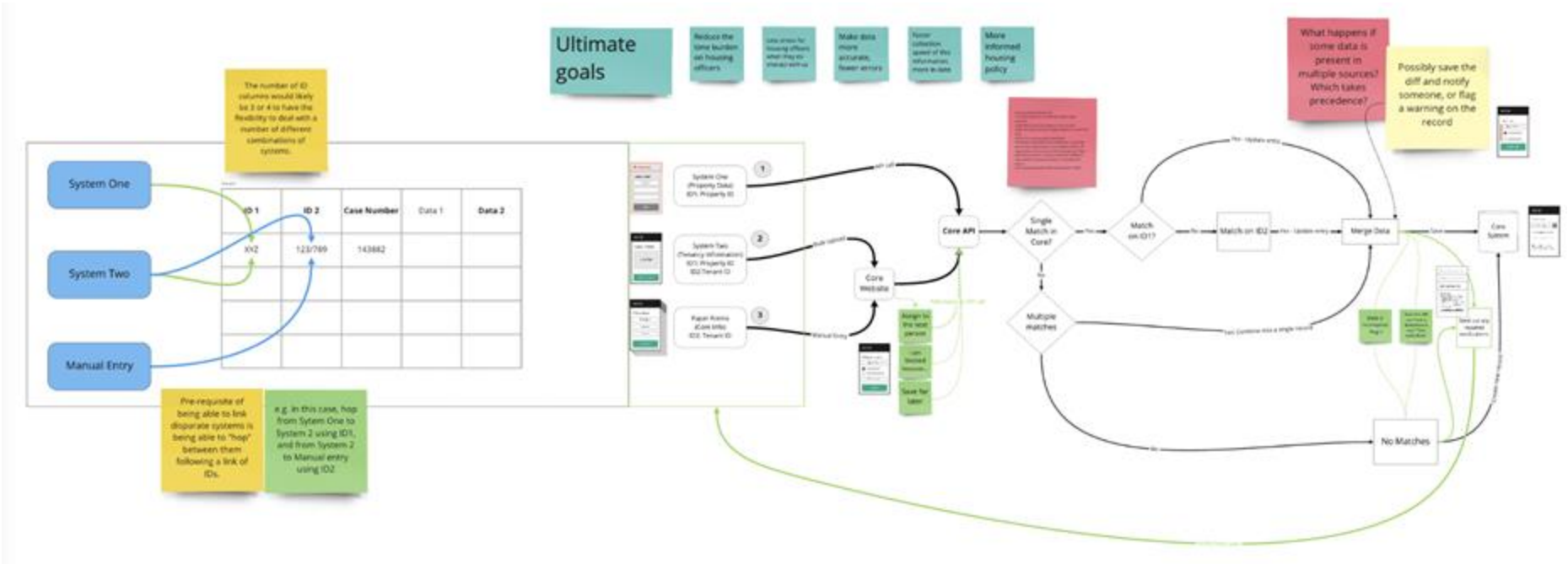
# To be journey flows

See full blueprint




MadeTech

# Linking data from different sources




See full blueprint

MadeTech