

# ID Inclusion & Data Sets Project

**OIX** OPEN IDENTITY  
EXCHANGE



## CONTENTS

Executive Summary .....	03
<b>1. Introduction</b> .....	05
<b>2. ID Challenged Segments &amp; Methodology</b> .....	07
2.1 The UK Identity Challenge .....	08
2.2 How we defined the ID Challenged .....	10
2.3 ID Challenged Segments .....	11
2.4 Key Traits & Variables .....	12
<b>3. Data Sets &amp; Prioritisation</b> .....	13
3.1 Data Sets which might help the ID Challenged .....	14
3.2 The ID Challenged .....	16
3.3 ID Challenged Heat Map .....	17
<b>4. ID Challenged and Alternative Proofing Techniques</b> .....	25
4.1 Alternative ID-Proofing Techniques .....	26
4.2 Assessment & Prioritisation Criteria .....	26
4.3 Prioritised Alternative ID-Proofing Techniques .....	27
<b>5. Summary of Findings</b> .....	29
Appendix 1 – Methodology .....	32
Appendix 2 – ID Challenged Segments & UK Maps .....	34
Appendix 3 – Data Sets .....	45
Appendix 4 – Alternative ID-Proofing Techniques & Methods .....	52
Appendix 5 – Additional insights about the ID Challenged .....	58
<b>Glossary of Terms</b> .....	59

**Author:** Dr Sarah Walton

**Editor:** Nick Mothershaw, OIX Chief Identity Strategist

This report was overseen by the **Digital Inclusion Steering Group**.

**Contributors:**

Carol Buttle, Pablo Romero, Chris Allgrove – GDS

Mark Sugden – DBS

Emma Lindley, Louise Maynard-Atem – Women in Identity

Jim Purves – Post Office Ltd

Katherine Holden – techUK

Harry Weber-Brown – TISA

Jonathan Middleton – UK Finance

**Experian** were commissioned to conduct the segmentation work.

## EXECUTIVE SUMMARY

This report identifies those in the UK population who cannot verify their identity through traditional means as they lack the ID documents typically used to verify their identity: passport and driving licence as well as lacking a data history.

**The hypothesis**, that access to identity services will be improved by introducing new data sets into the ecosystem and leveraging alternative proofing techniques, was found to be positive (in theory) in so far as there are data sets that likely include some of these people.

### This report discovered the below:

- There are an estimated **5.9 Million**<sup>1</sup> individuals in the UK who are ID Challenged.
- The introduction of additional data sets to the ecosystem will likely increase access to government and private sector services.
- Alternative ID-Proofing techniques might support some of the ID Challenged to get a Digital ID.

### The ID Challenged were segmented into five main segments:

Rural Solitude	Vintage Veterans	Budgeting Families	Urban Renters	Community Culture
<b>287,429</b>	<b>836,367</b>	<b>1,238,117</b>	<b>991,339</b>	<b>291,074</b>
Estimated adults	Estimated adults	Estimated adults	Estimated adults	Estimated adults

<sup>1</sup> Of those, 3.64 Million were analysed in detail. It is reasonable to assume all ID Challenged people can be helped by the approaches detailed in this report.

**Prioritised Data sets to target for access: ranked in order of numbers in group and ease of access**

Data Set (Ranked based on ID Challenged Reach)	Data Set Source	ID Challenged Coverage (estimated)	Ease of Access
1. NHS Patient Number	NHS	Most of ID Challenged	5
2. National Insurance No.	DWP	Most of ID Challenged	3
3. Council Tax bill	Local Authorities	2,140,508	1
4. Online Banking	Open Banking (OBIE)	1,979,564	2
5. Learning Records Service	DfE	1,200,000	4

Online banking is already being considered through the Open Banking and Council Tax bill and Learning Records Service (LRS) are more likely to be more accessible than NHS & HMRC data.

The data sets have been prioritised based on ease of access (1 being easiest).

The ranking is subjective, based on the views of contributors. Whilst the Council Tax bill data set is highest due to ease of access, online banking allows for full verification for a Claimed ID and would score higher in terms of use.

**Alternative ID-Proofing techniques:**

Rank	Proofing Technique/ ID-Documents
1	Account Log-in controlled environment
1	Digital Verifiable Credential issued by Authoritative Source with Digitally Signed Certificate
2	API validation against authoritative source - yes / no response
3	Richer data sharing APIs <sup>2</sup>

**This report suggests the following next steps:**

- Further investigation into enabling access to the priority data sets is undertaken.
- Further investigation into the appropriate alternative ID-Proofing techniques for each dataset.
- The statistics are used as a benchmark to monitor progress of the size of the UK ID Challenged population size on an annual basis.

We recognise that digital identity inclusion is a complex field and whilst we are looking at it from a socio-economic perspective, there are other angles that are important but fall outside of the scope of this report. For example, we have not considered aspects such as disabilities or ethnicity.

<sup>2</sup>An API, or "application programming interface", is a set of protocols that allow the transfer of data from one system to another.

# 1.

## I Introduction

There is a proportion of the UK population who will not be able to verify their identity as they lack traditional ID-proof documents or credit history to do so. For this group alternative ID-proofing documentation combined with alternative proofing techniques and data sets were assessed. There are a number of considerations regarding risk and mitigation of alternative ID-proofing techniques and data sets considered in this paper.



**There are two main challenges for users:**

- 1 No ID documents or data history
- 2 Can't get a digital ID due to digital accessibility

This document looks at the first challenge (those without passport or driving license and lacking data history) and how data might help users in this situation. Although the scope of this paper does not include vouching and assisted digital, these methods will be required for some people. Assisted digital support includes access through channels such as telephone, face to face and help from others to use online services.

**Hypothesis:**

Access to identity services will be improved by introducing new data sets into the ecosystem and leveraging alternative proofing techniques.

**This report considers the below:**

- Definition of the ID Challenged and estimated number of people (in collaboration with Experian)
- Prioritised Data sets
- ID-Proofing techniques and risks
- Map alternative ID Proofing techniques to data sets
- Priority Proofing Techniques for the ID Challenged

For the purposes of this document the following versions of guidance documents are referred to:

- **GPG45:** UK Government Digital Service - *How to Prove and Verify Someone's Identity - Good Practice Guide (GPG) 45, version 5.1*<sup>3</sup>

GPG45 will be used in this paper to refer to this guidance.

The term verification is used differently in different sectors. For the purposes of this paper, the following terms are used:

Term	Definition	How referred to in GPG45
<b>Validation</b>	Validating that a user exists based on documentary or electronic evidence	Evidence Strength and Validity
<b>Verification</b>	Verifying the person trying to create a digital ID is the person they are claiming to be	Verification

This discussion report makes a number of suggestions, which have the potential to improve access to identity services for the ID Challenged and are considered worth further exploration by the ID Steering Group.

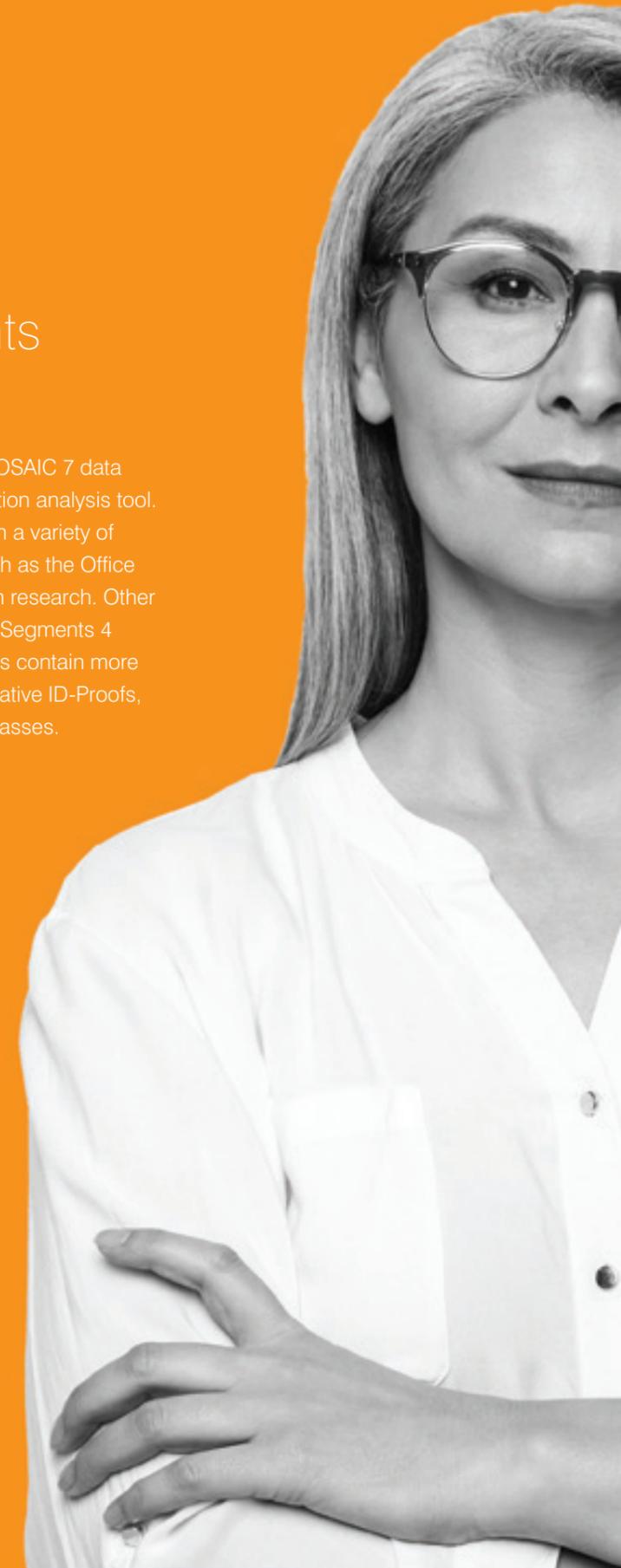
It also raises questions where further investigation is required to enable a more conclusive assessment of risk and reach of the data set or ID-proofing technique for the ID Challenged segments of the UK population.

<sup>3</sup> Good Practice Guide (GPG)

# 2.

## ID Challenged Segments & Methodology

To define the ID Challenged, we worked with Experian's MOSAIC 7 data sets. MOSAIC 7 is the latest version Experian's UK population analysis tool. It comprises data about the UK population that is based on a variety of different existing datasets created by national surveys, such as the Office of National Statistics' Census, as well as bespoke Experian research. Other data sets were considered, such as the Financial Strategy Segments 4 (FSS4). Experian MOSAIC 7 was chosen as these data sets contain more proxies (approximate matches) to identify those with alternative ID-Proofs, such as DWP letter of entitlement and concessionary rail passes.



## 2.1 THE UK CHALLENGE

There are over 8 million accounts signed up to GOV.UK Verify and 19 government services using Verify. According to the National Audit Report (NAO), GDS reported a verification success rate of 48%<sup>4</sup> (Feb 2019)<sup>5</sup> in the context of GOV.UK Verify. To be successful users will typically have the below ID proofs and be able to provide evidence of these online:

- Passport
- Driving Licence
- Online Banking + non-bank Credit Account

### According to GDS and IdPs, failure is in most cases is due to the following reasons:

- No photo ID
- Accessibility issues
- Do not own a smart phone (or the phone does not have a good signal, enough storage and is unable to download apps)
- Change of name
- Lack address history (at least 1 year)
- Lack suitable credit history
- Under 20 years
- Over 75 years
- IdPs have recorded high failure rates for the over 50s who do not have credit accounts
- Journey abandoned for several reasons (such as time out)
- Unsuccessful fraud attempt

Location is also seems to be a factor (especially those for those in the North East, Yorkshire, Wales).

<sup>4</sup> The verification success rate measures the proportion of people who succeed in signing up for Verify in a single attempt out of those who try. Those people have their identities successfully confirmed by a commercial identity provider. Some failures to sign up are not counted as part of this measure, such as the number of people dropping out before they finish their applications.

<sup>5</sup> National Audit Office (NAO) Report, March 2019. This report had the latest published data for GOV.UK Verify at the time of writing.

The DWP Confirm your Identity (CYI) service runs alongside GOV.UK Verify and successfully verifies a proportion of Claimed IDs, some of whom have failed to create a digital ID via the GOV.UK Verify route, but there is still a high proportion who fail.

Reported in March 2019 that Universal Credit is Verify's biggest government customer. Of the 70% of Universal Credit customers that attempt to sign up using GOV.UK Verify, 38% Universal Credit claimants can successfully verify their identity online<sup>6</sup>.

Those who do not attempt to sign up – or fail because they do not have the necessary documents or knowledge – are required to attend DWP offices to do a manual verification of identity using other documentation. DWP found the following reasons for failure to get a digital ID on GOV.UK Verify<sup>7</sup>:

30%

Neither driving licence, nor passport

14%

Thin file ("unbanked") and have no identity data history

14%

Could not answer knowledge-based verification (KBV) questions based on credit file data

Knowledge-based verification (KBV) questions are questions only the Claimed ID should know. Knowledge-based verification questions can be used to verify the individual is who they claim to be, against a claimed identity, which has already been validated in combination with other forms of ID-proofing. According to GPG45 knowledge-based verification challenges should be specific enough to be able to prove that that person is who they say they are.

Although it is not known how many Claimed IDs who couldn't answer knowledge-based verification questions through the Verify route get through CYI, we do know that 14% of people struggle through the Verify route due to a thin credit file. As CYI asks different knowledge-based questions based on information from their P60, or information that exists in other DWP data sets, this might suggest that those data sets offer potential to support those who are ID Challenged verifying their identity online.

DWP 2021/22 plans to iterate and add more data sources to further improve the outcome for citizens.

<sup>6</sup> National Audit Office (NAO) Report, March 2019. This report had the latest published data for GOV.UK Verify at the time of writing.

<sup>7</sup> These figures are pre-Covid

## 2.2 HOW WE DEFINED THE ID CHALLENGED SEGMENTS

The ID Challenged are found across the whole population, in all socio-economic demographics. Three steps were taken to define those who are ID Challenged:

### Step 1 - Filter

UK Households were ranked with higher propensity to **have neither driving license, nor passport.**

### Step 2 - Focus

Key drivers were identified to segment the ID Challenged Universe:  
**Age, Digital Literacy, Financial Resilience.**

### Step 3 - Features

**Key indicators (traits or variables) were identified to quantify how many people in each ID Challenged segment might have an alternative ID-Proof or be an alternative data set.** For example, the percentage of households who have DWP benefits, or those with qualifications (which would point to being in the Department of Education Learning Record Dataset).

## MOSAIC 7 Segment Long List – 22 Segments

All 66 Mosaic segments contain some ID Challenged individuals.

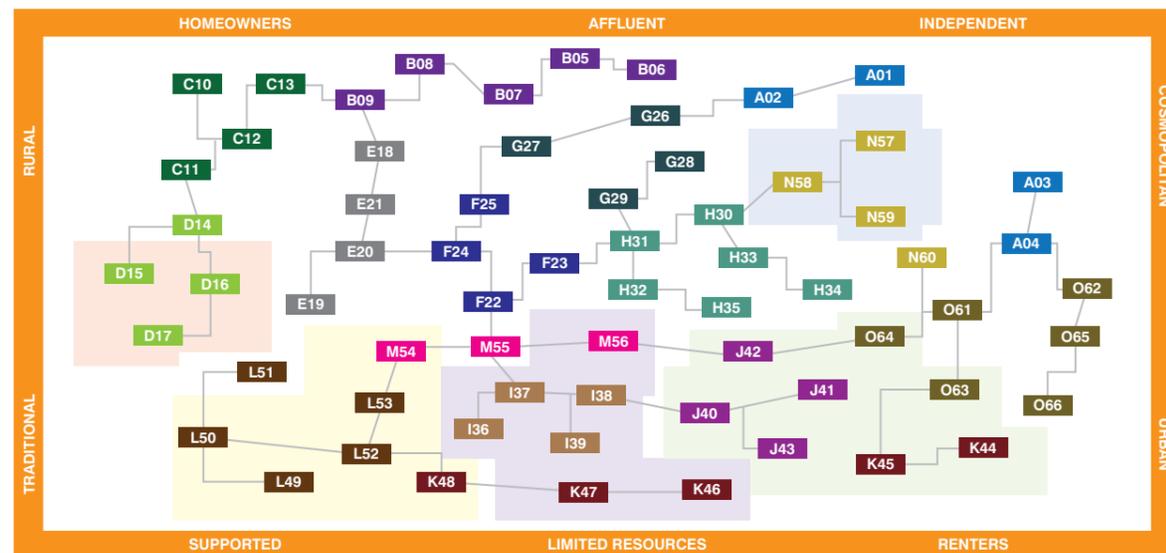
Using the key drivers (age, financial resilience and digital literacy), Experian analysts rank ordered the Mosaic segments to identify those with the largest proportion of the ID Challenged.

The top 22 segments were selected as a cut off to represent the segments of the UK population that contain the ID Challenged targeted for support as part of this analysis – these segments cover between 35% - 40% of the UK population. We looked at the overall adult UK population who have a high propensity to drive and have a driving license, have a higher propensity to travel abroad and have a passport, as well as considering the key traits of the proportion of the population who fail to verify their identity on GOV.UK Verify.

Based on the findings from Identity Providers, the expectation was that some of the financially resilient segments in the 50-55 age group would be in the target ID Challenged segments. However, when we explored deeper, we found that there was a higher-than-average propensity to have a driving licence or travel abroad (Mosaic proxy for passport holders). As a result, we excluded these segments from the ID Challenged Universe. From the long-list, segments were clustered into smaller groups. See Fig. 2.3.

The result was 5 ID target ID Challenged Clusters.

Fig. 2.3. Clusters of ID Challenged groups



We did further analysis using ONS and DVLA statistics on young people (16-24 years) to enable the analysis of data sets with younger people in them (such as the Learning Record Service) as Experian analysis only included UK adults 18 years and above. As the Experian population estimates are based on household data, people in this group tend not to be named on Council Tax bills, rental agreements, etc. The 16-24 age group is not mutually exclusive as 18-24 year-olds will also be found in segments 1-5 above. See Appendix 2.

As previously noted, the ID Challenged are found across all segments of the population, in all socio-economic demographics. The first five segments are where the highest populations of ID Challenged exist. From the household estimates, Experian estimated the number of ID Challenged (those with neither driving license, nor passport) in each segment.

Segment	Population Count	ID-Challenged Estimate	% ID Challenged
Segment 1	1,865,213	287,429	15%
Segment 2	3,866,698	836,367	22%
Segment 3	6,010,278	1,238,117	21%
Segment 4	5,553,721	991,339	18%
Segment 5	2,332,320	291,074	12%
Rest of UK	31,024,818	2,283,145	7%
<b>TOTAL</b>	<b>50,653,048</b>	<b>5,927,471</b>	<b>12%</b>

**2.4 Step 3: Key Traits / Variables for ID-Proofing Techniques and Data Sets**

We created a list of variables to enable an estimation of the percentage of the ID Challenged in each Segment with a higher propensity to have an alternative ID-Proof or reside in a data set that might either validate or verify (or both) a Claimed ID exists or is bound to an identity.

For example, we can further cut the data to estimate how many people in each segment are likely to have a passport, DWP benefits letter, or be in the Learning Records Service Data Sets or have a mortgage account (as in the example below). In some cases we needed to assign a proxy in Mosaic (such as travel abroad for passport holders).

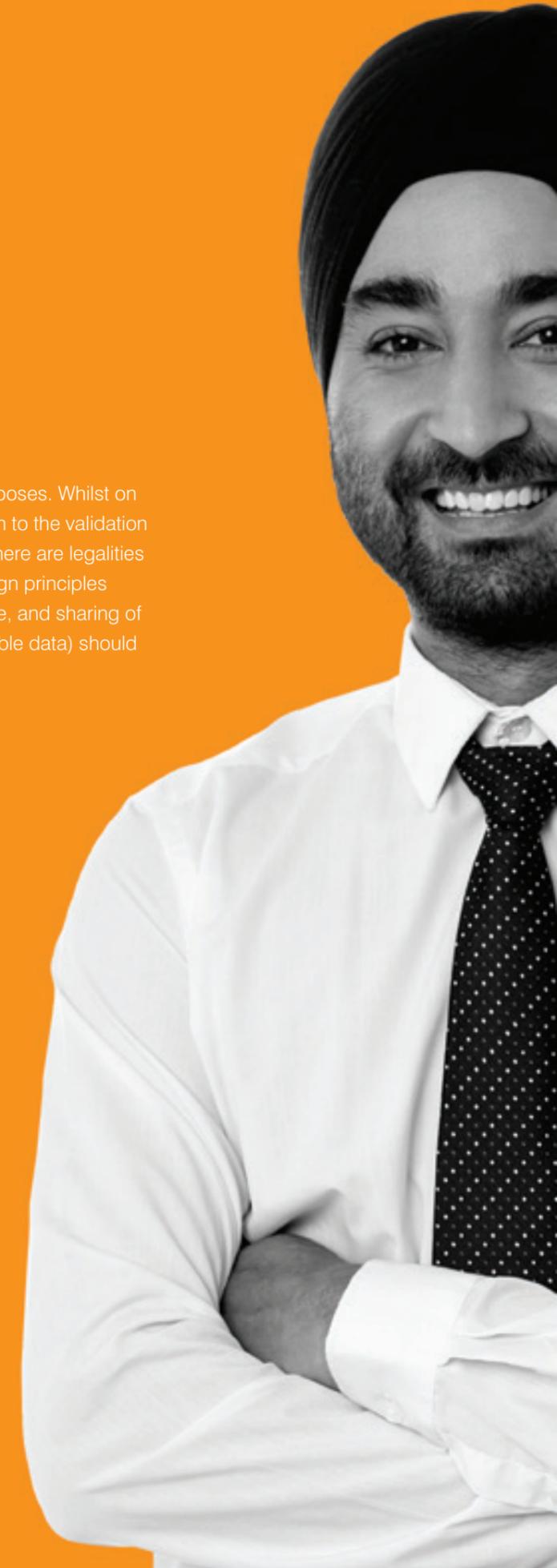
Sub Category	Mosaic Category	Document Type	Mosaic Variable	Seg 1%	Seg 2%	Seg 3%	Seg 4%	Seg 5%	UK Avg
Accommodation	Tenure	A mortgage account (including buy to let mortgage accounts)	Owned	58%	40%	40%	15%	68%	64%
		A rental or purchase agreement for a residential property	Rented	12%	6%	7%	57%	21%	18%
	Benefits	LA Letter - Social Housing	Housing Benefits	14%	18%	17%	15%	9%	11%
Household	Marital Status	A marriage or civil partnership certificate	Married	41%	42%	30%	11%	29%	39%

Fig. 2.4. An example of Variables (and Mosaic proxies)

# 3.

## | Data Sets

The data sets considered here are for “exploratory only” purposes. Whilst on the face of it the data sets look like they might offer a solution to the validation aspect of the user’s journey towards verifying their identity, there are legalities about accessing and/or sharing information. Privacy by design principles should be adhered to protect consumers and where possible, and sharing of signals, such as in Method 1 (rather than personally identifiable data) should be aimed for to protect the end user in line with GDPR.



The below data sets were considered and ranked in order to highlight those with the highest potential and value in exploring further. For details about the data sets, see section 4.1.

**Data sets were prioritised using the following criteria:**

1. ID Challenged segment counts in the data set
2. Data set enabled validation & verification
3. GPG45 strength scores
4. Level of technical difficulty (existing APIs, that enable the transfer of data from one system to another)

Risk was also assessed as was political and legislation level of challenge. Some data sets would require legislation change and others would require a policy change to be used for identity purposes.

Data Sets from the below organisations were considered. Data sets were assigned proxies within Mosaic to enable estimations of counts within each segment. As segments are not mutually exclusive, counts cannot be added up.

### 3.1 Data Sets which might help the ID Challenged

**Data sets have been ranked in the below order:**

1. NHS Patient Number
2. National Insurance Number
3. Council Tax Bill
4. Online Banking
5. Learning Records Service (LRS)
6. Council Tax Reduction Scheme
7. Occupational Pension
8. Deferred Pensions
9. Concessionary Travel Pass
10. NHS Staff
11. National Fraud Initiative (NFI), for further exploration

We anticipate that some of the DWP Benefits (DWP) Data Sets will offer a high reach to a large proportion of the ID Challenged population, but further investigation is needed into each of the data sets to determine those most helpful to the ID Challenged.

Please note that for some data sets not all the information was available at the time of writing. In some cases, attributes relating to the data in the data set were not known and so strength scores could not be assessed. In other cases, the appetite to share data has not been assessed yet. And in other cases, such as the NHS staff data sets, no Mosaic proxy existed to enable an estimation of counts per/ segment. Where the score is not assessed, ranking is largely based on the counts and level of difficulty. In the case of NFI data sets, they have ranked high as some of the data sets contain very high population counts.

Over 30 data sets were considered. The below table lists the data sets which ranked highest based on the number of ID Challenged people they are likely to help.

Data Set (Ranked based on ID Challenged Reach)	Data Set Source	ID Challenged Coverage (estimated)	Ease of Access
1. NHS Patient Number	NHS	Most of ID Challenged	5
2. National Insurance No.	DWP	Most of ID Challenged	3
3. Council Tax bill	Local Authorities	2,140,508	1
4. Online Banking	Open Banking (OBIE)	1,979,564	2
5. Learning Records Service	DfE	1,200,000	4

Data Set (Ranked based on ID Challenged Reach)	Data Set Source	Evidence Strength	Validation	Verification	ID Challenged Coverage (estimated)
Online Banking	Open Banking (OBIE)	3	Y	Y	1,979,564
NHS Patient Number	NHS	2	Y	-	Most of ID Challenged
National Insurance No.	DWP	2	Y	-	Most of ID Challenged
Council Tax Bill	Local Authorities	2	Y	-	2,140,508
Learning Records Service (LRS)	DfE	2	Y	-	1.2M
Council Tax Reduction Scheme	Local Authorities	2	Y	-	791,977
Occupational Pension	Local Authorities	3	Y	Y	699,235
Deferred Pensions	Local Authorities	3	Y	Y	663,812
Concessionary Travel Pass	Local Authorities	2	Y	-	536,797
NHS Staff – Lower Access; No proxy	NHS	2	Y	-	Likely High proportion of 600,000
NHS staff – Higher Access; No proxy	NHS	3	Y	Y	Likely Low proportion of 800,000
DWP Benefits (DWP)	DWP (Multiple data sets)	2/3	Y	Y	22.8M <sup>9</sup>

Fig. 3.1. Data Sets in priority order based on subjective assessment and taking into consideration evidence strength and potential to verify or validate or both.

<sup>9</sup> DWP benefits statistics: February 2021

NHS Staff with higher access to NHS Systems, such as doctors, are likely not to be ID Challenged. Whereas those staff with lower access to NHS Systems, such as hospital porters, are more likely to be ID Challenged.

For more information about the individual data sets, see Appendix 3.

It is worth noting that the Electoral Register is already available and is used in some financial services. From a policy perspective it can be used for ID purposes. It is not listed in GPG45 as the listing is self-declared and so open to fraud. As it might be used to assist over 2.4 Million of the ID-Challenged, consideration might be given to inclusion in GPG45.

### 3.2 ID Challenged – 3 Million Adults will struggle to get a Digital ID

The five ID Challenged Segments, as well as the sixth segment (Rest of UK), which shows the ID Challenged in the rest of the UK population, could be helped by using online banking as a part of the ID Proofing process, particularly the younger age groups in each segment.

Over 3.2 million people will likely be accounted for within the online banking data set across the five ID Challenged segments. A high proportion of ID Challenged people will have online banking<sup>9</sup> (the exact proportion is unknown).

Across the five segments, we have made a rough estimation that 1,979,564 ID Challenged people use online banking as well as 1,324,224 of the Rest of UK ID Challenged, making 3.3m people who might be able to be brought into the Digital ID ecosystem through the use of Open Banking. This estimate assumes that roughly the same proportion of people who have online banking in each segment is consistent with the proportion of ID Challenged who use online banking (this would benefit from further research to validate).

See appendices for details about the Open Banking Data Set.

Segment	Adults (In Segment)	ID Challenged (In Segment)	Using Online Banking (In Segment)	ID Challenged who may use Online Banking*	ID Challenged with Mobile / Smartphone
Rural Solitude	1,865,213	287,429	1,041,267 (56%)	160,960	1,265,445 (68%)
Vintage Veterans	3,866,698	836,367	1,879,551 (48%)	491,456	1,039,314 (56%)
Budgeting Families	6,010,278	1,238,117	3,192,307 (53%)	656,202	1,039,314 (56%)
Urban Renters	5,553,721	991,339	2,953,198 (53%)	525,409	1,508,469 (81%)
Community Culture	2,332,320	291,074	1,169,132 (50%)	145,537	1,403,176 (71%)
Total	19,628,230	3,644,326	10,235,445 (52%)	1,979,564	-
Rest of UK	31,024,818	2,283,145	18,114,936 (58%**)	1,324,224	-

Fig. 3.2. Proportion of the ID Challenged who might be helped with the OBIE Online Banking Data Set

\*The proportion of ID Challenged is unknown. Assumption: % who use online banking is roughly consistent with overall segment percentage.  
 \*\*Rest of Pop. Online Banking use has been estimated based on weighting up a survey.

<sup>9</sup> Online Banking and Personal Finance - This is the proportion of individuals in the segment who have used the Internet for any banking or finance related activity in the last 6 months.

It is important to note that the number of ID Challenged who use Online Banking in each segment is unknown. These figures are an estimation. We have assumed that the proportion of ID Challenged people who use Online Banking in each segment are reasonably consistent with the overall percentage who use online banking in each segment. It is possible that the ID Challenged might be less likely to use online banking, and so further research would be needed to validate this assumption.

Use of online banking apps was used as a key indicator of digital capability required to perform the tasks to create a digital ID without the help of others. This is highlighted in the ID Challenged segment pen profiles (see Appendix 2). There are further digital literacy indicators listed in the Target ID Challenged segment data dashboards in the appendices. Mosaic proxies were assigned for digital literacy and access, including use of online business services and E-commerce, social networks, and access to broadband and smart phones (detailed in Appendix 5).

### 3.3 ID Challenged Heat Map

The heat map shows the ID Challenged Segments which could be helped most with the below data sets. See fig.3.4 for the full table of data sets considered.

Data Set	Rural Solitude 287,429	Vintage Veterans 836,367	Budgeting Families 1,238,117	Urban Renters 991,339	Community Culture 291,074	TOTAL 3,644,326
Online Banking	160,960	491,456	656,202	525,409	145,537	1,979,564
NHS Patient Number						Most ID Challenged
National Insurance No.						Most ID Challenged
Council Tax Bill	203,751	384,728	581,914	713,764	259,055	2,140,508
Learning Records Service (LRS)	9,255	5,269	84,687	131,451	24,013	1.24M
Council Tax Reduction Scheme	57,485	250,910	284,766	178,441	20,375	791,977
Occupational Pension	54,611	175,637	235,242	178,441	55,304	699,235
Deferred Pensions	83,354	460,001	37,143	39,653	43,661	663,812
Concessionary Travel Pass	83,354	460,001	37,143	39,653	43,661	536,797

Fig. 3.3. Heat Map to show data sets which might help the highest number of ID Challenged

The below table details all of the data sets considered.

Within each of the ID Challenged segments there are people who possess a passport or driving license as it was not possible to identify the exact overlap (please see section 1 on methodology of calculation for ID-Challenged). As the ID Challenged was in part defined as people who have neither passport, nor driving license, those within the segments will not have both documents.

Data Set	Data Set Source	Evidence Strength	Population Coverage	ID Challenged Coverage: Segment Estimations (mutually exclusive)	ID Challenged Coverage:
Passport (Mosaic proxy: travel abroad)	HMPO	4	42M	Segment 1: 86,228 Segment 2: 259,273 Segment 3: 371,435 Segment 4: 396,535 Segment 5: 133,894	875,930
Driving Licence	DVLA	3	50M	Segment 1: 235,691 Segment 2: 585,456 Segment 3: 879,063 Segment 4: 684,023 Segment 5: 209,573	2,593,806
Student Loan Account (Mosaic proxy: University Degree)	Student Loan Company	3	20M	Segment 1: 6,898 Segment 2: 1,338 Segment 3: 18,819 Segment 4: 106,172K Segment 5: 23,053K (based on 18-35 year olds)	154,943
Extended Customer Attribute Data Set MOSAIC Proxy – Online Banking	Open Banking (OBIE)	3	50M (based on 76% of UK population OB penetration, 2020 figure)	Segment 1: 160,960 Segment 2: 491,456 Segment 3: 656,202 Segment 4: 525,409 Segment 5: 145,537	1,979,564

Data Set	Data Set Source	Evidence Strength	Population Coverage	ID Challenged Coverage: Segment Estimations (mutually exclusive)	ID Challenged Coverage:
Learning Records Service (LRS)	DfE	2	28M	Segment 1: 9,255 Segment 2: 5,269 Segment 3: 84,687 Segment 4: 131,451 Segment 5: 24,013 Segment 6: 1M	Segments 1-5: 254,675  Young Home Sharers: Roughly 1M
Birth Certificate	GRO	2	67M (9M digital)	None in ID Challenged segments as digitised to 12 years of age.	0
National Insurance Number	HMRC	2	54.5M	54.5M <sup>10</sup>	-
HMRC Tax Bill (Mosaic Proxy: Company Director)	HMRC	2	-	Segment 1: 8,622 Segment 2: 8,363 Segment 3: 37,143 Segment 4: 29,740 Segment 5: 26,196	110,064
DWP Benefits (DWP) No Mosaic proxy <sup>11</sup>	DWP (Multiple data sets)	2	1.84M (based on NFI benefits data set counts)	No proxy	22.8M

<sup>10</sup> Estimated from ONS mid-2018.

<sup>11</sup> Statistic source: [GOV.UK](https://www.gov.uk)

Data Set	Data Set Source	Evidence Strength	Population Coverage	ID Challenged Coverage: Segment Estimations (mutually exclusive)	ID Challenged Coverage:
NHS Patient Number (NHS Letters – correspondence with NHS Number)	NHS	2	67M	-	60M
NHS Staff Higher Access No proxy	NHS	3	800,000	Low proportion as ID Rich	-
NHS Staff Lower Access No proxy	NHS	2	600,000	Higher proportion; no proxy Higher proportion in Segments 1,3,4,5	-
Council Tax Bill (Mosaic Proxies: Rented & Owned added within each segment)	LA	2	25M	Segment 1: 201,199 Segment 2: 384,728 Segment 3: 581,763 Segment 4: 713,763 Segment 5: 259,055	2,140,508
Housing Benefit (Mosaic – housing benefit)	LA	2	-	Segment 1: 40,240 Segment 2: 150,546 Segment 3: 210,479 Segment 4: 148,700 Segment 5: 26,196	427,461
Occupational Pensions (Mosaic Proxy: Occupational Pension) As based on a proxy, this is an ID Challenged population estimation across all UK occupational pensions.	-	2	20M	Segment 1: 13,652 Segment 2: 86,062 Segment 3: 2,352 Segment 4: 5,353 Segment 5: 7,189	114,608
Railcard (Mosaic proxy: train)	National Rail	2	-	Segment 1: 11,497 Segment 2: 50,182 Segment 3: 74,287 Segment 4: 69,393 Segment 5: 26,196	231,555
Oyster Card (Mosaic proxy: underground travel)	Oyster	2	-	Segment 1: 11,497 Segment 2: 33,454 Segment 3: 74,287 Segment 4: 89,220 Segment 5: 37,839	246,297
Concessionary Bus Pass Older Person (Mosaic Proxy: retired)	LA issued	2	-	Segment 1: 83,354 Segment 2: 460,001 Segment 3: 37,143 Segment 4: 39,653 Segment 5: 43,661	536,797

<sup>10</sup> Estimated from ONS mid-2018.

Data Set	Data Set Source	Evidence Strength	Population Coverage	ID Challenged Coverage: Segment Estimations (mutually exclusive)	ID Challenged Coverage:
<b>National Fraud Data Sets</b>	NFI (Multiple data sets – see below)	Unable to score without info.	-	-	-
Electoral Register Estimation based on 2019 turn out of 67.3%	Local Authorities	-	47.4M	Segment 1: 193,439 Segment 2: 562,874 Segment 3: 833,252 Segment 4: 667,171 Segment 5: 195,892	2,452,628
Council Tax Mosaic proxy – mortgage & rental agreement	Local Authorities	-	28.8M	Segment 1: 203,751 Segment 2: 384,728 Segment 3: 581,914 Segment 4: 713,764 Segment 5: 259,055	2,143,212
Council Tax Reduction Scheme Mosaic proxy – Council Tax Benefit	Local Authorities	-	4.4M	Segment 1: 57,485 Segment 2: 250,910 Segment 3: 284,766 Segment 4: 178,441 Segment 5: 20,375	791,977
Occupational Pension Mosaic proxy – Occupational pension	Local Authorities	-	4.7M	Segment 1: 54,611 Segment 2: 175,637 Segment 3: 235,242 Segment 4: 178,441 Segment 5: 55,304	699,235
Deferred Pensions (Very rough estimation based on Mosaic proxy – retired)	Local Authorities	-	2.26M	Segment 1: 83,354 Segment 2: 460,001 Segment 3: 37,143 Segment 4: 39,653 Segment 5: 43,661	663,812
Concessionary Travel Pass Mosaic proxy – retired	Local Authorities	-	9.7M	Segment 1: 83,354 Segment 2: 460,001 Segment 3: 37,143 Segment 4: 39,653 Segment 5: 43,661	536,797

>>

Data Set	Data Set Source	Evidence Strength	Population Coverage	ID Challenged Coverage: Segment Estimations (mutually exclusive)	ID Challenged Coverage:
Blue Badges / Parking permits No Mosaic proxy	Local Authorities	-	29M	Highest proportion in Segments 1 & 2. Counts not possible as no Mosaic proxy.	-
Housing Rents Housing association tenants. No proxy.	Local Authorities	-	2.6M	No proxy.	-

See Annex for % breakdown across variables with Mosaic proxies (compared to UK average %)

Fig. 3.4 Data Sets and estimated ID Challenged segment counts and totals.

Companies House – at present the data is not robust enough for ID Verification purposes, but they may be a contender in the future.

**Detailed Criteria for Further Data Set Prioritisation:**

- Level of technical difficulty (APIs available)
- Number of ID Challenged in data set
- Overall Strength Score
  - ID Validation (depends on proofing method) 1-4
  - ID Verification (score: 1-4)
  - Evidence Strength (score: 1-4)
- Granularity of Knowledge Based Questions
- Tried & tested and in use
- History over time (score: 1-4)
- Data maintenance
- Risk
- Political level of challenge considered after ranking as an overlay. In some cases political challenge was defined by a need for a legislation change to extend the use of data sets to be queried for the purpose of ID Proofing (i.e. the National Fraud Initiative (NFI) data sets). In other cases there was either a need for policy change (i.e. Learning Records Service) or both legislation and policy underpinning legislation would be needed to safely manage the risk to the Claimed ID and the Data Owner/ Data Controller.

Where ID validation, historical check & ID verification is deemed to be technically possible, it may not be possible politically or ethically (i.e. NHS protection of vulnerable citizens). The prioritisation approach takes these factors into account. However, political challenge will often depend on the nature of the data set query and possible outcomes from such a query. For example, queries about medical conditions are not appropriate to disclose as they put vulnerable people at risk and would require interpretation by medical experts.

**Assumptions used for risk and strength assessment:**

- Assessment is based on the authoritative source of all of these data sets having verified the ID of the Claimed ID in order for them to be in the data set to begin with.
- In the case of Learning Record Service (LRS), that would be the school/college that the person got their qualification from. The level of ID verification may score lower than others depending on the data set.

Please see Appendices for details about the data sets.

Data Sets	ID Validation Check	Historical Check	ID Verification	Risk Level To ID Issuer	Risk To Citizen	Options (& possible uses of the data: i.e. KBV Questions)	Strength (Based on GPG45)	Reach: ID-Challenged Segments (above UK average)
<b>DWP – multiple Data Sets</b> (including universal credit, etc.)	Y (score: 2)	Y In some cases. Years vary depending on data set	N	Medium	(Y/N response) Low Richer APIs Gov: Medium Private: High	Y/N Response & Richer APIs e.. Static Knowledge Based Questions - Tried and tested (Confirm Your ID) - Further work might assess individual data sets against population statistics	<b>Score 1</b> N / A / DOB <b>Possible Score 2</b> If NINO level of duplication is acceptable	<b>Segments 3,4,5</b> Counts per dataset possible where Experian proxies for data sets exist See <i>dashboard</i>
<b>DWP – CIS</b> (Customer Information System Data Set)	Y	Y (up to 3 years)	N	Medium	Low (If carried out by authoritative source)	Current use: Aliveness Check & by Las For future consideration: - Static Knowledge Based Questions - ideal and tested (Confirm Your ID) - Attributes enable ID Validation - ID Verification possible for those who are on the system and have been verified by DWP JCP officers	<b>Score 2</b> DOB level verified (i.e through 2 accepted document types, etc.) Fraud marker	<b>Segments 1,2,3,4</b> Above the UK average counts for social housing, which could be cross-checked against CIS
<b>Learning Records Service</b> API validation against authoritative source – yes / no response	Y	In some cases	N	Low	Low	- Data from 2010 (oldest now 25 years – will change in future) - Students registered at a school from 14 years - Static KBQs - Not comprehensive data	<b>Score 1 =</b> Name / DOB	<b>Primarily Segments 3, 4 &amp; 6</b> See counts below
<b>GRO – Birth Register</b> API validation against authoritative source – yes / no response	Y	N	N	Low	Low	- Data from 1837: England & Wales only - Name, DOB (address at birth) - No unique identifier	<b>Score 1</b> Name / DOB	<b>Segments 3,4,5</b> Not as useful for segments 1 & 2 as high % in Scotland
<b>Student Loan Account</b> API validation against authoritative source – yes / no response	Y	Y In some cases Years vary	Y	Low	Low	- Name, DOB, Address - Unique Identifier (Account number) - GPG45 / AML verification	<b>Score 3</b>	<b>Segments 4 &amp; 5</b> have the highest proportion, but all segments are well below the UK average

>>

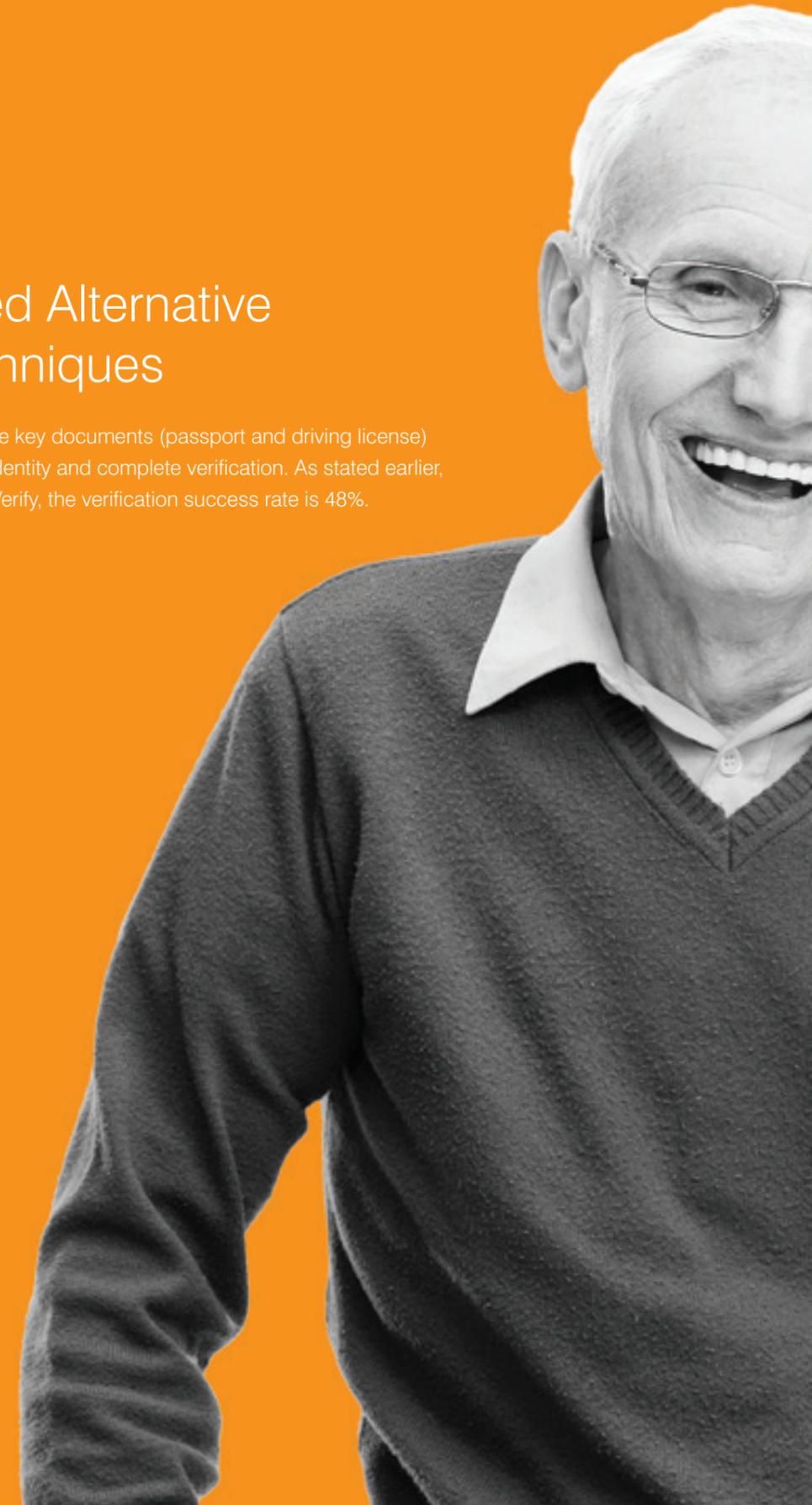
Data Sets	ID Validation Check	Historical Check	ID Verification	Risk Level To ID Issuer	Risk To Citizen	Options (& possible uses of the data: i.e. KBV Questions)	Strength (Based on GPG45)	Reach: ID-Challenged Segments (above UK average)
NHS – Patient	Y	Y Level of difficulty – HIGH	Y (only with a vouch)	Medium	High	- KBQ – ONLY about non-medical / sensitive data. Open conversation - KBQ – not possible based on medical records. Highly unlikely - Level of difficulty too high	Score 2 = if the NHS number counts as a unique identifier Score 1 = if the NHS number not deemed an adequate UI	All Segments
NHS – Staff: Higher level Access	Y	Y	Y	Medium	Low	- 800,000 Staff – full ID check aligned to GPG45 - KBQ possible - Validation & Verification possible	Score 4	Unlikely to meet ID Challenged segment profiles
NHS – Staff: Lower Grade Access	Y	Y	N	Medium	Low	e.. Estimated 600,000 (porters, etc., and would include thin file) - Hospitals do full ID check aligned to GPG45 Score 3 - Photo ID – bound to actual person not reliable for this staff group	Score 3 Primary Care Secondary Care might not align with score 3	Unlikely to meet ID Challenged segment profiles
NFD – 22 Data Sets (concessionary bus pass, blue badge, etc.)	Y	Y In some cases	N In some cases More detail about data sets required	Low	Low	See breakdown of counts. Attributes not available at this stage Full coverage	Attribute data not available at this stage	See counts in NFI Data set table
Private Sector – Travel Oyster Card, National Rail Card	Y	Y In some cases & if data kept after expiry	N	-	-	Needs further investigation	-	-
Pensions / Occupational e.g. Armed Forces, NEST, Civil Service, NHS	Y	Y	N	-	-	Needs further investigation	-	-
Extended Customer Attributes Data Set	Y	Y	Y	Medium	Low	- KBQ possible - Validation & Verification possible	Score 3	Segments 4 & 5 see counts in data set table

Fig. 3.5 Data Set Prioritisation Assessment Table

# 4.

## ID Challenged Alternative Proofing Techniques

To recap, the ID Challenged lack the key documents (passport and driving license) and credit history to validate their identity and complete verification. As stated earlier, based on GDS statistics of use of Verify, the verification success rate is 48%.



### 4.1 Alternative ID-Proofing Techniques

The below methods could perform part of the ID-Verification process (validation only), or all of it (Validation and Verification) – in the case of Methods 3, 8 and 9.

METHOD 1: API validation against authoritative source – yes / no response

METHOD 2: Richer data sharing APIs

METHOD 3: Account Log-on in controlled environment

METHOD 4: Digital Verifiable Credential issued by an authoritative source

METHOD 5: Letters with added security features (i.e. QR Code)

METHOD 6: Manual validation against authoritative source

#### A Note on the Manual Vouch

For many citizens who fall into the ID Challenged group, a manual vouch may be the only way they can verify their identity. The Manual Vouch (in-person) has not been considered in this report as it is a traditional, universal method of verifying a claimed ID and a common fallback method rather than alternative ID-Proofing technique. A vouch is defined in GPG45 as ‘a declaration from someone who knows the claimed identity.’ A person vouches for another person by claiming that they know them to match the claimed identity. A vouch is another type of evidence about a claimed identity. Vouches have been combined with various forms of ID validation to complete the full ID verification journey for many years. The Manual Vouch can be used in collaboration with several of the validation methods listed above which perform only the validation part of the customer journey, such as API Validation against an authoritative source eliciting a yes / no response.<sup>12</sup>

### 4.2 Alternative ID-Proofing Techniques: Assessment of Risk

We prioritised the following six alternative ID-proofing techniques based on a criterion which includes the following:

- Whether the Issuer has a contract with a 3rd party on behalf of the Data Subject
- Whether the method validates only or validates and verifies
- Risk Assessment to the ID recipient, ID issuer, data subject
- Ease of Use for customer

We also took into consideration risk (to issuer, recipient & ID Subject).

See Fig. 4.2. for risk assessment.

ID-Proofing techniques using method 3 (Account Log-on in controlled environment), method 4 (Digital Verifiable Credential issued by authoritative source), and Methods 1 & 2 (API Validation) ranked higher. Methods 3 and 4 validate and verify (VV), whereas Method 1 only validates (V), and depending on the nature of the richer API data, in many cases, Method 2 will also only validate – unless the knowledge-based questions were based on a pre-verified source. See Fig. 4.1.

<sup>12</sup> List of people who can perform a manual vouch.

### 4.3 Prioritised Alternative ID-Proofing Techniques

The six methods were ranked accordingly:

Rank	Proofing Technique/ ID-Documents	Method	Validation (V) / Verification (VV)
1	Account Log-in controlled environment	M3	VV
1	Digital Verifiable Credential issued by Authoritative Source with Digitally Signed Certificate	M4	VV
2	API validation against authoritative source - yes/ no response	M1	V
3	Richer data sharing APIs	M2	V (VV in some cases)
4	Letters with added security features (digital or physical)	M5	V
5	Manual validation against authoritative source	M6	V

Fig. 4.1 ID-Proof Ranking

The top two methods have ranked highest as they validate and verify an identity. However, as API interfaces enable ease of use for the customer (if service design is consumer-centric) and if they already exist are reasonably cost effective to deploy (they are expensive to build), APIs rank highest among the validation only methods.



Proofing Technique / ID Documents	Method	Issuer has a contract with 3rd Party on behalf of data society	Data "Integrity" Risk (risk to recipient)	Validation (V) / Verification (VV)	Data Breach Risk (risk to issuer / sender)	ID Fraud Risk (risk to the recipient)	Data Breach Risk to Citizen
API validation against authoritative source - yes / no response	M1	Yes	Very low	V	Medium	Low	Low
Richer data sharing APIs + Confidence level / Evidence score	M2	Yes	Low as from source (as long as source reliable)	V	Medium (Data compromise breach leak risk to the service)	Medium	Medium
Account log-in in controlled environment (ie OB-TPP contract)	M3	Yes	Low	VV	Low	Low	Low
Digital Verifiable Credential issued by Authoritative Source with Digitally Signed Certificate	M4	No contract (but in the system)	High (unless storage location is strongly linked to recipient to make it low)	VV	Medium	Low	Low
Letters with added security features (ie QR Codes)	M5	No	High	V	High	High - easier to counterfeit a letter	High
Manual validation against authoritative source	M6	No	Low	V	Medium	Medium	Medium

Fig. 4.2 – ID-Proof Risk Assessment Table (VV – validates and verifies. Single V – only validates)  
Risk Levels: Very Low, Low, Medium, High

# 5.

## Summary of Findings and Next Steps

### Summary of Findings

The findings in this document can be summarised as follows:

- There are an estimated **5.9 Million** individuals in the UK who are ID Challenged.
- The introduction of additional data sets to the ecosystem will likely increase access to government and private sector services.
- Alternative ID-Proofing techniques might support some of the ID Challenged to get a Digital ID, giving them increased access to online services.



**Priority alternative ID-proofing techniques are:**

Rank	Proofing Technique/ ID-Documents
1	Account Log-in controlled environment
1	Digital Verifiable Credential issued by Authoritative Source with Digitally Signed Certificate
2	API validation against authoritative source - yes / no response
3	Richer data sharing APIs (particularly those that can verify as well as validate a Claimed ID)

Although some API Methods can only validate, in some cases combined with a vouch, the two API methods offer an effective route to verify an identity and so they rank the highest in the group that only validates a Claimed ID in third place. It should be noted that those APIs which can both validate and verify would rank higher (in second place).

**The top priority data sets at the time of writing are:**

- NHS Patient Data Set
- National Insurance number Data Set
- Council Tax Bill Data Set
- Online Banking Data Set

It is important to note that the number of ID Challenged who use Online Banking in each segment is unknown. The figure of 1,979, 564 people who may use online banking across the five ID Challenged segments is based on the invalidated assumption that the proportion of people who have Online Banking in the segment are reasonably consistent with the ID Challenged who may use online banking in the segment. It is possible that the ID Challenged might be less likely to use online banking. Further research would be needed to validate this assumption.

**Data Sets which would be worthwhile assessing in more detail:**

- National Fraud Initiative (NFI) Data Sets, which are available today for ID checking purposes (it was not possible to assess strength scores with the data available)
- DWP Customer Information System (CIS)
- DWP Data Sets (further work required to identify & prioritise data sets)
- Learning Records Service (LRS) Data Set (further work required directly with DfE to assess policy and attributes in more detail. At the time of writing, DfE validation was not available).

**Data Sets to reassess in 2-3 years in terms of counts in the ID Challenged segments:**

- GRO: check if legislation change has been reformed to include private sector)
- Learning Records Service (LRS): check if Department of Education policy has changed or is likely to.

The GRO data set has the potential to be useful to validate the younger segments of the ID Challenged population in years to come – if both funding is available to continue the digitisation drive and legislation is addressed. On this point it is noted that the current purpose is not being driven by ID validation for the purpose of obtaining a digital ID, so engagement now by the Identity community is important to influence the direction of travel of the GRO digitisation programme if it is to be future proof and extendable for ID proofing purposes.

Although the scope of this paper did not include vouching and assisted digital, these will be required for some people.

**Next Steps:****The OIX Inclusion Steering Group to assess how to apply the research findings in the following areas:**

- Agree Data Sets to explore further
- Decide which ID-Proofing Techniques to investigate further
- Further exploration into DWP Data Sets, particularly the CIS data set
- Further exploration into National Fraud Initiative Data Sets and attributes to assess strength scores
- Stay close the GRO developments. If legislation is passed and births backdated, this will likely be a useful data set in the future
- Further research to validate the number of people in each segment who use, or may use, online banking
- Consider using the segment counts as a benchmark to monitor on a regular cadence to monitor the movement of the ID Challenged segments over time.



## APPENDIX 1 – Methodology

### ID Challenged segment focus drivers

#### Driver 1. Age:

- Retired 65+  
Pre-retirement 55-65
- Mid-life: 25-54
- Young Adult: 18-24
- Teens: 13-17

#### Driver 2. Digital Literacy: *Literate & Not Literate*

- Digitally Literate – have access and devices/ digital skills capability & online banking services (need all of these characteristics to fall into this group)
- Not Digitally Literate - no access/ capability/ devices & use online banking / financial services.

Some digital characteristics were added to traits (such as what phone type is commonly used – to assess the likelihood of a person having NFC functionality on their mobile phone). Near Field Communication (NFC) technology helps users to make secure transactions and exchange digital content. NFC used to read information, including the holders photo, from an e-Passport chip can verify a Claimed ID in collaboration with a selfie of the user.

#### Driver 3. Financially Resilience: *Resilient & Not Resilient*

Financial Literacy was defined in three segments: financially resilient, not financially resilient in work and not financially resilient out of work.

**Financial Resilient Definition:** Have the ability to respond to financial change.

- Savings
- Credit history/ can access mainstream credit
- No signs of water/ fuel poverty & can pay household bills

**Not Financially Resilient Definition:** Cannot respond to financial change.

- No savings
- No consumer borrowing / can't access mainstream credit
- Water/ fuel poverty / can't pay household bills

In Work: Gig Economy / Cash workers / Houses in Multiple Occupation (HMO) residents (who cannot produce a Council Tax bill as proof of ID).

**Out of Work :** Benefits / not on benefits.

### Level of accuracy of estimates of id challenged individuals

In terms of the level of confidence one can have in the estimates of the numbers of ID-Challenged individuals, confidence is difficult to ascertain as there are numerous steps in the derivation of ID challenged, from the inferred extrapolation of the survey data through to the use of “not having travelled abroad in the last 12 months” as a proxy for not having a passport, which cannot be quantified.

The combination of attributes to define ID-Challenged is: “*have not travelled abroad in the previous 12 months*” AND “*do not have a driving licence*”. The volume estimates for the number of individuals meeting both criteria were based on a survey carried out by TGI where panellists were asked these two questions. Their answers were then aggregated by MOSAIC Type. We have inferred that the general population would follow a similar trend depending on which Mosaic Type they are classified as.

For example, Experian asked 10,000<sup>14</sup> individuals if they have a driving licence or a passport. They then identified the individuals who have neither and aggregated the sum of those individuals by Segment (i.e. 10% segment A have neither, 25% segment B have neither, and so on). Experian then made the assumption that 10% of all UK individuals in Segment A have neither a passport or driving licence and are therefore ID-Challenged.

There were limits to the data sets for our purposes. Teens (13-17 years of age) were in our target groups. Experian can highlight households with people between 12-17 years, but cannot analyse minors below 18 years of age.

For the 18-22 years group ONS data was analysed. As MOSAIC 7 is household based it cannot offer insight into individuals. As many of this age group are living in shared accommodation with parents or houses of multiple occupancy (HMOs), they were lost in the data. For the purposes of estimating the number of ID Challenged individuals in the Learning Records Service Data Set, manual analysis using DVLA driving licence data and Office of National Statistics (ONS) data was used to calculate a rough estimation of the number of ID Challenged. Please see Appendix 2, Young Home Sharers for more detail.

In addition to defining the ID Challenged, we wanted to understand which sub-segments of the Target ID Challenged segments are digitally literate, what channels they use and alternative ID-Proofing documents they might have access to, as well as what government or private sector data sets, they may be validated against (proving the Claimed ID exists) or verified against (proving the Claimed ID is who they say they are). To do this we identified ‘traits’ and mapped these to the chosen data sets and ID-Proofing documents.

Identifying traits that define the ID Challenged enabled us to map to potential alternative ID Proofing techniques, data sources, and access channels for each segment – and to assess which alternative ID-proofing technique is most accessible (taking risk into consideration) for the sub-segment of the ID Challenged Universe.

<sup>14</sup> This is an example only and not the actual sample size.

## APPENDIX 2 – ID Challenged Segments & UK Maps

### Segment 1 Rural Solitude

There are an estimated 287,429 adults who are ID Challenged in this segment. Mostly age 66+, although not all. Many are in work, a lower than UK average number with qualifications, some own small properties but most are living in housing association accommodation in rural Scotland / Cornwall / Wales. 32% use online banking apps.

#### Rural Solitude | Portrait



Rural Solitude are households who live in affordable properties in village and countryside settings. Many residents are mature in age, but families with children are also included.

Those of working age are employed in the local economy in roles usually found close to home. Their intermediate and lower-level occupations earn below average wage that don't leave a great deal of spare at the end of the month. Older residents are retired and rely on modest pension incomes.

**Adults**  
**1,865,312**

**% of UK Adult Population**  
**3.7%**

**66+**  
475,267 / 25%

Age

**Baby Boomers**  
**1946 - 1964**  
690,122 / 37%

Generations

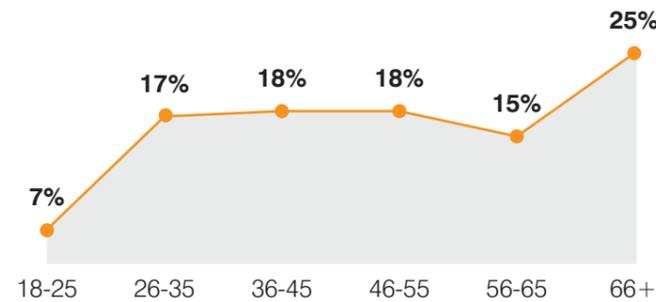
**Employed Full-Time**  
852,524 / 46%

Employment

**68%**  
1,265,445  
71% UK Coverage

Smartphone / Mobile

**Age Bands | % of Segment vs. UK Average**



**ID Poor**  
Experion ID Poor Estimate



**287,429** Adults

**Online Banking Apps**  
Digital Literacy



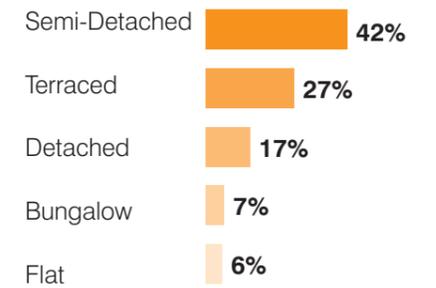
**606,116** Adults

#### Property

More likely to live in a bungalow or semi-detached properties in rural areas. Some own their properties and a large majority are housing association residents.

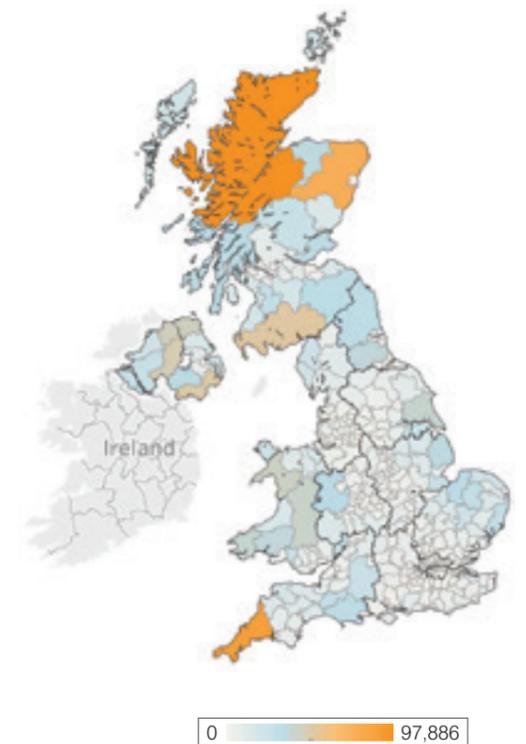
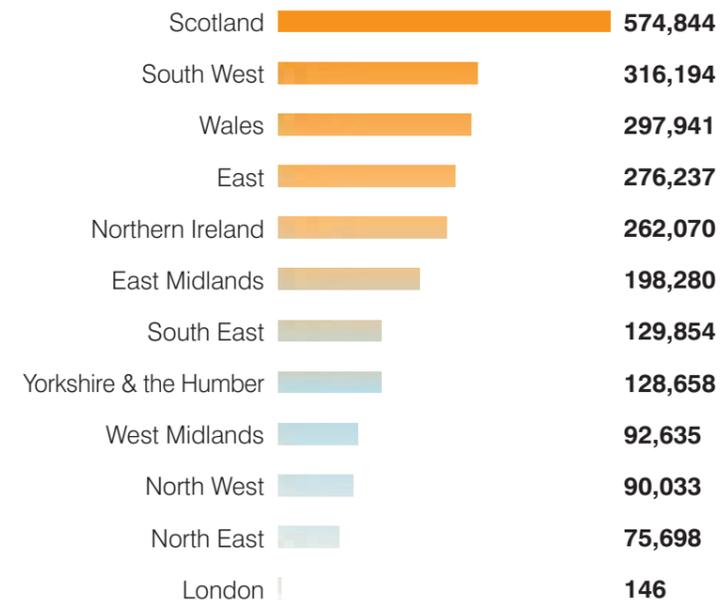
#### Education

High proportion with no qualifications and some with GCSEs. Less likely for rural solitude to study after GCSE level.



#### Rural Solitude | Geographic Profile

**Regional Profile | vs. UK Segment Average**



### Segment 2 Vintage Veterans

An estimated 836,367 adults are ID Challenged in this segment. Mostly age 66+, retired, no qualifications, rural and the majority live in housing association accommodation in Scotland / NW /NE of England. 31% use online banking apps.

#### Vintage Veterans | Portrait



Vintage Veterans are retired (or near retirement) individuals aged over sixty five who live in accommodation that is modestly sized.

Many of this group reliant on the state pension alone and some are eligible for additional benefits. Though their financial resources are very low, most are experienced at budgeting and match their expenditure to their income. Homeowners may have more to spare than those paying rent, though they have the cost of home maintenance. Some have modest savings and are able to run a small car.

**Adults**  
**3,866,698**

**% of UK Adult Population**  
**7.6%**

**66+**  
911,762 / 49%

Age

**Baby Boomers**  
1946 - 1964  
1,073,268 / 58%

Generations

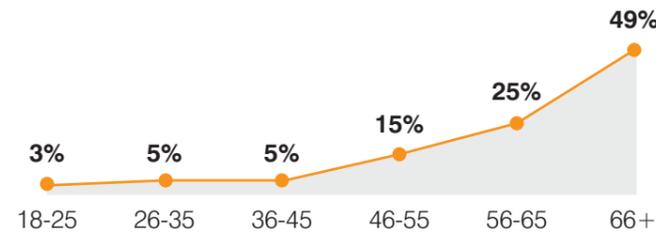
**Retired**  
1,034,499 / 55%

Employment

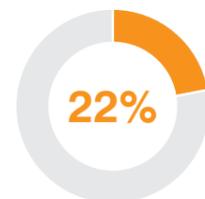
**56%**  
1,039,314  
71% UK Coverage

Smartphone / Mobile

**Age Bands | % of Segment vs. UK Average**

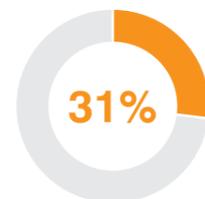


**ID Poor**  
Experion ID Poor Estimate



**836,367** Adults

**Online Banking Apps**  
Digital Literacy



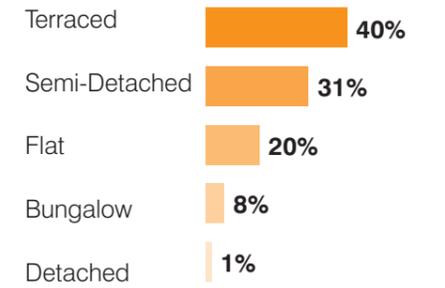
**1,198,824** Adults

#### Property

Highest prevalence of bungalows and also 1 bed terraced properties in suburban areas. Largest housing association population amongst all segments.

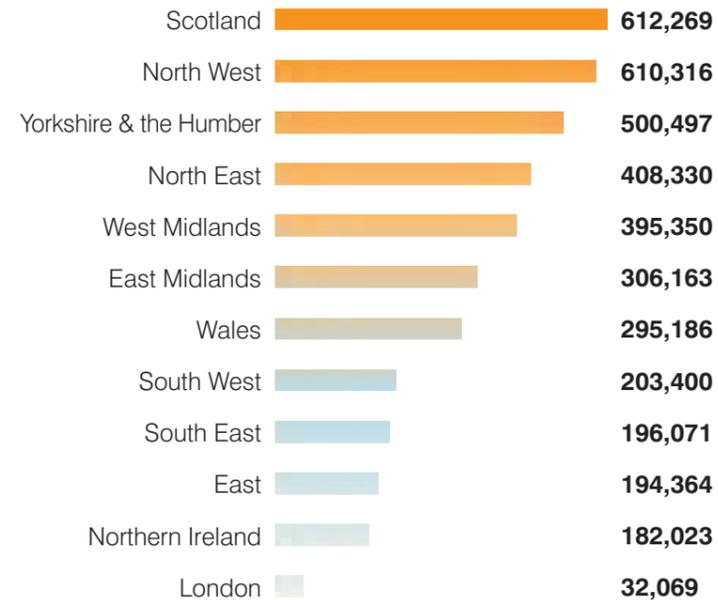
#### Education

Majority with no qualifications. Lowest segment going to university.



#### Vintage Veterans | Geographic Profile

**Regional Profile | vs. UK Segment Average**



0 99,263

### Segment 3 Budgeting Families

An estimated 1,238,117 adults are ID Challenged in this segment. Mostly age 26-45, in work, some with GCSEs, but very few with A'Levels. They are urban, many living in housing association accommodation in the North of England. About 35% use online banking apps.

#### Budgeting Families | Portrait



Budgeting Families are households who have limited incomes and budget carefully. Most adults are aged in their twenties, thirties and forties. Many live as couples often with children present.

Those adults in employment earn modest wages from a range of lower-level jobs. Not everyone works full-time - some work part-time or stay at home to look after children and some are studying or looking for work. With low disposable incomes and many priorities competing for cash, parents work hard to make their money stretch as far as possible. Short term loans are sometimes used to spread payments and some may use other expensive credit options.

**Adults**  
**6,010,278**  
 % of UK Adult Population  
**11.9%**

**36-45**  
 509,000 / 27%

Age

**Generation X**  
 1965 - 1980  
 767,416 / 41%

Generations

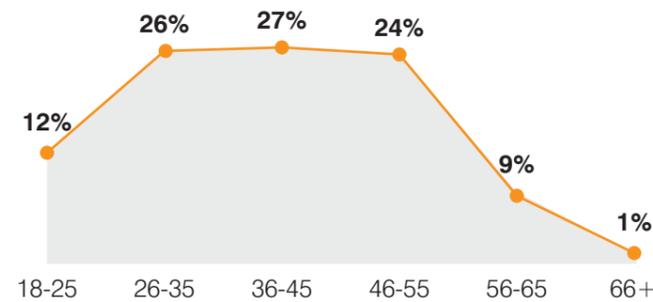
**Employed Full-Time**  
 944,178 / 51%

Employment

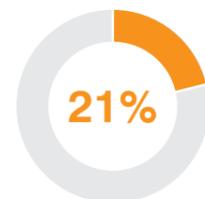
**81%**  
 1,505,353  
 71% UK Coverage

Smartphone / Mobile

Age Bands | % of Segment vs. UK Average



ID Poor  
 Experion ID Poor Estimate



1,238,117 Adults

Online Banking Apps  
 Digital Literacy



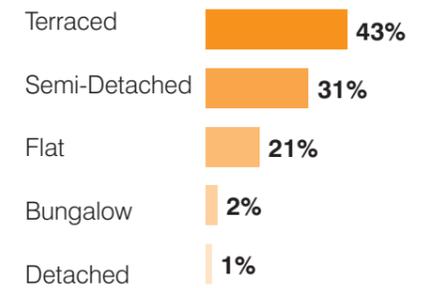
1,198,824 Adults

#### Property

Likely to live in terraced properties and 2-3 bedroom semi-detached houses in suburban areas. Low proportion of renters and a large percentage of housing association residents and frequently living with presence of children.

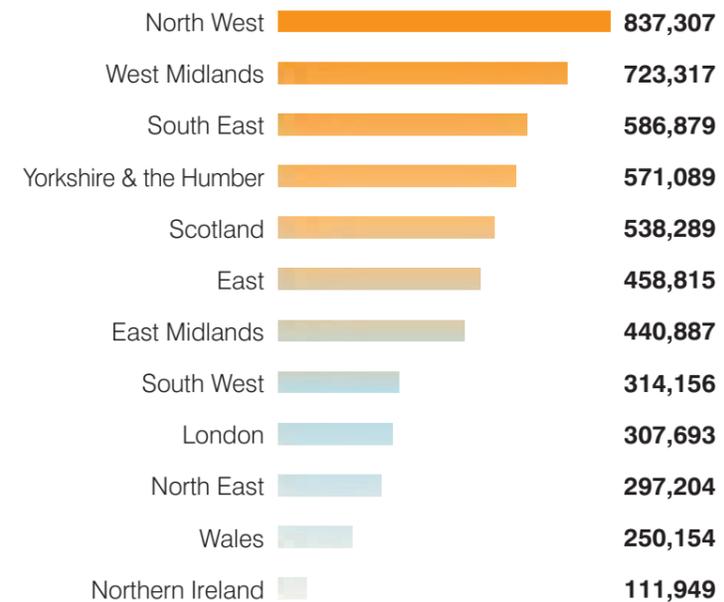
#### Education

High proportion reaching GCSE level as well as above average with no qualifications. Very low continuation of education after A Level.



#### Budgeting Families | Geographic Profile

Regional Profile | vs. UK Segment Average



0 150,895

### Segment 4 Urban Renters

An estimated 991,339 adults are ID Challenged in this segment. Mostly age 26-35, in work, some with qualifications & lower than average university goers. They are urban – the vast majority are living in London with some in Birmingham; there are many renters and a higher than average proportion live in housing association accommodation. About 35% use online banking apps.

#### Urban Renters | Portrait



Urban renters are young single people in their twenties and thirties who rent affordable living spaces. Some live alone while others share with housemates or partners.

Many residents work full-time and earn relatively low wages in technical, semi-routine or routine occupations. Some may be studying or looking for work. Levels of movements are high, and most residents have been living at their address for a few years or less.

**Adults**  
**5,553,721**

**% of UK Adult Population**  
**11.0%**

**26-35**  
628,503 / 34%

Age

**Millennials**  
1981 - 1996  
851,284 / 46%

Generations

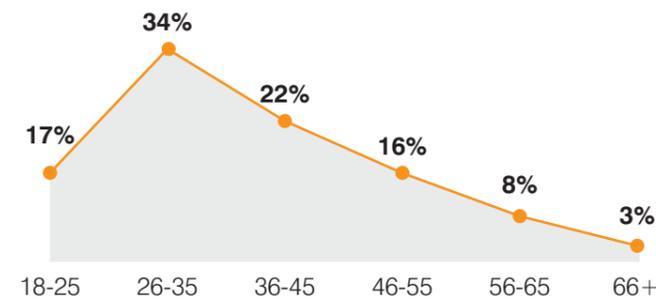
**Employed Full-Time**  
1,239,657 / 66%

Employment

**81%**  
1,508,469  
71% UK Coverage

Smartphone / Mobile

Age Bands | % of Segment vs. UK Average



ID Poor  
Experion ID Poor Estimate



991,339 Adults

Online Banking Apps  
Digital Literacy



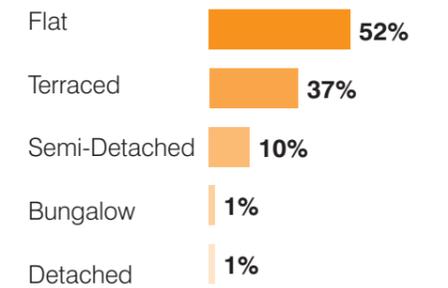
1,929,627 Adults

#### Property

Likely in flats or terraced properties in urban areas. High prevalence of renting and also higher than average housing association residents.

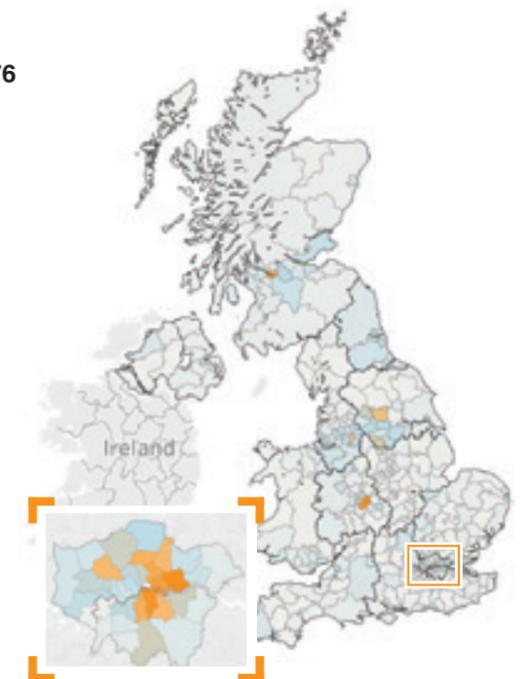
#### Education

Moderate proportions reaching GCSE and also vocational qualifications. In line with UK totals for A Levels whilst under-represented at degree level.



#### Urban Renters | Geographic Profile

Regional Profile | vs. UK Segment Average



0 130,903

### Segment 5 Community Culture

An estimated 291,074 adults are ID Challenged in this segment. Mostly age 26-45, in work with lower than average qualifications. They are urban dwellers, many in multinational settled communities – with over half living in London. About 27% use online banking apps.

#### Community Culture | Portrait



Community Culture are households living in residential city suburbs. These urban communities are comprised of people at different stages of life, including families and retired people.

Employment ranges from roles earning good salaries in city centres to routine jobs that provide basic wages. There is a high use of public transport, especially for travelling to work. Many are settled in their homes and have lived there for many years. These neighbourhoods are often multinational in character.

#### Adults

**2,332,320**

% of UK Adult Population

**4.6%**

**36-45**  
418,615 / 22%

Age

**Millennials**  
1981 - 1996  
560,831 / 30%

Generations

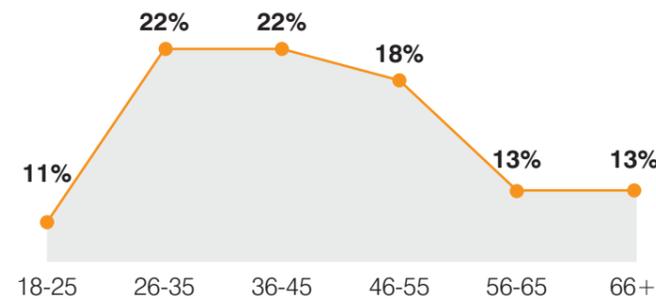
**Employed Full-Time**  
1,022,981 / 55%

Employment

**75%**  
1,403,176  
71% UK Coverage

Smartphone / Mobile

Age Bands | % of Segment vs. UK Average



#### ID Poor

Experion ID Poor Estimate



**291,074** Adults

#### Online Banking Apps

Digital Literacy



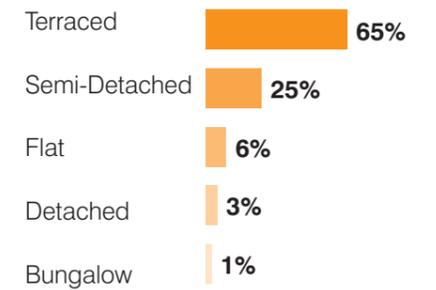
**618,820** Adults

#### Property

Renting and owning larger 2-3 bed terraced properties in dense urban areas. Often family homes with children in household.

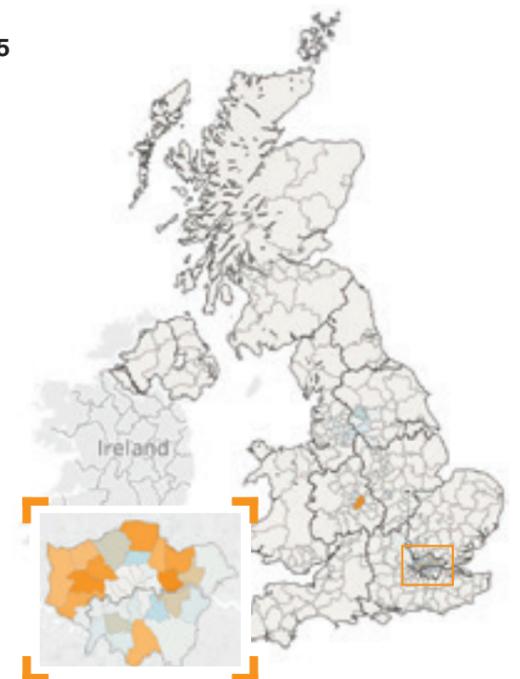
#### Education

Larger proportion with vocational qualifications, lower than average studying a degree and slightly higher than average with no qualifications.



#### Community Culture | Geographic Profile

##### Regional Profile | vs. UK Segment Average



0 176,697

## Rest of UK – Segment to show the number of ID Challenged outside Segments 1-5

There are ID Challenged people across the whole UK adult population. In addition to the five key segments, which include the highest proportion of the ID Challenged in the UK, Experian estimated 7% of the rest of the UK population are ID Challenged. That's roughly 2,283,145 people. This segment were not analysed in detail as the methodology does not enable that, but it is a reasonable assumption that this segment can be helped with some of the alternative ID Proofing techniques and approaches.

The clusters of 'types' which make up this group are varied (which makes analysis challenging). A higher proportion of the remaining ID Challenged fall into clusters of 'types' which tend to be more affluent, with higher-than-average incomes and are found across all age ranges and in both urban and rural areas.

### 16-24 years – Young Home Sharers

For the purpose of estimating the number of ID Challenged within the Learning Records Service Data Set further manual analysis was undertaken.

Young Home Sharers are aged 16-24. ONS data was reviewed, which is based on individuals, unlike MOSAIC 7, which is based on households. This enabled 18-24 year-olds to be drawn out, who are invisible to a large extent in the household data – as they are not the primary bill payers, nor the primary name on rental agreements and so typically will lack the ID-Proofing documents. Some analysis was also performed on 13-17 year olds.

Roughly 10-15% of people between 16-24 years old are ID Challenged.

### Age 16-24 ID-Challenged estimates:

The ID-Challenged estimates for those people within the 16-24 age group with neither a passport, nor a driving licence are below. These are rough<sup>15</sup> and based on ONS Air, Sea, Tunnel statistics for a proxy for passport holders.

**4.5 Million** – no passport and no full driving licence

**2 Million** – no passport and no driving license of any type (i.e. no provisional license, nor full license)

As a large proportion of 16-24 year olds use a mobile phone to go online, it is likely a high proportion could probably use NFC (near field communication) and passport to verify their identity. Near-field communication allows two devices placed within a few centimetres of each other to exchange data and is used by Verify and NHS Log-on to take a selfie as part of the online verification customer journey.

The Photographic ID Research Report found that: 'Younger people were more likely than the general population to hold a form of photo ID. Ninety-nine per cent of those aged 18-29 held a form of photo ID, slightly higher than either those aged 30-69 (98%) or 70+ (98%).'<sup>16</sup>

Of those aged between 19-24, most are urban, highly transient, likely to live in shared accommodation and will often not be named on council tax/ rental letters. These individuals are mostly found in Segment 4: Urban Renters.

Age 13-17 – Most live at home, do cash work if any; an estimated 85-90% travel by air/sea/tunnel and 95% use a mobile phone to go online.

<sup>15</sup> Young Home Sharers: method of calculation (16-24 year-olds): The latest ONS statistics for the passport proxy (travel by air/sea/tunnel) available by age were for 2015 so they were weighted up based on the incremental increase between 2014 and 2015. The DVLA statistics by age were used (which are accurate) and provided directly from the DVLA team. Then the ONS statistics for UK overall population by age were used to calculate the negative – those who do not have a driving licence or a passport. These ONS UK population statistics are estimates and as the year by year age groups for 2021 were not estimated by ONS; the calculation weights up the 2019 data based on the expected population increase in the UK. Please bear in mind that this does not make for a high level of confidence in the estimations due to the proxy and the estimations. The ID Challenged estimates for Young Home Sharers can only be used as a rough indication.

<sup>16</sup> The Photographic ID Research Report, Cabinet Office, 31 March 2021.

## APPENDIX 3 – Data Sets

### Extended Customer Attributes (Online Banking) Data Set

A data set of extended customer attributes is currently being developed by the UK Open Banking team (OBIE) in collaboration with seven UK banks. The data set includes attributes such as name, address, DOB and might draw on customer banking activity for knowledge-based verification questions. Customers will have been verified as part of the 'know your customer' (KYC) and anti-money laundering (AML) processes which are standard in the banking sector. There are roughly 6.4 Million of the ID Challenged people across the five segments in this data set. Hence, it scores high.

The aim of this work is to design an open Extended Customer Attribute (ECA) Standard to support the sharing additional customer data, which is not required under PSD2 (a set of laws and regulations for payments for the EU) nor the CMA (Competition and Markets Authority) Order, and which can be surfaced by banks and provided to relying parties on a commercial basis. This data can be used for a number of identity verification use cases and is of particular relevance to the UK government's work (in particular the proposed trust framework for digital identity) as well as the OpenID Foundation's eKYC & Identity Assurance work<sup>17</sup>.

It has been agreed that a common, open ECA Standard such as that already developed by OBIE<sup>18</sup> would be beneficial to support a number of use cases, providing clarity around the data elements, presentation of consent to the customer, and provision of this data to relying parties. The open Extended Customer Attribute (ECA) Standard is modelled on the principles used within the PSD2 customer journeys for TPPs (trusted third party providers), for example, gathering of consent and authentication.

It is suggested that the Steering Group stay close to developments on the open Extended Customer Attribute (ECA) Standard to support the sharing additional customer data, especially as the proportion of the population using online banking is speculated to rise.

### National Fraud Initiative (NFI) Data Sets

The National Fraud Initiative comprises 22 data sets, which include concessionary bus passes and blue badge holders. Those which have potential for ID verification are listed in the table. As attribute data was not available at the time of writing, GPG45 scores could not be assigned.

Based on the high counts in some of the NFI data sets (such as the electoral register and council tax). It is suggested that NFI data sets are explored in more detail so that scores can be assessed based on attributes for each data set.

### NFI Usage and Legislation Reform

The Cabinet Office uses the powers granted exclusively by the Local Audit and Accountability Act 2014 (the 2014 Act) to collect and match data as part of the National Fraud Initiative. The 2014 Act requires that the Code of Data Matching Practice is agreed by Government and followed by all organisations that participate in the Cabinet Office's data matching exercises (NFI). The National Fraud Initiative (NFI), conducted by the Cabinet Office, involves data matching and the mandate currently only covers the use of the data to help in the prevention and detection of fraud: Misrepresentation or non-disclosure.

<sup>17</sup> OpenID Foundation

<sup>18</sup> Open Banking Standard

**NFI has undertaken a request for expansion of these powers to allow the NFI data to be used for wider purposes, as detailed below:**

- a. Recovery of debt owing to public bodies
- b. Prevention and detection of crime (other than fraud);
- c. Prevention and detection of errors and inaccuracies; and
- d. Apprehension and prosecution of offenders.

Dependent on the parliamentary schedule and legislative approval, a decision on the extension of the NFI powers is expected by the end of May 2021.

### DWP Data Sets

DWP has a number of data sets, including Universal Credit, Employment Support Allowance and Jobseekers Allowance, Child Maintenance, Retirement, Bereavement & Care. More granular investigation is required into individual data sets, numbers of people in those data sets and attributes that align to the Experian variables. This work can be done with interaction with data set owners if DWP datasets are deemed a priority. Pension Credit and Winter Fuel Payment would likely align most with segments 1 & 2 for example. The full list is on GOV.UK.<sup>19</sup>

The State Pension dataset would certainly reach a chunk of the ID Challenged segment in Segment 1 and segment 2. The UK average % of the population for people receiving the state pension is 23% and segment 1 has 30% on State Pension and segment 1 contains 22%. An estimated 86,0000 people are highly likely to be on state pension in segment 1. Roughly 184,000 people are likely to be in the State Pension data set in segment 2.

Further investigation into individual DWP data sets is considered worthwhile. DWP have not indicated whether they would or should share data at the point of writing this report.

### DWP CIS Data Set

The Customer Information System (CIS) is a system driven by a data set used by the Department for Work and Pensions to store basic identifying information such as name, address, date of birth, National Insurance number and fraud markers if relevant. As well as identifying information, CIS keeps a limited record of benefits claimed over the last two to three years if applicable.

The information CIS holds is used to help DWP staff quickly find out about UK citizens when there is a need to contact the department or their service front offices. It is suggested that further work is done to ascertain the level of difficulty for IdPs to query the CIS data set for ID validation and ID verification in combination with another proofing technique.

As Claimed IDs are held in this data set, along with the documents produced and details of a vouch, it is worth investigating further for ID-Proofing purposes more widely.

Local Authorities can query the CIS data set to Validate and Verify an ID if the Claimed ID is present in an LA office.

<sup>19</sup> Gov.UK

### Learning Records Service (LRS), Department for Education

In order to estimate younger people in this data set, further manual analysis was undertaken using DVLA and ONS data. Within the five ID Challenged segments there are young people who are not 'visible' as they are not identified on the household documents.

Please see Appendix 2 – Young Home Sharers (16-24) for further detail.

The Learning Records Service went live in 2010. It is a record of all qualifications from UK awarding bodies recognised by the Department for Education. Qualifications awarded to UK citizens from the age of 14 are added by UK schools. It is not fully comprehensive. At the point the report was written it had not been established how many unique citizens are accounted for in the data set. Deeper understanding of the numbers of people in the data set would be useful. Based on the data set including all 14 year-olds from 2010 onwards who are registered at a school in the UK it might be useful to reach the younger ID Challenged who are 25 years old and below in 2021.

An API exists (currently thought to be used exclusively for back office transfers), although APIs are not open and politically it may take some time to agree extended use for ID validation purposes as conditions of use are tightly bound to Department for Education policy.

Citizen awareness and service provider awareness of the service is low. It includes attributes such as name, address, DOB, qualification type (GCSE, Degree, etc.). However it is less comprehensive for university qualifications as it is not mandatory for universities. It does not include a Photo ID and is not cleansed, so the data becomes out-of-date quickly and does not manage change of name or change of address. However, it could be useful for static knowledge-based verification questions, such as 'where were you living in 2010?'

#### Primarily applicable for segments 3, 4 & 14-25 year-olds<sup>20</sup>.

#### Segments 1 – 5 for people 18-25 Years, approximate population counts are:

**Segment 1:** 9,255 people (46% of segment have a qualification and 7% are 18-25 years)

**Segment 2:** 5,269 people (46% of segment have a qualification and 7% are 18-25 years)

**Segment 3:** 84,687 people (57% of segment have a qualification and 12% are 18-25 years)

**Segment 4:** 131,451 people (78% of segment have a qualification and 17% are 18-25 years)

**Segment 5:** 24,013 people (75% of segment have a qualification and 11% are 18-25 years)

A rough estimation of the number of people in the data set (based on all schools adding all registered students) is 7 Million people who are between the ages of 14 – 25 years in 2021. Of those, and based on 10-15% of the 16-24 year olds being ID Challenged, it is estimated that between 700,000 – 1 Million people in the data set are ID Challenged.

There are a small proportion who do not attend school and will not be added to the data set, so this is not inclusive of all 14-25 year olds in the UK. As it is not mandatory for universities to update the LRS, the counts are quite low, coupled with the political challenge of opening the API for use for an ID purpose, at this stage this data set requires further investigation. If the policy of the Department for Education changes, this is a promising data set to explore further.

<sup>20</sup> Please see Appendix 2 for more information about Young Home Sharers (16-24).

### NHS Patient Data Set

The NHS patient data set is used for the purpose of linking a person to a medical record. The data set includes Name, Address, DOB, NHS Number. The NHS Number is not entirely unique, so it can't be classed as a unique identifier, but as there are only very few duplicates (it is thought no more duplicates than rare duplication of the NINO), it may be of use for ID purposes. If this is deemed to be an adequate unique identifier, then this data set would have a strength score of 2. If not, a score of 1. There is no Photo ID. The process aligns to GPG45, but there is no fraud or credit checking. An API exists and is currently used to query the dataset to support patients to register for NHS Log App in as well as in person in a GP Practice.

### NHS – Staff Data Sets

The NHS has two categories of staff in their 1.4 Million strong staff data set: i) higher level access, and ii) lower level access.

Primary care is any service directly accessed, which includes general practitioners (GPs) and health visitors. Secondary care includes hospital services and mental health services.

#### i) Higher Access NHS Staff:

NHS Staff with higher level access to national NHS systems have Photo ID cards and access to the Integrated ID Management which comprises standard access & national systems access. Their identity will have been verified in line with GPG45 guidelines. NHS staff verification process is score 3 in hospitals and for higher access staff (verified photo ID and proof of address over time).

Hence the 800,000 NHS staff on the Integrated ID Management system could be both validated and verified using this data set. However, its use is questionable as this group would likely not fit the ID Challenged profile of any of the segments. They are doctors, nurses and specialists within a socio-economic demographic group that did not rank high in the Experian analysis and so discounted. In other words there will be very few ID Challenged people in this data set.

#### ii) Lower Access NHS Staff:

NHS staff with lower level access to NHS systems include porters, cleaners and other staff who would likely fit the ID Challenged profile. They would likely be found across all 5 segments, but particularly in segments 3, 4 and 5 with higher proportions of people in work.

This group comprises roughly 600,000 people. They do not have a photo ID card. A standard NHS employers ID Check will have been performed for this group (photo & address or 5 forms and no-photo ID).

Every employer should be seeking evidence. NHS HR was unable to confirm what ID checks the GP practice does (primary care).

The process for both NHS staff groups supports the ID Challenged, such as a 16 year-old apprentice without a driving licence and who's never been abroad, by performing a vouch with a photo & school records.

It is unknown exactly how many people have been vouched, although it is likely 60% of the 1.4M directly employed (GPG45 type ID proofing) would not have been vouched. So roughly 560,000 people might have been vouched for. A proportion of those vouched may include non-UK Nationals who find it more difficult to validate and verify their identity, especially if they are new to the UK.

### Student Loan Account

Student loan accounts are held with the Student Loan Company. The SLC is funded entirely by the UK government and the devolved administrations. It is a non-profit making government-owned organisation that administers loans and grants to students in colleges and universities in the UK. SLC is an executive non-departmental public body, sponsored by the Department for Education. It could be categorised as an authoritative source as it is the responsibility of a government department. In which case, if the SLC were to issue digital verifiable credentials, or if its data set were queried to validate an identity or verify an identity that has been verified according to the MLR 2017, it would produce a score of 3 and be useful for verification of a claimed identity.

However, its ID Challenged reach is limited. Segments 4 & 5 have the highest proportion of people with a student loan account (this relied on a proxy – university degree or higher), but all segments are well below the UK average. The UK average is 33% and segment 4 has 21% and segment 5 contains 24%.

A population estimation would be rough as not all people with a degree will have a student loan account. A very rough calculation might be 23,0000 people in segment 5 (taking the estimated % of the segment with a degree and then cutting that with the % in the age band up to 35 years when many people would have paid a student loan off as well a taking off a proportion for those who do not need a student loan). Please treat this calculation with caution. A similar calculation for segment 4 might amount to roughly 106,000 people between the ages of 18-35 years. Although there is a higher % of people in segment 5 likely to have a student loan account, the count is higher in segment 4 as there are three times as many people in segment 4 compared to segment 5.

The SLA ranks higher as it can also verify a claimed ID, but lower as its reach is lower, which brings its priority as a dataset to focus on down the overall ranking.



## General Register Office (GRO)

GRO (part of HMPO) team members were consulted. Birth Certificates carry name, date of birth, address at birth and a system number (this is unique). Not everyone is registered – a small proportion of the population, including travellers who have home births with no medical intervention. It is appropriate for validation an identity exists, but not verification. Anyone can obtain anyone else's birth certificate, provided that they can provide sufficient information to identify the entry and pay the required fee.

GRO holds over 285 million records dating back to 1837. Around a half of the records are paper-based, although various attempts at digitisation have been made over the years. The result is that civil registration records are currently held in a variety of digital and non-digital formats, with differing levels of electronic capability. A full electronic data set is available from 1 July 2009 for all births and deaths. From this date it relies solely on input data. Hence validation is possible via API. However as the full set only includes young people up to the age of 12, GRO data is not of use for the ID Challenged until 2022 onwards. Each year approximately 700,000 births are recorded. This is a priority data set for the future if the data continues to be input and APIs are made available for the Private Sector.

**BIRTHS:** 700,000 per/year 12 years = **8.4M**

**DEATHS:** 500,000 x 12 = **6M** (not linked to births or marriage data and no name change recorded.)

Marriages are recorded from 2010. GRO considered it impossible to link across birth, death and marriage data sets as there is no unique identifier to carry across from one record to another – each registration entry for a different registration event such as birth, marriage, civil partnership and death has a different system identifier.

GRO is currently working with DCMS/ GDS / DWP / DVLA to consider digitisation of records prior to 2008. This digitisation process will also include a large proportion of scanned in paper documents. For DWP and DVLA purposes this is effective as the purpose was to combat the need for customers to produce a paper birth certificate.

Funding from GDS will enable GRO to capture birth data electronically back to 2000. This group would include some of the younger ID Challenged. It would be useful for DVLA Provisional Driving Licence checks for younger people. GRO is looking to increase digital records for people up to 21 years of age (this would meet DVLA's data requirements) and would like to go back to 1995 for births and deaths, but will require funding to increase that capability. Additionally, GRO would like to increase the number of field-captured electronic records (as opposed to scanned documents), which can be queried.

A Life Events Verification API exists. It is available via an online service and is currently used for HMRC administration checks for child benefits applications. A match is made to Child benefits number (DWP then puts this on the CIS data set). The same API is also used for Universal Credit for a child. This API could be used for ID-validation, although it needs some more work.

The same API could be re-purposed for the use of ID-Validation checks, but the policy which wraps around it would need to be tested for the public and private sector (if, and only if, legislation was in place to include private sector usage underpinned by robust policy appropriate to mitigate risk).

### GRO Usage and Legislation Reform

In terms of legislation, GRO has powers disclose information to government departments/ Local Authorities (public sector only). The GRO is listed in legislation as having a discretionary power as part of 2017 Digital Economy Act. Death data (not birth data) is shared with a limited number of commercial organisations under powers contained in the Police and Justice Act 2006 – e.g. Experian and Equifax. However, GRO lacks the power to disclose birth data with the private sector more widely – a change in legalisation would be required for private sector use for ID validation purposes. A bill is being proposed by DCMS to give GRO powers to share with the Private Sector. This is unlikely to be in 2021.

## Local Authority Data Sets

The Welsh Government and several Welsh Local Authorities were consulted about data sets and ID proofing processes and techniques. Every Local Authority has their own business processes and particular data sets, however some general approaches are common to most.

There are also Library Data Sets, which might also serve for validation (depending on what documents the individual has, they need to produce proof of address and something from an official source, like a bank card, so could score 1 or 2). From initial consultation with local authorities in Wales, which has a large proportion of ID Challenged citizens, there are numerous other data sets, which may help with validation.

Local Authorities administer numerous data sets. Those identified for potentially increasing the potential for supporting the ID Challenged are council tax and housing benefit. The data overlaps with a number of the National Fraud Initiative data sets (including housing benefit and council tax as well as others). Without further analysis of the attributes of the respective data sets, it not possible to conclude categorically whether the Local Authority or NFI data sets are more use for ID purposes. Purely in terms of ease, it is suggested that the NFI is investigated in more detail as it is a centralised repository of Local Authority data and so would make API querying of data sets more straightforward.

The process a local authority uses to verify an ID requires two forms of ID. All staff are trained to recognise a passport & true likeness. In the past banks have accepted Housing Benefits letter from the Claimed ID. If the Claimed ID has no passport, driving licence, or credit history they are sent by some local authorities, to the local DWP Job Centre where a vouch can be performed if the Claimed ID is known over time. A declaration is signed to that effect.

In Wales if someone does not have internet access or capability, Local Authority offices will complete the application (for example, Council Tax Reduction) on behalf of the person online and check the CIS Data Set for National Insurance Number and additional checks – or refer the Claimed ID to a JCP for verification checks. Some LA services are now co-located within Job Centre Plus.

Based on conversations with the Welsh government and Welsh Local Authorities about the verification process, DWP's CIS Data Set is a more promising source for the ID Challenged verification.

## Private Sector Data Sets

There are a number of data sets which we created proxies for in the ID Challenged segment analysis. Oyster cards and National Rail photo cards and passes were included. Contacts in the relevant organisations were not consulted and were assessed on process to validate identity information available in the public arena at the time of writing. Level of difficulty can be assessed with further consultation with the organisations (Transport for London (TfL) and National Rail).

Oyster Cards, based on a Mosaic proxy of underground travel – estimation of 246,297 people. Railcards (with Photo ID), based on a Mosaic proxy of train travel - estimation of 231,555 people.

As train travel is relatively expensive and these are not extremely high numbers – and counts are higher in other data sets which are more likely to embrace the ID Challenged (such as the NFIs concessionary bus pass data set which has nearly 600,000 people in it) – these are not considered priority data sets.

## APPENDIX 4 – Alternative ID-Proofing Techniques & Methods

### METHOD 1 & 2: API validation against authoritative source - yes/ no response & Richer data sharing APIs

These API<sup>21</sup> ID-Proofing techniques validate an identity exists. Richer APIs can verify an identity if they allow for knowledge-based “verification” questions that are based on a credit history. However, based on current GPG guidance (referenced earlier in this report), knowledge-based “validation” questions would not verify an identity if they were not based on historical data that has previously verified the Claimed ID (such as bank account data). They do not verify an identity.

Both YES/NO APIs and Richer APIs require an API to exist to enable the query of data sets which hold data about a data subject.

Method 1 offers a Yes/No response to a query.

Method 2 offers richer data in response to a query, such as confidence level, an evidence score, other attributes that can be used for KBVQs, the source of the knowledge-based questions (to enable assessment of whether they can merely validate or verify an identity), or the type of document checked by the authoritative source.

The API Validation Methods are dependent on APIs being available to interface with Data Sets. APIs are expensive to build from scratch, but if they already exist they can be a very useful way to query existing data with relative ease – both for the systems that use them and the customers that use the front-end services APIs can help to drive.

Another consideration for these methods is the standards and policy which are wrapped around an APIs. Policy and data risk were key criteria for prioritising both the data sets and ID-proofing techniques.

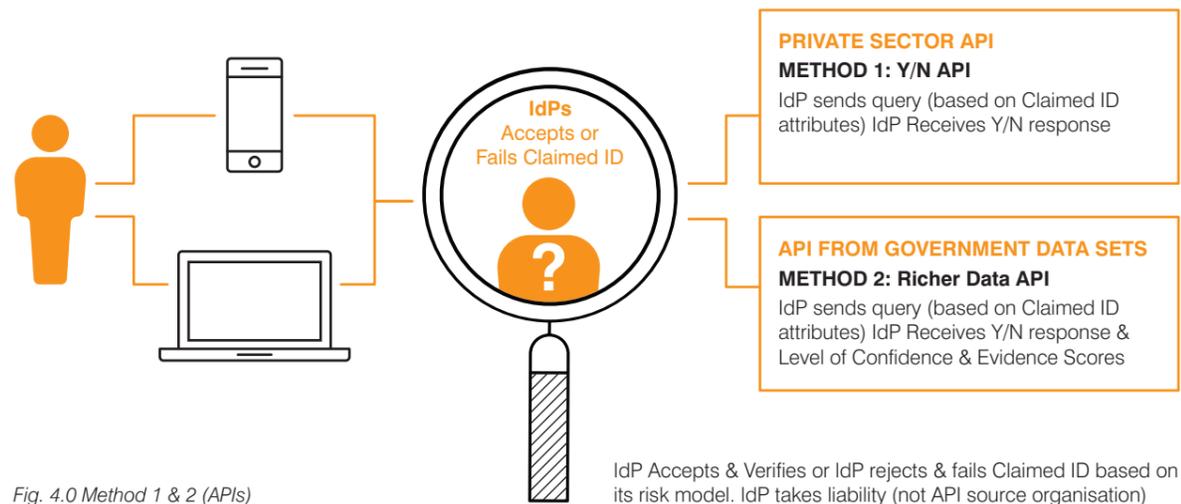


Fig. 4.0 Method 1 & 2 (APIs)

Fig 3.0 shows a high level representation customer journey, and is one example of how ID-Proofing Methods 1 and 2 might work, taking the Learning Records Service (a DfE data set) as an example. The risks associated with individual data sets are covered in section 4. Some considerations for use of this service should be noted.

<sup>21</sup> An Application Programming Interface (API) allows two applications to talk to each other.

There is a low level of customer awareness that the service exists and the data set has been live since 2010. As the data set is not kept up-to-date over time, its reliability in terms of offering a useful set of questions the use could answer (and remember the answer to) is to some extent limited and there is a risk that over time the information is increasingly compromised, making it easier for a claimed identity to be impersonated. In the case of a data set like this, the most appropriate use would be for knowledge-based “validation” questions as the data is not kept up to date over time. Static knowledge-based questions could validate that the identity exists for younger people, although as you will see later in this report, there are not currently a very large proportion of the ID challenged in this data set.

This is just an illustration of what a journey might look like. Other data sets (such as DWP benefits data sets or National Fraud Initiative data sets might also be queried using Methods 1 & 2).

See Section 3 for Data Sets considered.

### METHOD 3: Account Log-on in controlled environment

Account Log-on in controlled environment might include a customer accessing non-banking services, such as government services or carrying out a DBS check, by using their Open Banking log on. The bank might be an attribute provider, act as a “component IdP” or it may act as an IdP itself in terms of liability in some instances.

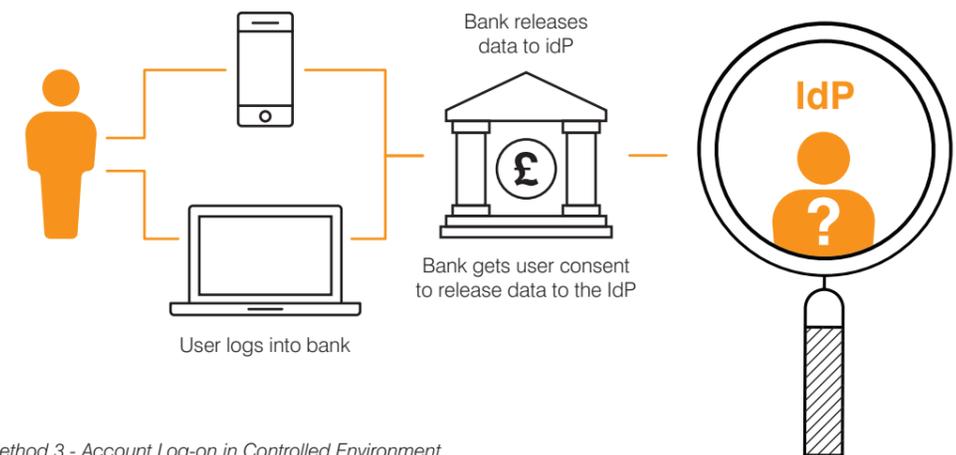


Fig. 4.1. Method 3 - Account Log-on in Controlled Environment

In addition to Open Banking, this technique could be applied to other open data initiatives, such as Open Finance, Open Savings and Investments, Pensions, Open government data.

#### METHOD 4: Digital Verifiable Credential issued by Authoritative Source with Digitally Signed Certificate

A Digital Verifiable Credential issued by an authoritative source, such as a government department, to a trusted wallet can both validate and verify the identity of an ID Subject.

It takes a user through the full customer journey to both validate and verify their identity. It ranked joint first as a proofing technique, although there are some risks (see risk table).

A Digital Certificate is usually created by an organisation (in this case, an authoritative source). It is a digital equivalent of a physical authenticator, such as a driving licence, with similar levels of assurance from the issuing authority. The certificate carries the digital credential and a digital signature.

A Digital Credential links two or more attributes together and can carry several facts. It is 'verifiable' because it is digitally signed. The digital signature verifies that all the attributes carried within the digital credential are linked. The certificate says who signed the certificate and that gives the link to the authoritative source.

A Digital Certificate is usually created by an organisation, and this will be signed by the authoritative source issuing the credential and certificate.

A Digital Signature is 'mathematical proof the data you are seeing is the same data as when the data was applied (i.e. it has not been changed) and links the data to the organisation that created it (which owns the key to unlock the data).

For this method to work, both a Digital Certificate & a Digital Signature are needed. The Claimed ID and authoritative source also need to have an established relationship.

The recipient needs to be able to read the Verifiable Credential. It is not possible to make the Verifiable Credential readable or not readable. Privacy is down to the individual and who they make the Digital Credential available to. The trusted wallet or IdP is the gatekeeper. A digital wallet is a software-based system that securely stores user payment information and passwords for numerous payment methods and websites.

For example, Jane Smith has had a COVID vaccine. The NHS issues a verifiable credential to a trusted wallet, a personal data store, or via an IdP.

The user would have a smart phone with multifactor authentication to use this method. This method assumes that by virtue of the Claimed ID having a trusted wallet, the user would likely be an early adopter of technology and have a reasonably high level of digital literacy and capability. Hence, there would be a smaller proportion of users who could use this method found in the ID-Challenged segments, particularly the older users.

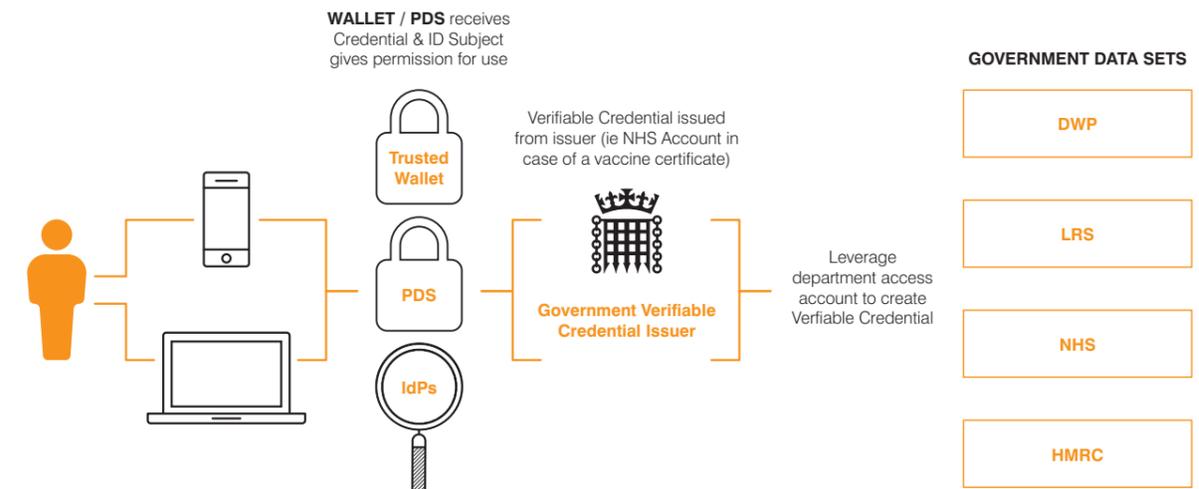


Fig. 4.2 – Method 4 – Verifiable Credential Issued by an authoritative source via Trusted Wallet (this example uses the case of an NHS Certificate being issued by the NHS)

**METHOD 5: Letters with added security features (i.e. QR Code)**

Letters with added security features such as QR codes can validate a Claimed ID, but not verify their identity.

There are a number of dependencies for this method to work. Use of a QR code in a letter would require an organisation to decode the QR code, which would need to contain strong identity proofing evidence, such as name and date of birth. Provided that such evidence were in place, the data carried in, or pointed to by, a QR code could be used by an IdP to validate the Claimed ID's identity. The Claimed ID would need to receive a letter in order to prove their address and that letter would need to come from an authoritative source.

There are various types of QR code with different amounts of data and levels of security. Quick Response (QR) codes are a popular and increasingly used way of storing data. Part of their popularity and wide adoption is that they are inexpensive to create. Because simple QR codes have some security risks there is a move towards using Secure Quick Response (SQR) codes. These types of codes are equipped with a reading restriction function, making them more appropriate for storing private information. This solution is more likely to guarantee the integrity of the source data and originating party validity with an appended cryptographic hash (see glossary). The coupling of a digital certificate minimises the possibility of spoofing, tampering and man in the middle attacks. All of these breaches are known issues with standard QR codes.

With standard QR codes, attackers are able to embed malicious URLs that contain malware which is able to extract data from a device when scanned or embed code that using redirection to phishing sites. Counterfeit QR codes are not uncommon where fraudsters place their own codes over legitimate ones. SQRs minimise these risks and because these codes can only be read by specific types of scanners, whilst not guaranteeing complete security does add an extra layer of protection.

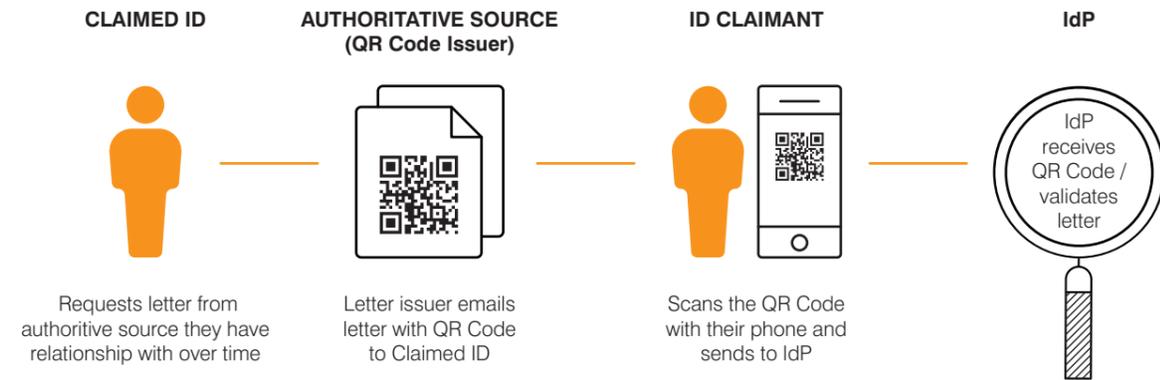


Fig. 4.3 – Method 5: Example Customer Journey using letters with QR Codes

**METHOD 6: Manual validation against authoritative source**

Manual validation against authoritative source can be used to validate an identity exists, but cannot verify the identity of a Claimed ID.

Verification is possible when the method is combined with a vouch or if the individual is already verified on the DWP Customer Information System (CIS). An authoritative source might include a DWP Job Centre Plus or a Local Authority front office such as a connect centre in Wales. In some Local Authorities where Universal Credit was rolled out early, such as Flintshire in Wales, there are very few ID-Challenged people passing through their offices. Most of these people are seen at JCP offices.

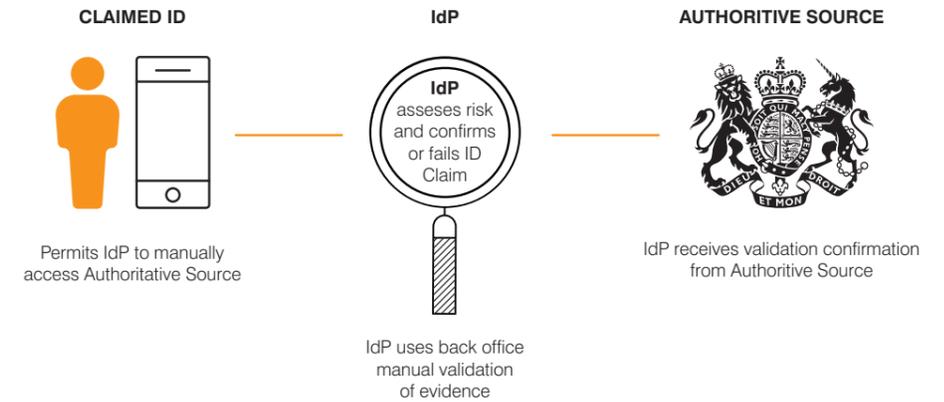


Fig. 4.4 – Method 6: Manual Validation against an authoritative source

## APPENDIX 5 – Additional insights about the ID Challenged

Data Sets Variables & Proxies for alternative ID-Proof count estimations from Experian's MOSAIC 7 data sets.

Sub-Category Mosaic Category Document Type Mosaic Variable Seg 1% Seg 2% Seg 3% Seg 4% Seg 5% UK Avg

### Accommodation & Household

Sub-Category	Mosaic Category	Document Type	Mosaic Variable	Seg 1%	Seg 2%	Seg 3%	Seg 4%	Seg 5%	UK Avg
Accommodation	Tenure	A mortgage account (including buy to let mortgage accounts)	Owned	58%	40%	40%	15%	68%	64%
		A rental or purchase agreement for a residential property	Rented	12%	6%	7%	57%	21%	18%
	Benefits	LA Letter - Social Housing	Housing benefit	14%	18%	17%	15%	9%	11%
Household	Marital status	A marriage or civil partnership certificate	Married	41%	42%	30%	12%	29%	39%

### Education

Sub-Category	Mosaic Category	Document Type	Mosaic Variable	Seg 1%	Seg 2%	Seg 3%	Seg 4%	Seg 5%	UK Avg
Education	Highest qualification	A student online loan account	University degree or higher	10%	2%	4%	21%	24%	33%
		An education certificate from a regulated and recognised educator	Has qualifications	46%	21%	57%	78%	75%	77%
		Learning Records Service - source DfE (no qualifications & inverse)	GCSEs	23%	7%	32%	20%	13%	14%
			Vocational qualifications	11%	16%	15%	43%	46%	31%
		UCAS	University degree or higher	10%	2%	4%	21%	24%	33%

### Financial & Bills

Sub-Category	Mosaic Category	Document Type	Mosaic Variable	Seg 1%	Seg 2%	Seg 3%	Seg 4%	Seg 5%	UK Avg
Financial	Apps	A bank account through online banking	Personal banking	32%	32%	35%	35%	27%	32%
		Current / savings	No account	6%	7%	7%	7%	9%	5%
		Loans	Have loan	17%	14%	18%	18%	15%	16%
		Retirement	Personal	19%	21%	19%	18%	19%	24%
Utilities / Bills	Electricity bill	A gas or electric account	Direct debit or pay...	81%	80%	72%	77%	84%	85%
		Gas bill	Direct debit or pay...	83%	81%	74%	78%	85%	86%
		Directors in...	No directors	97%	99%	97%	97%	91%	94%

### Digital Literacy and Access

Sub-Category	Mosaic Category	Document Type	Mosaic Variable	Seg 1%	Seg 2%	Seg 3%	Seg 4%	Seg 5%	UK Avg
Digital	Business and Industry Websites	Digital Literacy - Capability / Access / Willingness	Business services	53%	58%	39%	40%	45%	46%
			E-commerce	9%	7%	9%	8%	12%	9%
	Communication	Digital Literacy - Capability / Access / Willingness	Use Social Networks	56%	47%	60%	57%	45%	51%
	Home internet	Digital Literacy - Capability / Access / Willingness	Broadband	93%	91%	92%	89%	89%	93%
	Internet devices	Digital Literacy - Capability / Access / Willingness	Smartphone / mobile	68%	56%	81%	81%	75%	71%

## Glossary of Terms:

### Attribute

Attributes are the describing characteristics or properties that define all items pertaining to a certain category applied to all cells of a column. For example, name, address, DOB.

### API

Transfer of data from one system to another. API stands for “application programming interface”.

### Cryptographic Hash

An algorithm that takes an arbitrary amount of data input—a credential—and produces a fixed-size output of enciphered text called a hash value, or hash. That enciphered text can then be stored instead of the password itself, and later used to verify the user.

### GPG45

UK Government guidelines for verifying an identity

### KBQ

Knowledge-based questions or knowledge-based verification are questions only the Claimed ID can know and can be used to validate and verify an identity in combination with other forms of ID-proofing. According to GPG45 knowledge-based challenges should be specific enough to be able to prove that that person is who they say they are.

### Malware

Software that is specifically designed to disrupt, damage, or gain unauthorized access to a computer system.

### NFC

“Near-field communication” allows two devices placed within a few centimetres of each other to exchange data.

### Phishing

A fraudulent practice of sending emails purporting to be from reputable companies in order to induce individuals to reveal personal information, such as passwords and credit card numbers.

### Spoofing

Type of cyber-attack in which someone attempts to use a computer, device, or network to trick other computers or networks by masquerading as a legitimate entity.

### Trusted Wallet

A digital wallet is a software-based system that securely stores user payment information and passwords for numerous payment methods and websites.

### Validate

Validating that a user exists based on documentary or electronic evidence.

### Verify

Verifying the person trying to create a digital ID is the person they are claiming to be.

### Vouch

A vouch is defined in GPG45 as ‘a declaration from someone who knows the claimed identity.’ A person vouches for another person by claiming that they know them to match the claimed identity. A vouch is another type of evidence about a claimed identity. A vouch can be in-person or digital.

### With thanks to those organisations consulted:

Linda Edwards, Deputy Head of Reform – Bills and Legislation, Her Majesty's Passport Office, General Register Office  
 John Duffy, Deputy Head of Reform – Strategy and OGD Engagement, Her Majesty's Passport Office, General Register Office  
 Richard Trevorah – Technical Director at tScheme  
 Chris Lewis and Mark Saunders – Synectics Solutions (on behalf of NFI)  
 Adam Lewis, Director – Hippo Digital  
 Sam Sullivan, Head of Statistics and Research – Pennaeth Ystadegau ac Ymchwil / Data Cymru  
 Paul Heskins, Service Design Lead – Identity, DVLA  
 David Rennie, Digital Identity & Data Assurance Specialist – IDEMIA  
 Melissa Ruscoe, Programme Head, NHS login – NHS Digital



**OIX** OPEN IDENTITY  
EXCHANGE

Open Identity Exchange UK Europe  
Suite 1, 3rd Floor  
11-12 St James's Square  
London SW1Y 4LB

**Email:** [info@openidentityexchange.org](mailto:info@openidentityexchange.org)

**Web:** [openidentityexchange.org](http://openidentityexchange.org)