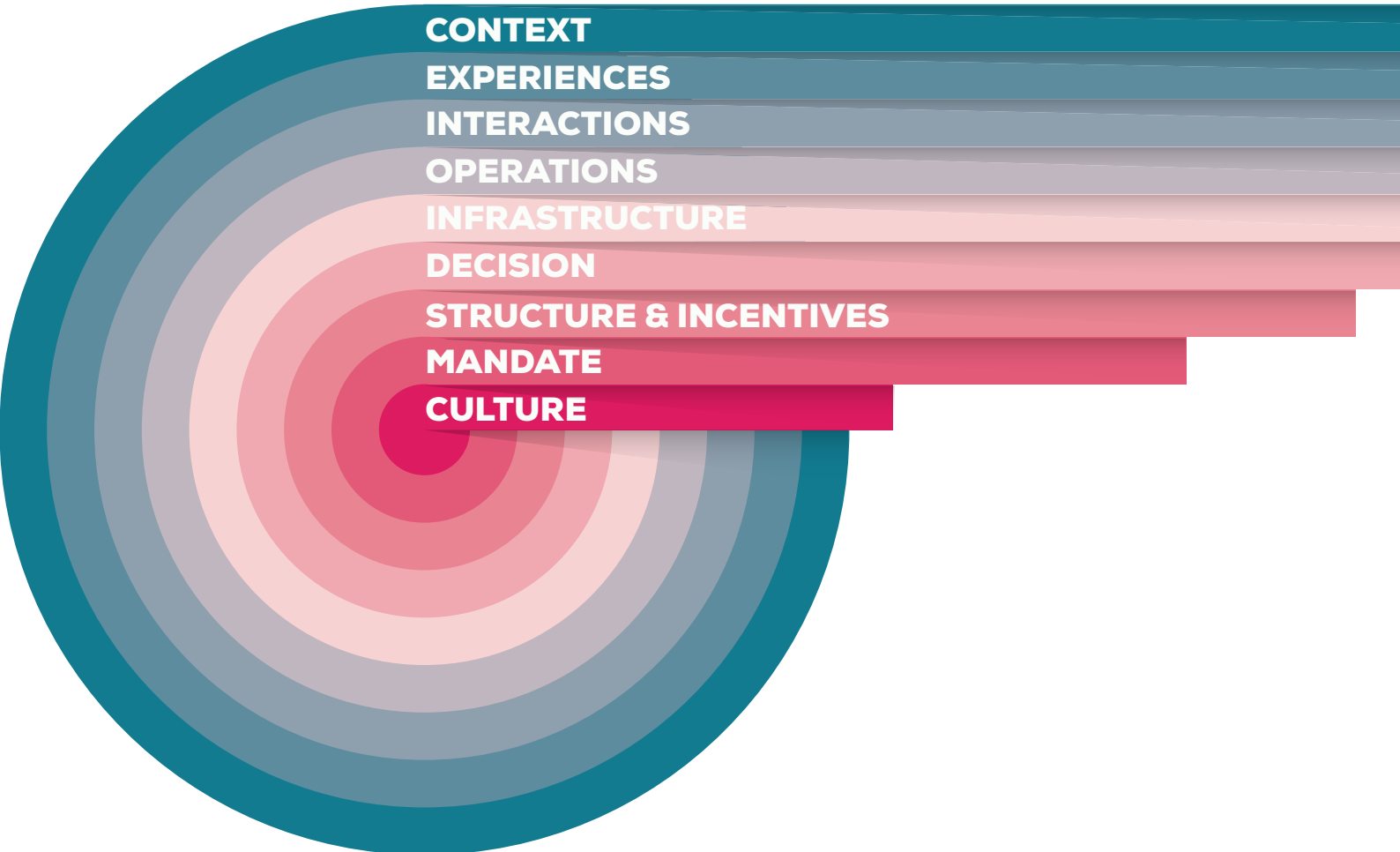
DfE DATA DISSEMINATION DISCOVERY REPORT

JULY 2018

Undertaken by hive^{IT} on behalf of the Department for Education



Foreword

Over the last 12 weeks Hive IT have been working with statisticians in the DfE to take a fundamental look at the needs of users of DfE statistics and data, and to make recommendations for how we can further strengthen our statistics "offer". This work is part of our wider package to review the coherence of DfE's statistics[1] and I welcome Hive IT's report and findings. Much of the work has confirmed what we already know about the use people and organisations make of our statistics, but the work has also given us some fresh perspectives on the user needs we need to take into account as we develop DfE's statistics. The work has also given us some helpful practical next steps to take.

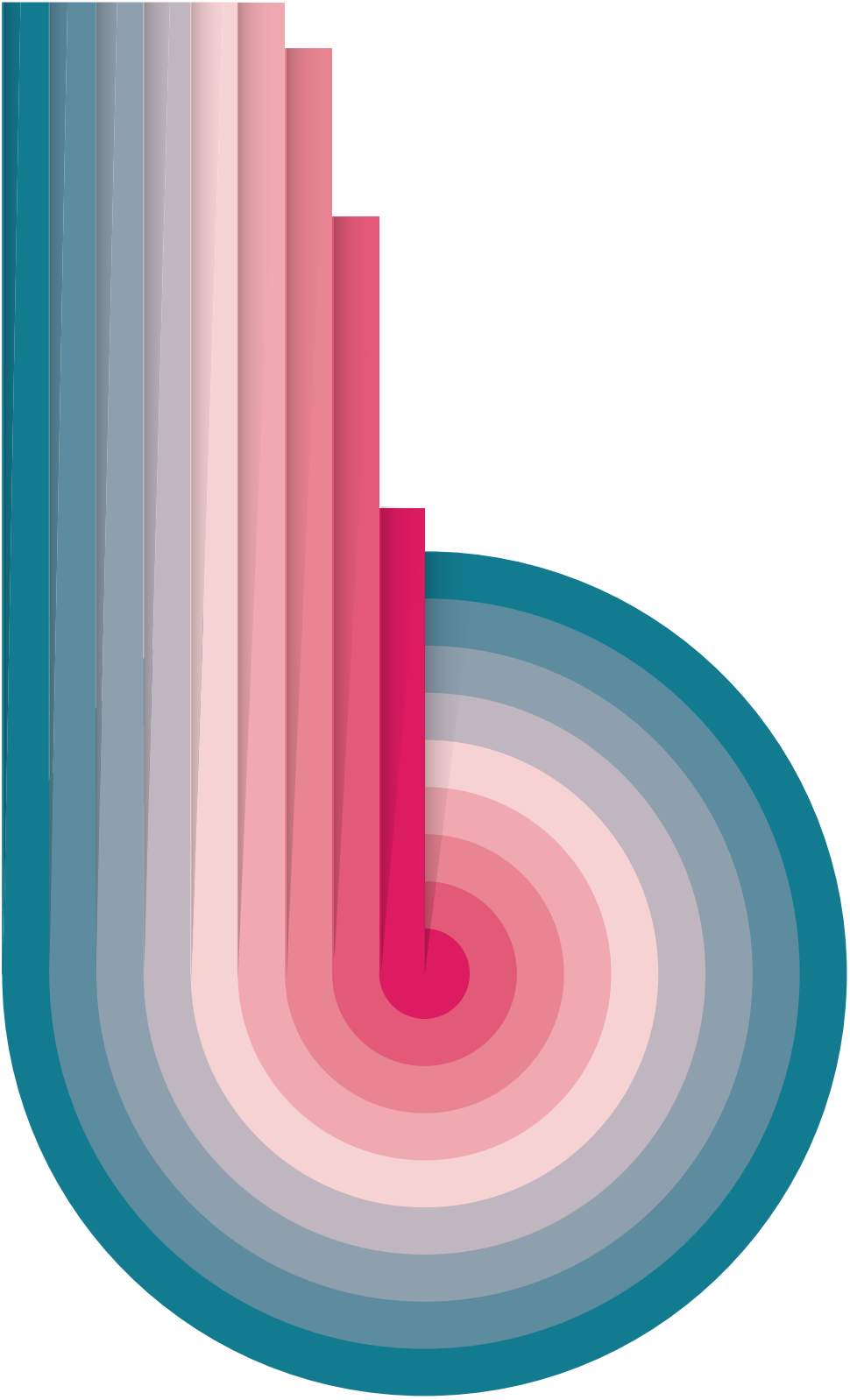
It has been good to hear about how valuable many users find DfE statistics and data; much of what we already do meets the needs of a range of users. But it is clear we have more to do to modernise and to more fully meet users' needs.

DfE will now be working to implement the two core recommendations made in this report. The first of these is about developing, standardising and accelerating the work we already have underway to move our statistics onto a "reproducible analytical pipeline". For DfE, that means data and statistics will be produced in a more automated and standard way. For our users, it means you can expect to see an increasing level of standardisation in the underlying data we provide with our statistics, and more consistent variable names and characteristic breakdowns.

The second phase of work is more complex, but at its core is about providing an easier "point and click" way of interrogating the underlying data we currently publish. Again, this fits well with the coherence principles[1] we published earlier this year. We will beginning this work over the next couple of months and will provide relevant updates via our statistics home page.

Neil McIvor
DfE Chief Data Officer and Chief Statistician

[1] https://www.gov.uk/government/organisations/department-for-education/about/statistics



EXECUTIVE SUMMARY

Context

The Department for Education is one of the largest producers of official statistics in government, publishing over 50 distinct statistics series each year – comprising around 80 individual publications – covering a wide breadth of policy areas, from GCSE results, to the social work workforce; from spending on the education system, to the destinations of leavers from higher education. These statistics are a public asset; they inform debate, monitor progress and provide transparency over the education system.

The current model for official statistics, with some exceptions such as the Compare School Performance website, is to publish a PDF document with key findings, backed up with excel tables and machine readable underlying data files. This model has been established for some time, and given the current pace of technology change, the Department wants to understand whether it continues to meet the wide range of their users' needs, or whether there are ways to improve their approach and increase flexibility.

Steps are already being taken within the DfE to explore alternative dissemination techniques and improve production processes., including developing a 'Reproducible Analytical Pipeline (RAP)[2]' to automate parts of the process. This project is intended in part to discover if their users will benefit from these efforts, explore their needs, and look at the case for a more wholesale change.

Hive IT were commissioned to carry out a 12 week research project to better understand the needs of users and producers of the Department for Education's statistics, and to identify options for modernising the dissemination of products. The aims of the project were to:

- Explore who uses DfE's official statistics outputs, how they currently use and access those outputs and if they meet their needs
- Explore how to best make those outputs available to users, learning from industry best practice
- Explore how different dissemination routes can free up statistics producers' time and help them tell clear and coherent stories with the data
- Explore how they can reduce the need for ad hoc data requests by publishing more flexible data outputs

This report will outline the findings of the research, the needs and frustrations of the users, and provide suggestions and recommendations as to what directions the DfE should take for the next stage of development. This work is part of the wider piece looking at the overall coherence of DfE statistics[3].

[2] dataingovernment.blog.gov.uk/2017/03/27/reproducible-analytical-pipeline/
[3] www.gov.uk/government/publications/standards-for-official-statistics-published-by-the-department-for-education

Project Methodology

The project was run using the GDS (Government Digital Service) agile methodology[4]. As a **discovery** project, it aimed to understand if users need alternative dissemination options and an improved production process, and to provide the building blocks for further work should this be the case.

Should this be the case, the next phase is called **alpha**, where any service is prototyped and tested in order to establish feasibility, problems, costs and risks, and understand whether to move the project on to **beta**, take a different approach or end it.

High Level Research Findings

Through our research we identified the following key findings:

- Many users are very appreciative of, and dependent on, the statistics produced by the DfE. For external expert consumers, in most cases their livelihoods and ability to do their jobs depend on it

- For internal users, even when skilled, their role means they feel the need to check their interpretation of DfE-published statistics and would like to be able to better self-serve. They want to be less of a burden and less reliant, and not face the delays this results in.
-

- For many producers of statistics within the DfE there is a commitment to modernisation. Tied to this, there is a lot of frustration with the apparent inflexibility of GOV.UK as a platform for disseminating statistics, and with the time spent on routine tasks such as QA and Suppression

- All identified user types have a desire to be able to customise the data according to their requirements, they want consistency in data, ease of use, the ability to view historical data alongside current data and to be made aware of changes to datasets or when new publications are released more effectively

[4] https://www.gov.uk/service-manual/agile-delivery

Headline recommendations

Hive IT recommend that as a result of this Discovery Project the DfE should:

- Undertake an Alpha project focusing on internal process changes. This explores developing the existing work done on automating routine analysis and making the processes more easily reusable (known internally as the the Reproducible Analytical Pipeline or RAP) and expands its use throughout the Department

- Undertake a separate second Alpha project focusing on the dissemination of data which provides a new platform for access to, and consumption of, statistics. This should make the statistics easier to find, access, navigate and understand.

- Implement a number of smaller changes which are external to the Alpha projects but which will have significant impact

The Future

It has been clear from our research, and noted time and again, that the data which the DfE are producing, its statistics and its narratives, are so very much appreciated and relied upon by their users. If the DfE are to address the user needs uncovered as part of this Discovery, they have the opportunity to be world-leading in their dissemination of statistics in an inclusive and user-friendly manner.

The DfE has the ability to make real and impactful changes for its users and Hive IT are proud to have been able to support the DfE in taking that first step to enabling change.

Contents

Part 1:
RESEARCH

A) Introduction

The emphasis of this Discovery project was understanding who the users of the Department's official statistics are, what they need, and where the DfE are currently meeting, exceeding or failing to meet those needs. This Discovery project also considered high level approaches that could be used to respond to those needs.

The Discovery Project was broken down into 4 phases and ran from the 19th March - 12th June. The phases of the project were as follows:

Phase 1
Internal research with the DfE teams producing statistics

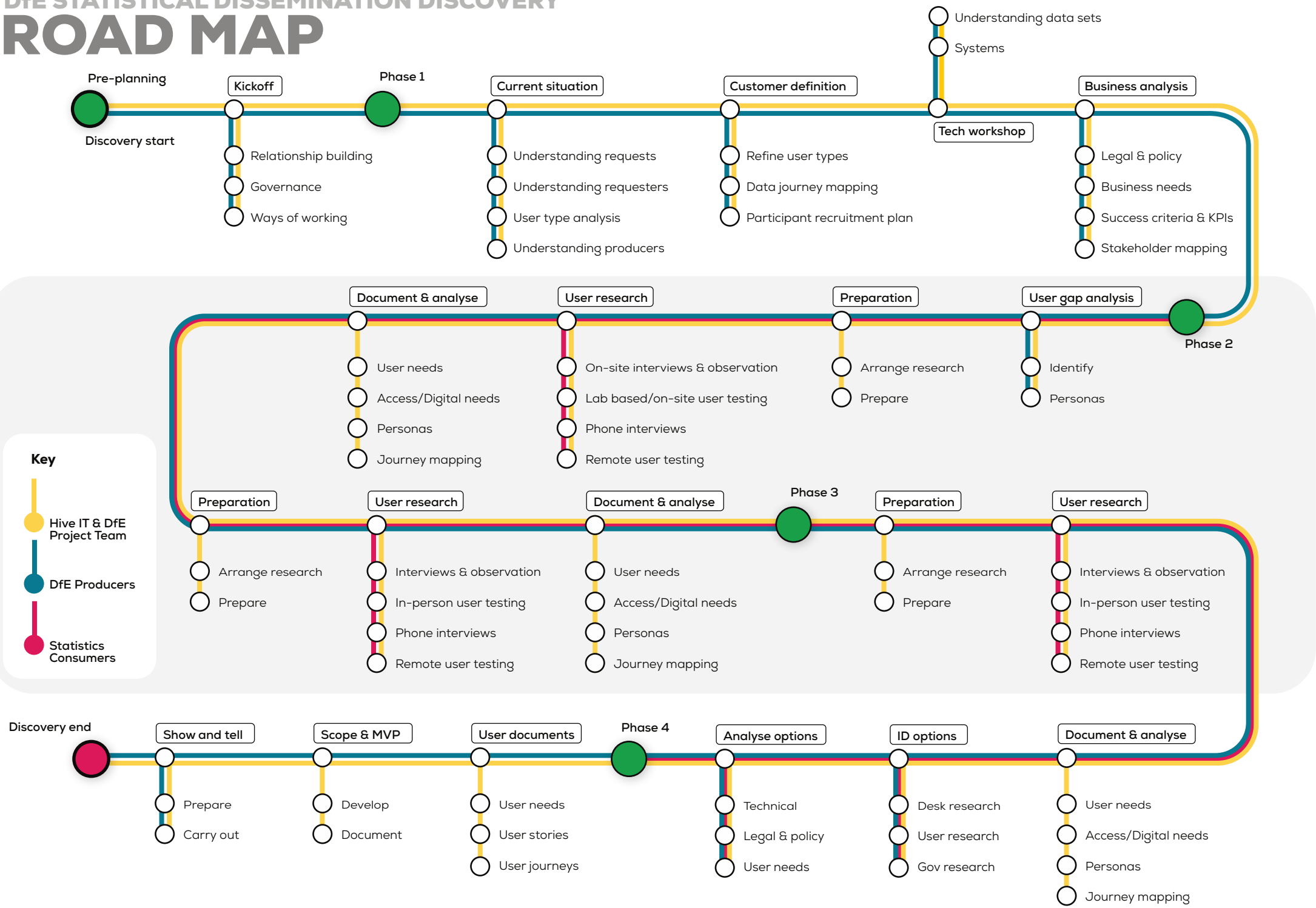
Phase 2
Internal and external research with statistics users

Phase 3
User Testing and best practice for statistics production and dissemination

Phase 4
Discovery wrap up and Alpha phase definition

DfE STATISTICAL DISSEMINATION DISCOVERY

ROAD MAP



Over the course of the project we engaged with 95 users directly, received 133 survey responses and had a team of 8 Hive IT members engaged on the project at varying times.

This document aims to provide details on:

- What the research uncovered
- What we would recommend as a result of the research
- The user types we identified
- How we engaged with research participants

The recommendations within this document (as a result of our research) are intentionally user-focused, talking about the needs that should be addressed rather than specific technical solutions that should be developed.

We believe that placing the user at the heart of everything we do means that solutions will ultimately shape themselves, through the collaboration of a team working to meet those needs. If we focus on the solution first, we may never consider whether we are meeting the defined needs.

B) The users of DfE Statistics

Participant Engagement

Participant engagement was identified at the very outset of the project as being a key dependency to the success of the Discovery. After all, it was only with the engagement of participants that any meaningful research could be completed.

Both Hive IT's and the DfE's previous experience of User Research was the key factor for extending the initial 6 week Discovery project over a 12 week period, ensuring that we had enough time to access the participants who met our Personas and User Types, and completing the research to the highest possible standard.

Through this Discovery project, we identified 14 types of specific users:

Within the DfE

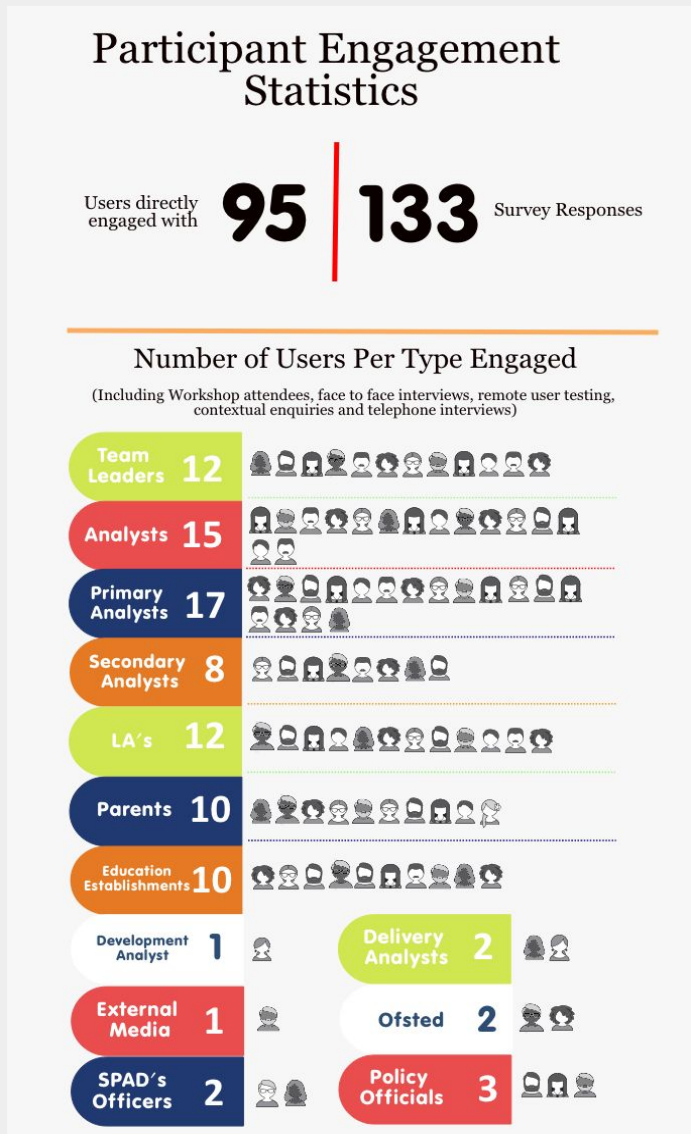
- Primary Analyst
    - Producer of DfE Statistics
- Secondary Analyst
    - Takes one or more outputs from primary analysts and other sources, carries out further analysis and interpretation
- Team Leader
    - Team Leader of analysts
- Development Analysts

    - Use statistics to support and develop policy
- Delivery Analysts
    - Use statistics to deliver services and policy commitments
- Special Advisor's Officers
    - Support Special Advisors
- Policy Officials
    - Deliver services and develop policy

Outside the DfE

- MPs Researchers
- Education Establishments
- Analysts (External)
- External Press
- Parents
- Ofsted
- Local Authorities
- Marketeers

Engagement Statistics



We identified a split between those who primarily consume data, and those who produce it. This exists both inside and outside the DfE - so for example an analyst in a Local Authority consumes DfE data, but then analyses it and creates their own outputs. This is evident both in behaviour and expertise.

These splits, as well as the common themes, user needs and aims that were identified allowed us to group the users under 4 defined high-level user types:

i) Expert external producers and consumers

Correlates to ONS 'Expert Analyst' Persona

The conduit between DfE statistics and non-expert consumers and the public.

- External Analysts (Commercial)
- External Analysts (Campaigning)
- Local Authorities
- External Media
- Sector Partners

These users tend to be skilled analysts, capable of analysing and understanding data using a variety of tools.

They produce analysis, tables, dashboards, charts, graphs and visualisations for others and act as a key conduit between the DfE data and non-expert consumers in Local Authorities (LAs), schools, charities, Non Governmental Organisations (NGOs) and External Media. Many are highly reliant on machine readable data, to import data into their own software and systems, and to access the full range of published data.

They are very appreciative of, and dependent on, the statistics produced by the DfE, and in most cases their livelihoods and ability to do their jobs depend on it.

Key problems are inconsistency, changes, and lack of machine readable data.

ii) Non-expert external consumers

Correlates to ONS 'Inquiring Citizen' Persona

The normal user, public and parents - only users of DfE statistics *through the interpretation of others.*

- School governors
- Parents
- Marketeers
- MP's researchers
- Public

These users tend to have no training and few skills in understanding or analysing data.

They are currently reliant on the interpretation and analysis of others, primarily the *Expert external producers and consumers.* This means their understanding of quality and progress in education comes from the press, local and national and OFSTED.

We found no evidence that they consume data directly from the DfE through the gov.uk statistical pages, and little that this takes place in other ways. It is important that the DfE decide if they are happy with this situation.

The key problem is the lack of awareness of what is out there.

iii) Non-expert internal consumers

Correlates to ONS 'Information Forager' Persona

The colleagues of DfE analysts, lacking analytical training but with a great appetite and need for data.

- Special Advisors' (SPAD's) Offices
- Policy Officials
- Internal press office

They tend not to be highly skilled or trained in data analysis, although this is not universal. They have a great deal of need for data and statistics to inform their, and their superiors', positions. Due to time pressures within the department, they often struggle to get the help they need. Even when skilled, their role means anything they find must be passed through Quality Assurance within the DfE and fact checked.

They place a high load on analysts, and are aware of this. They want to be less of a burden and less reliant, and not face the delays this reliance implies.

Key problems are availability of resource, and not having the skills or user-friendly tools to self serve.

iv) Expert internal producers and consumers

The experts; analysts within the DfE, with access to record level data, responsible for disseminating data and analysis to everyone else.

- Primary Analysts
- Development Analysts
- Delivery Analysts
- Secondary Analysts
- Team Leaders

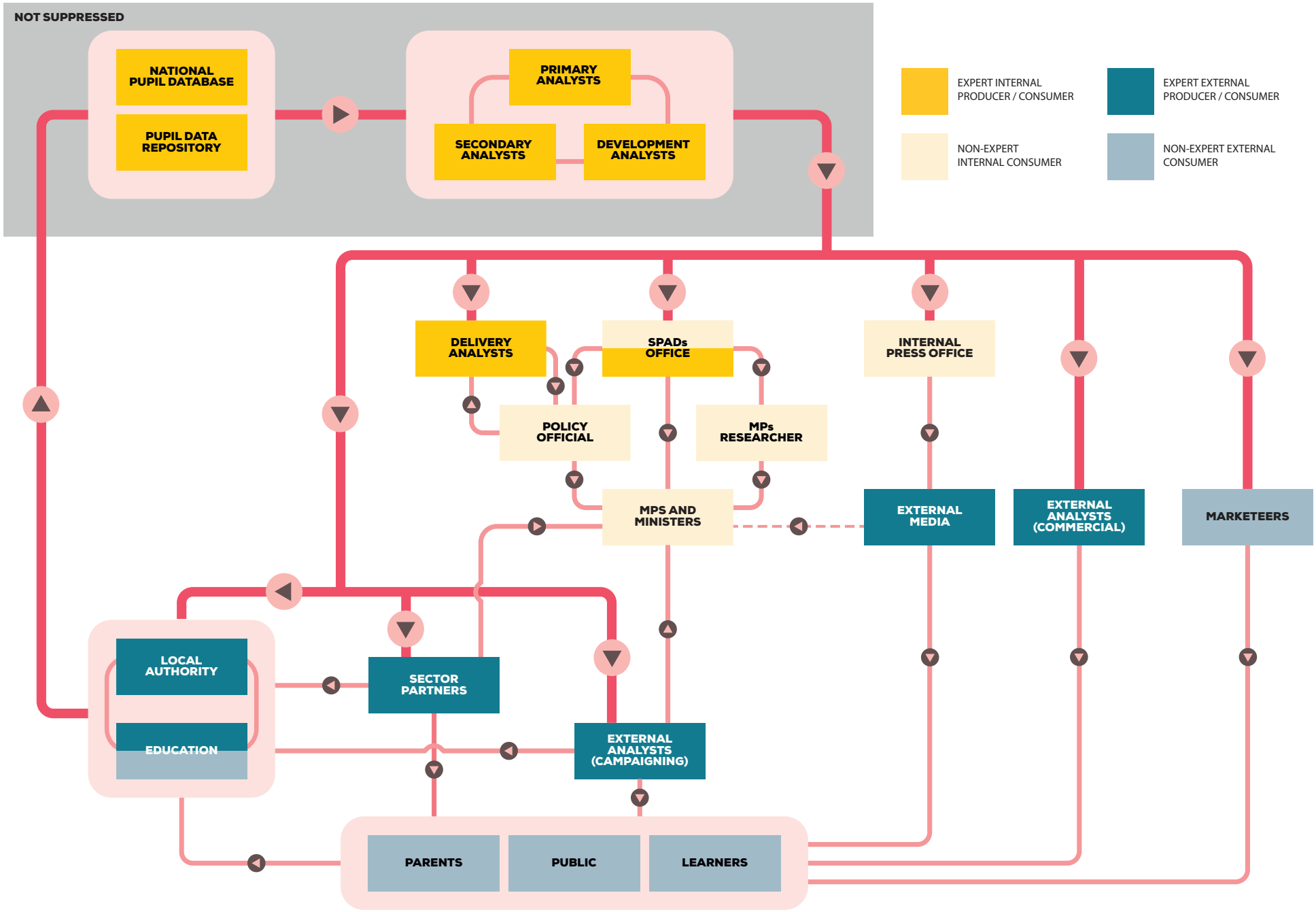
All highly skilled, although there is a wide range of knowledge in R and SQL, as well as current best practice inside and outside government. The lack of time to increase knowledge is a common frustration amongst this group.

As a whole, there is a commitment to modernisation, increasing skills, and finding new ways to disseminate and explain data. Tied to this is a lot of frustration with the inflexibility and restrictions of GOV.UK as a statistics dissemination platform, and with the time spent on routine tasks.

Key problems are time pressures, excessive time on routine tasks, and the lack of an agreed path or time to be innovative in their analysis and explanations.

DATA FLOW

BETWEEN USER TYPES



C) The Current Situation

What the DfE does well

Our research in the Discovery project wasn't focused on uncovering frustrations which users encountered with the DfE Statistics, but instead understanding how users interact, use, and access the data. While this ultimately showed us their frustrations, it also gave us an insight into the things that the DfE are currently doing well and should continue to do.

It is important to highlight these to ensure that the DfE recognises where it is succeeding, so that if necessary it can put controls in place to ensure that these successes are continued and maintained.

Trustworthy and relied upon

The DfE is recognised as the official source for education statistics. The DfE's data is trusted for its quality and accuracy.

> *"It's hard to trust other sources..."*
>
> ***Researcher (Analyst)***

Research has shown that Local Authorities absolutely rely on the DfE's data, some even mentioning that they could not do their job without it.

> *"Publications from the DfE are really important to us..."*
>
> ***Local Authority***

Attitudes and the desire for change

Individuals within the DfE are protective of the data that goes into their publications; this *'ownership'* means that staff really care about the quality and accuracy of *their* publications.

Staff are passionate about making improvements to develop their workflows and automate routine tasks which would allow them to use their time more creatively and in effect benefit the wider audience.

> *"It should be simple to access government statistics online..."*
>
> ***Internal Team Leader***

Recognising issues

Internally the DfE are great at recognising when things are no longer working well. Staff are diligent in recognising inefficiencies and identifying opportunities for improvement.

This self-reflective culture is a catalyst for continuous improvement.

Protection and risk aversion

It goes without saying that the DfE handles a lot of sensitive information from data on looked-after children to teachers' wages. Staff have a responsibility to protect the individuals who are reflected in the data they collect.

It is obvious that staff are devoted to making sure that the statistics produced cannot be traced back to an individual, ensuring that the data cannot be used in a malicious way.

Every team member we spoke to showed an affinity for data they manage; there is a sense of ownership.

This attitude to the management of risk and containment is a great foundation for trust between those who generate the data (e.g. learners and teachers), those who collect and manage it (producers), and those who consume it (consumers).

Accessibility considerations

Producers care about the accessibility of their publications. There are clear considerations for accessibility in the publications that are produced by the DfE.

In the production of summary documentation PDFs, accessibility is checked using tools built into production software (mainly Microsoft Word Accessibility Checker).

This process helps to ensure that users with visual impairments are catered for, but does not help users on the autistic spectrum or with a low reading age, for example. With guidance, we're confident that the DfE could improve in this area.

Areas for improvement

Despite the valuable and valued work the DfE produces, there were some consistent problems that were identified during the research.

Understanding each user's frustrations was important and formed the basis of our overall recommendation. As well as resulting in the identification of 77 user needs, they have been distilled down into the three main themes below, alongside the most common frustrations in that theme.

Visibility, findability and understanding of data

Many users struggle to find the right data. There is a lack of awareness of what is available, and outside the expert cohorts, little ability to understand what is found.

For example:

Users don't know when new data is published

Most Expert users know where to look to find the publications they use regularly, but they don't always know *when* to look for them.

Lack of skills and knowledge to analyse data

Data is disseminated in CSV and Excel XLSX formats but not all consumers have the skills or tools required to find their own answers to questions.

Lack of historical data

Expert external consumers have a requirement for historical data. Whilst this is present in some cases, in some cases it is served in separate publications.

PDFs don't provide the answer I'm looking for

For non-expert users, or those unfamiliar with a publication, the PDF is often the first port of call. However, if those users have a specific question, it can be hard to find answers.

Consistency and access to data

When it occurs, inconsistency and changes between years and publications causes problems, as does the occasional lack of machine readable data.

For example:

Inconsistency

A recurring theme throughout the research was the level of inconsistency across the publications; the same publication can even vary one epoch to the next. These inconsistencies are present in the various output formats (PDF, CSV, Excel tables etc) that make up each publication.

Things like suppression and rounding being handled differently frustrate consumers, especially external ones who see the DfE as entity; not a series of teams.

Data publication frequency and delays

External consumers (mainly experts) recognise that the DfE needs time to collect and analyse but can't understand why some statistics take so long to

publish. Due to complexities in the production process, statistics are often published upto 12 months after they are collected.

'Excessive' amount of data in the underlying datasets

Users are required to download the entire dataset, then remove and manipulate extraneous data reducing it to a specific subset. Many expressed a desire to be able to customise the data they downloaded.

Time and resource

The time spent on routine tasks leads to time pressures within the DfE, prevents statistics producers from spending their time more productively on analysis and communication, and means they lack the time to learn and innovate.

For example:

Manual production processes are very time consuming

Manual processes, such as suppression, disclosure control and Quality Assurance are important parts of the production of DfE statistics. However, many are repetitive and could be partly automated, saving time while improving quality.

Requests for facts and information are very time consuming

These include fact checking, Parliamentary Questions (PQs) and Freedom of Information (FOI) requests.

DfE statistics producers spend a large proportion of their time responding to a variety of requests, in addition to any publication work which they are already committed to undertaking.

Providing more information in the underlying data, and a better service to consume this information, should decrease the number of and amount of time spent answering these requests, as well as making it easier for unskilled users to understand the data used in request responses

Part 2:
RECOMMENDATIONS

A) Recommendations Overview

Hive IT recommend that as a result of this Discovery Project the DfE should undertake two projects, covering the internal production and external dissemination of data, and implement a number of smaller, quicker changes.

The projects should be run using the GDS (Government Digital Service) agile methodology, following on from this discovery project with two alpha phases. An alpha phase is intended to build prototypes of a service, test them with users and establish that the service in question is technically feasible. It allows the teams to find problems, identify solutions, estimate risks and establish whether the service should be further developed into a beta phase, where the service is developed from a prototype into a scalable, fully functional version.

Alpha phases:

- Undertake an Alpha project focusing on internal process changes which explores developing the existing Reproducible Analytical Pipeline (RAP) approach for producing official statistics and expands its use throughout the Department

- Undertake a separate second Alpha project focusing on the dissemination of data which provides a new platform for access to, and consumption of, statistics

- Implement a number of smaller changes which are external to the Alpha projects but which will have significant impact

The Recommendations are based on addressing user needs, and follow a review of current best practice both inside and outside government. They are supported by technical investigations, having an awareness of cost, and being informed by the current data process.

B) Alpha 1
The 'Process Alpha'

Process Alpha Overview

The teams who produce statistics currently use a variety of methods and software to take the record level data collected through established collection processes, combine, analyse and then publish it in a variety of formats. In most cases this requires moving the data across a variety of tools to produce the final publications - for instance from SQL, into Excel to manipulate and create tables, then numbers and charts are exported to Word to write the final summary document. This process can be unnecessarily time consuming and if any changes are made to the data, the entire process often needs to be manually repeated. In addition, outputs from this process are often not machine readable and to create these files is challenging and is not always carried out.

The Absence, Exclusions and Special Educational Needs (SEN) statistics teams are currently using features from what is known as a Reproducible Analytical Pipeline (RAP) to carry out parts of this process. They have used a combination of SQL and open source software, R and Rmarkdown, to develop reproducible code and a set of functions to take the machine readable underlying data, and then use it to produce the elements needed for each Statistics release. Their aim has been to do this in such a way that each release should be entirely reproducible from the underlying data and if the data changes, the outputs can be updated with minimal manual intervention.

Process Alpha Key Recommendation

We recommend that the existing Reproducible Analytical Pipeline (RAP) is built upon and expanded throughout the Department by the DfE.

The existing RAP work should be developed over the course of the Process Alpha into a Minimal Viable Product (MVP), which includes the addition of automated quality assurance and disclosure control.

This MVP should then be used to prove the concept sufficiently and justify expanding its use and scope. We recommend reviewing and improving the code, adding tests and rolling it out to produce an additional National or Official Statistic, by a team who are not currently using the RAP.

In order to remove the burden of repetitive, time-consuming tasks as much as possible, the RAP should also be expanded to handle Disclosure Control (DC) and Quality Assurance (QA) in addition to aggregation of data.

What the Process Alpha addresses

- A standardised production process would solve a lot of the inconsistency issues experienced by consumers; it would allow teams to publish data in the same ways, using the same methods and following strict naming conventions. Machine readable versions of publications would be produced as standard.

- After data is collected, it would be released sooner, thanks to automated suppression and quality assurance. This would also free up more time for analysts, allowing them to focus on providing more accessible summaries; telling better stories with the data.

- By making breakdowns of data easier and faster to produce and reproduce, more data can be published, therefore requests for additional data should be reduced

- Analysts would be able to devote more time to informing colleagues and providing insight on important policy issues and current affairs.

- Consistently named variables, IDs and column headings would mean that data could be cross-referenced allowing for deeper analysis.

- Historical data fed into the pipeline would produce a consistent way for time-series data to be combined into one resource.

- Process Alpha will not only provide a consistent production pipeline but also aims to create a standardised framework in which improvements to data dissemination can be trialled and implemented.

- The framework would also highlight and facilitate training requirements and allow colleagues to focus on personal development with a view to supporting the implementation of RAP across the department in future.

Future enhancements

Further future enhancements, consisting effectively of a beta phase, would be

- Beta 1) Roll out the expanded (Automated Disclosure control) RAP to those teams already using RAP
- Beta 2) Roll out the expanded RAP to all other teams
- The RAP would then be continually iterated on and improved by the RAP teams

C) Alpha 2 - The 'Dissemination Alpha'

Dissemination Alpha Overview

During our research, user testing and interviews showed a lot of frustration (both within and outside the DfE) with GOV.UK as a platform to disseminate statistics. User testing showed the findability of the data needed was very low, with non-expert users struggling to find the correct page, and even expert users struggling to find the correct set of data and file type within each page. Almost every user and analyst, even when familiar with the statistics being tested, downloaded and opened multiple files before they found the correct one.

The PDFs were only used to scan for particular pieces of information, rather than read as a whole, and never printed - behaviors the PDF format is very badly suited to.

Non-expert consumers were entirely dependent on the producers to analyse and interpret the figures accurately. This meant that if an area they were interested in was not covered sufficiently, internal users had to commission analysis or fact-checking, and external analysts struggled and were forced to send in FOI requests, call or email the named statistician, or interpret the data as best they could. This was a particular issue with what had been chosen to be included in Excel tables. Producers currently face the unenviable choice of producing massive, carefully formatted Excel tables covering as many use cases as possible, however rare - or focusing their time on the core use cases. This makes the files much easier to use and more suitable for most users, but edge cases aren't served.

Expert external users were found to be highly dependent on high quality machine readable data being available, and where only Excel files were available wasted a great deal of time converting and importing these. Separate files for different years also caused difficulty and wasted time, as in many cases they were concerned with analysing change over time.

Dissemination Alpha Key Recommendation

We recommend that a new service that allows access to and consumption of statistics is explored to meet the many varying user needs identified during Discovery.

The new service should be prototyped using iterative methods over the course of the Dissemination Alpha into a Minimal Viable Product (MVP). This MVP should then be used to prove the concept sufficiently and justify expanding its use and scope.

We recommend that the scope of the 'Dissemination Alpha' includes the design and prototyping of the new service, including the development of functionality to a sufficient level to conduct user testing around a single dataset. The aim will be to refine the proposed solution and inform further expansion beyond Alpha.

Recommended MVP features for the Dissemination Tool

From the learning gained during Discovery, we recommend that the tool developed as part of the 'Dissemination Alpha' provide external users with the following features:

1. The ability to view the data through multiple lenses or funnels
2. A platform for Primary Analysts to tell the story of their publication more effectively through web pages, with a consistent structure and tables, figures, graphs, charts and other visualisations being interactive and using the data on the platform
3. The ability for consumers of those publications to drill down and filter from the publications using the publication's figures, tables and visualisations as a starting point, to explore the data from their own point of view
4. The ability for users to view data over time, as well as a snapshot
5. The ability for users to customise and filter the data they are viewing at any point
6. The ability for users to access the data in the way that most suit their needs
7. The ability to access methodology, definitions and contact details for any data they are viewing
8. The ability to feed back to the producers responsible at any point
9. The ability to access a permanent link to any view or customisation of data
10. The ability to be notified when data changes

What the Dissemination Alpha addresses

- A single department-wide dissemination platform used to manage publication tasks would introduce greater consistency across the department and allow producers to focus on quality.

- Data published through a single platform would make it easier for consumers to find and drill-down to the statistics they are looking for.

- A web-based interface designed and controlled by the DfE would allow all consumers to access and analyse the data without the need for specialised tools. This would lower the barrier to entry. Users would be able to perform their own analysis and produce visualisations on the platform, or in the case of experts, download to use with their own tools.

- A web base interface should facilitate better accessibility practices and should be optimised to allow access using assistive devices in a way that PDFs cannot be.

- Consistently named variables, IDs and column headings would mean that data could be cross-referenced allowing for deeper analysis. Methodologies and formulas *in context* will allow consumers to recreate data easily.

- 'Deep-link' URLs allow users to share their specific analysis with others whilst also providing a method for DfE producers to provide richer, easier to understand responses to ad-hoc, FOI requests and others with minimal effort

- The ability to automatically export any of the data in multiple machine readable formats would allow consumers to download only the data they need. File names would be algorithmically generated ensuring they remain consistent.

- As new statistics are released they would appear on the platform whilst still allowing access to historical data which would allow easier time-series based analysis. Allowing consumers to subscribe to receive notifications would ensure that they always know when new data is published.

Most prominent Quick Wins

The following changes were identified during our research, as the 6 most impactful changes which could be implemented by the DfE external to any Alpha project.

- Restructure of DfE statistics publication page to prioritise the most recent and relevant data and be easier to navigate and understand
- Department-wide agreement on consistency and format
- Every dataset to have machine readable data and to establish clearer best practice in making machine readable files available
- Highlight to users the ability to subscribe to notifications
- Create an improved DfE Statistics landing page, leading to and available from the individual publication pages.
  This should include a themed hyperlinked index of all the data created.
- Set up a repository / wiki and encourage a code-sharing culture

Summary

Over the course of the project, Hive IT have focused on understanding how users interact with, use, and access the data DfE produce. Whilst the main core of what the DfE produces is data and its publications, this project has highlighted the integral importance of the people that produce and use that data.

As with any research project of this scale we have come across frustrations with how interactions are conducted and how data is produced. However, we have gained valuable insight into the things that the DfE are currently doing well and should continue to do.

DfE staff are devoted to the quality of the data that is produced, and this is reflected in how valuable the DfE data is considered by users, a statement that has been iterated in numerous interviews with end users. Not only that, but the continued efforts and dedication to making sure sensitive data is protected and can't be misused demonstrates the DfE staff members' sense of ownership and responsibility to the data they provide. Furthermore, they are committed to making the widely available data cater to accessibility needs.

Without the engagement from DfE staff, external partners and other key user groups, this project would not have been possible. The involvement from DfE staff has been integral to us reaching the number of users we have and gaining such valuable insights so that we can recommend improvements for the future. We believe this self-reflective culture, and the DfE staff members' ability to recognise things that are no longer working well, equates to this positive environment for change.

There is already a lot of inherently good practices and positive change occurring within the DfE and there are quick wins to be had to make further improvements. The culture of self improvement already present at the DfE means that these small improvements can be made quickly.

Our findings are not an exhaustive list of improvements that could be made; we have focused on areas that can show real improvement quickly, in the form of quick wins, and advised that 2 Alpha phases should follow the release of our research.

Ultimately we have uncovered a real desire from the DfE to service their users better, whilst maintaining the quality and value of the data that they produce. We believe that by combining this desire with the recommendations we have provided, the DfE have the opportunity to be world-leading in their dissemination of statistics.