

Pensions Dashboards:

Getting to the heart of matching



A thought leadership research report for pension trustees, providers and administrators from ITM, PASA Expert Partner for Dashboards



Altus
An  equisoft company

itm
Solution delivered...

Pension Fusion is the Integrated Service Provider (ISP) solution launched by ITM and Altus in November 2021, leveraging our combined experience of over 35 years as independent technology leaders within UK pensions. We became the first ISP to connect to the PDP central digital architecture in April 2022.

Our early start has enabled us to build a suite of sophisticated match criteria and solutions within the product, from which pension trustees, providers and administrators can choose.

These solutions have been developed from analysis of real schemes' data to understand the specific challenges of matching for pensions dashboards. We now want to share our findings with the industry.



Contents

	Foreword	Page 04
	Welcome	Page 05
	Summary	Page 06
	The matching challenge	Page 06
	Our research and results	Page 07
	Matching decisions for trustees	Page 10
	What should schemes be doing right now on matching, and what's next?	Page 11
1.	Introduction and background to the research	Page 12
1.1	Why is digital matching important for the success of dashboards?	Page 12
1.2	What are trustees' new duties, what help is there, and why is research needed?	Page 13
1.3	Understanding the research	Page 15
1.4	What was the focus of the research?	Page 15
1.5	What's in this research report?	Page 15
2.	Research process and findings	Page 16
2.A.	Defining the research approach	Page 16
2.B.	Refining the research dataset	Page 18
2.C.	Running the research match tests	Page 20
2.D.	Unpicking the research findings	Page 27
3.	Recommendations and next steps	Page 28
	Annexes	Page 30
I.	Key concepts and terminology used consistently in this research report	Page 30
II.	Further details of the research findings	Page 32
	About the research team	Page 34

Foreword

from Chris Curry, Principal, Pensions Dashboards Programme

Pensions dashboards have the potential to transform how people access and view information on their pensions. For the first time, everyone will be able to use a simple and clear online service that shows their State, occupational and personal pensions, all in one place. Once available, people of any age will be able to get a holistic view of their estimated retirement income.

Creating dashboards for the UK is a highly collaborative endeavour. MoneyHelper, at the Money and Pensions Service (MaPS), and other commercial providers, are building the front-end dashboards which users will actually see. At the Pensions Dashboards Programme (PDP, part of MaPS), we've created the central digital services which will securely orchestrate the search for users' pensions. Meanwhile, data providers are connecting their client pension schemes' data so it is searchable, and, amongst other things, are implementing scheme-specific processes to match the users of dashboards with their pensions.

In order for dashboards to match people to their pensions, there needs to be clear criteria on the data that will be used to do this – particularly to ensure that personal details held by pension schemes and providers are accurate. In August 2022, the Pensions Administration Standards Association (PASA) issued data matching convention guidance. The Pensions Regulator also issued some initial guidance in June on matching people with their pensions.



This large scale matching research by ITM / Altus is invaluable for bringing to life any matching issues and will help schemes and their technology partners devise sophisticated matching solutions. We're grateful for ITM / Altus undertaking this work for the benefit of the industry and showing the collaborative spirit that's needed from all of us to make a success of dashboards.

Welcome

from ITM / Altus

We very much hope you find this data matching research helpful for your preparations for pensions dashboards.

From our long experience working with pension records, we know that data matching could be the single biggest point of failure, or hopefully success, for the entire dashboards initiative.

Working with their technology and specialist data partners, pension trustees and providers now can, and must, understand the accuracy of key personal data items, right across their membership or customer base, and implement appropriately sophisticated match criteria.

We look forward to continuing this critical matching debate as we move forward with dashboards through the Autumn and towards schemes' compliance from April 2023 onwards.



Mark Lecompte
Chief Executive Officer, ITM

 [marklecompte5](#)

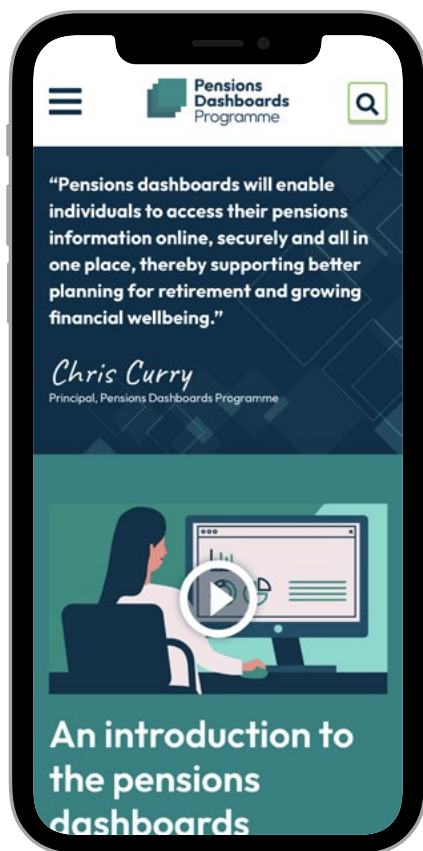


Kevin Okell
Managing Director, Altus

 [kevinokell](#)

Summary

1) The matching challenge



It's a great aspiration, Chris, and one we all hope will work.

But, of course, it all depends on pension trustees and providers (or suppliers on their behalf) actually being able to digitally find people's pensions with confidence.

To do that trustees must tell their data provider what match criteria¹ they want them to use.

Moving past the matching 'starting point'

Industry guidance has anticipated the starting point for many schemes will be to think about matching on data items such as Last Name, DOB and NINO, but it's clear as a one-size-fits-all solution this is not comprehensive. Which trustees can confidently say they hold, at all times, accurate, up-to-date, and correctly spelt Last Names, DOBs and NINOs, on every single one of their deferred and active records?

- NINOs held may be temporary or otherwise inaccurate, for example with two digits transposed.
- Last Name changes over time are often not notified to scheme administrators, particularly by deferred members.
- The day and month parts of DOBs can easily be incorrectly entered when they are first input, and only the member themselves would recognise the mistake!

Pension administrators are not always able to identify such inaccuracies in these data elements which they are given, or are not given, by employers, deferred members and others.

Whilst these realities are well known to pensions administrators they are sometimes not visible to trustees. The "Common data" Record-keeping Guidance from The Pensions Regulator has historically often been implemented by trustees and administrators as a test of the presence of data, rather than its accuracy.

So, trustees are going to need to decide on more sophisticated criteria for making matches, and possible matches, between people using dashboards and the records they hold. And they're going to need sophisticated insights on their scheme data to help them decide what those match (and possible match) criteria should be for their particular scheme.

Thought leadership research

This is why ITM is publishing this thought leadership research report: to help trustees and others in the industry better understand their matching challenges, and ideas for solutions. We've moved beyond a theoretical discussion of matching by investigating real data matches. We have looked at the real personal data held by some large schemes, and applied various Altus / ITM Pension Fusion match criteria² to discover just how successful, or unsuccessful, the consistent digital finding of pensions might be in reality.

¹ **A note on terminology** - match criterion, match criteria and match policies: We refer to "match criteria" as a set of data elements to compare against, e.g. LN / DOB / NINO, specified on either an exact match or a more sophisticated matching basis. A "match criterion" would be if you tried to match on just one data element, e.g. NINO, but that's not recommended so we don't tend to use that term. This report shows it's likely schemes will want their suppliers to use multiple match criteria, which we refer to as a "match policy", specifying the different match criteria to be used for both matches made and possible matches. If any one match criteria within a match policy results in a success then you have a match made result, or a possible match result.

² Over a year ago, in July 2021, the Pensions Dashboards Programme (PDP) announced ITM as an alpha data provider. Then, in November 2021, ITM and Altus announced they were joining forces to create the first Integrated Service Provider (ISP) solution, Pension Fusion. In April 2022, Pension Fusion was the first ISP to connect to the PDP central digital architecture. Over the last year, as part of this iterative development work, Altus / ITM have now built a suite of sophisticated match criteria and solutions within the Pension Fusion ISP product, from which client pension trustees, providers and administrators can choose.

Other ISPs may have alternative approaches to making matches and possible matches which are equally valid: in this report we're seeking to illuminate the general matching challenge rather than focusing on particular matching solutions for particular schemes.

2) Our research and results

How did we do the research?

Our core approach was to test matching on real data, at scale. So we chose the real active and deferred memberships of

- a) some large defined contribution (DC) schemes, including master trusts as they're first to connect, as well as
- b) some smaller defined benefit schemes.

In total, these schemes have **more than a quarter of a million** members.

We simulated all of these members logging in to a dashboard to see how successful various match criteria would be. Any member could log in, so trustees need to be comfortable that their chosen match criteria can find pensions right across their membership.

Not everybody will use dashboards, but anybody could. So as soon as you're connected, you need users to be able to find any pension in your scheme

By comparing the simulated Find Requests against the schemes' original data we then tested a range of different match criteria, to see how they would perform, both individually, and **in combination**. We had a focus on criteria for both "matches made" and "possible matches". And finally, but very importantly, we wanted to understand more about minimising the risk of matching against the wrong pension for "possible matches", and all but eradicating that risk for "matches made".

What did we find?

LN / DOB / NINO is a good start, but not enough

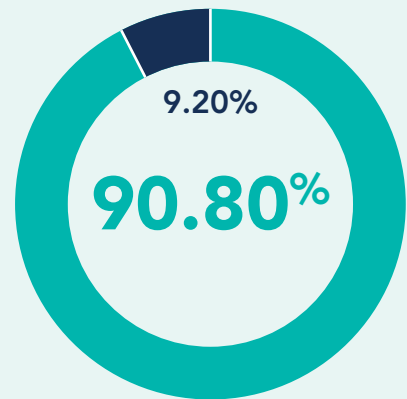
As anticipated in both TPR's initial Guidance on matching and PASA's Data Matching Convention (DMC)

Guidance, on the large schemes we researched (with a total of more than a quarter of a million member records), searching for an exact match on Last Name, DOB and NINO was a good starting point.

Figure 1:

90.80%

records in our research dataset matched our simulated Find Requests using Last Name, DOB and NINO



■ Unmatched on LN/DOB/NINO ■ Matched on LN/DOB/NINO

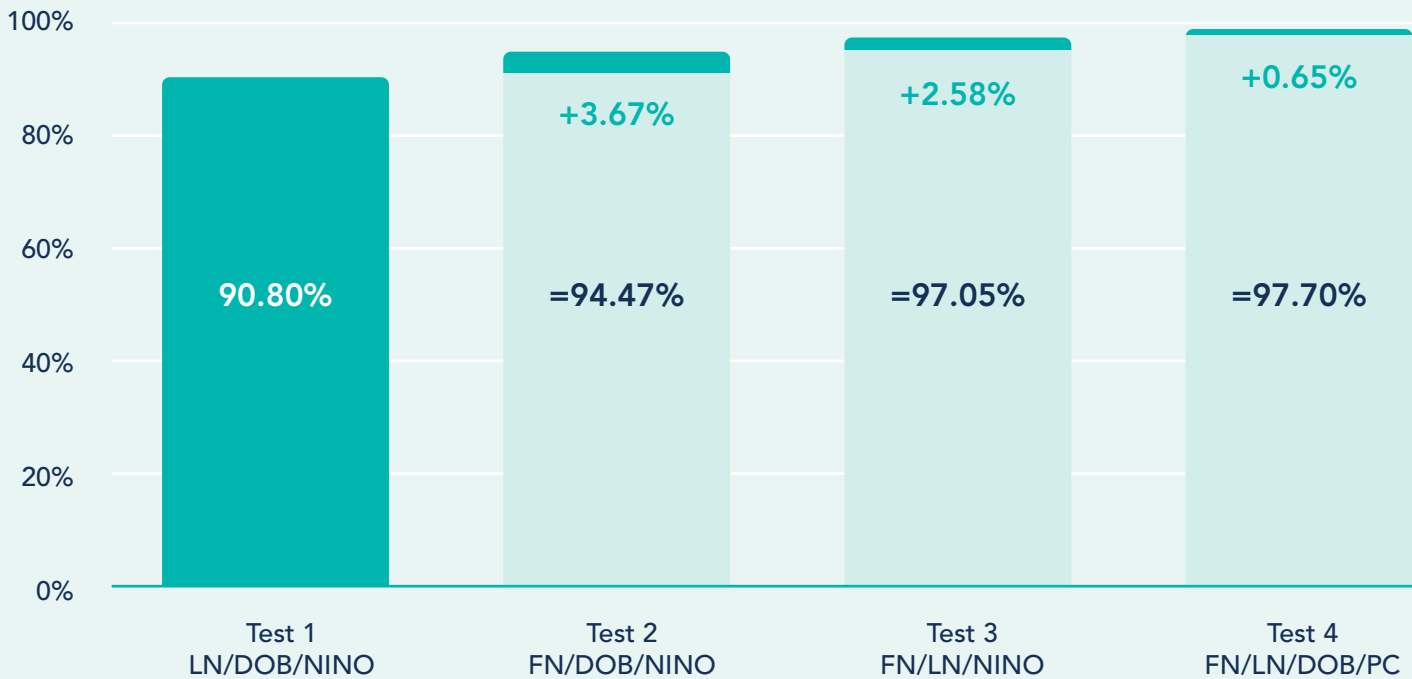
Important notes:

- This find rate reflects, that on 90.80% of our sample schemes' member records, the Last Name and DOB held by the schemes are exactly the same as those which would be included in Find Requests received from the central digital architecture, as verified by the central Identity Service.
- It also assumes these same 90.80% of users would have voluntarily, and correctly, input their NINO (which remember isn't centrally verified), and that these self-asserted NINOs are the same as the NINOs held by our sample schemes.
- From our 20 years of pension data experience we know it's likely these assumptions will be true for only a percentage of the member population, meaning the base level find rate for almost all schemes will be lower than 90.80%. So this can perhaps be seen as a "best case" for the simple, single, match criteria of LN / DOB / NINO.

Summary

Figure 2:

Incremental proportions of simulated Find Requests matching sa



The remaining pensions

Finding the remaining pensions really matters. Using our above best case scenario, if schemes attempt to make matches solely on LN / DOB / NINO, and thus fail to find roughly 1 out of every 10 pensions, across the whole UK pensions universe, that could equate to nearly 10 million pensions not being found: a poor outcome for savers.

Increasing the find rate towards 100%

After the base level of 90.80% matching, we then ran sophisticated additional test match criteria against the 9.20% unmatched population, aiming to improve this matching %. We then repeated this process multiple times using different test match criteria, each time against the leftover unmatched population. This is illustrated in Figure 2.

Figure 3 shows that, through this combination of different match criteria, we found (against our research dataset) it was possible to increase the find rate from 90.80% to 99.26% or more.

Later in this report we discuss how you could define which of these match criteria are best suited to "possible match" responses, as opposed to "match made" responses, and also discuss whether the final criteria - FN / LN / DOB - might over-reach as a possible match response for some members.

Sample schemes' records using different test match criteria

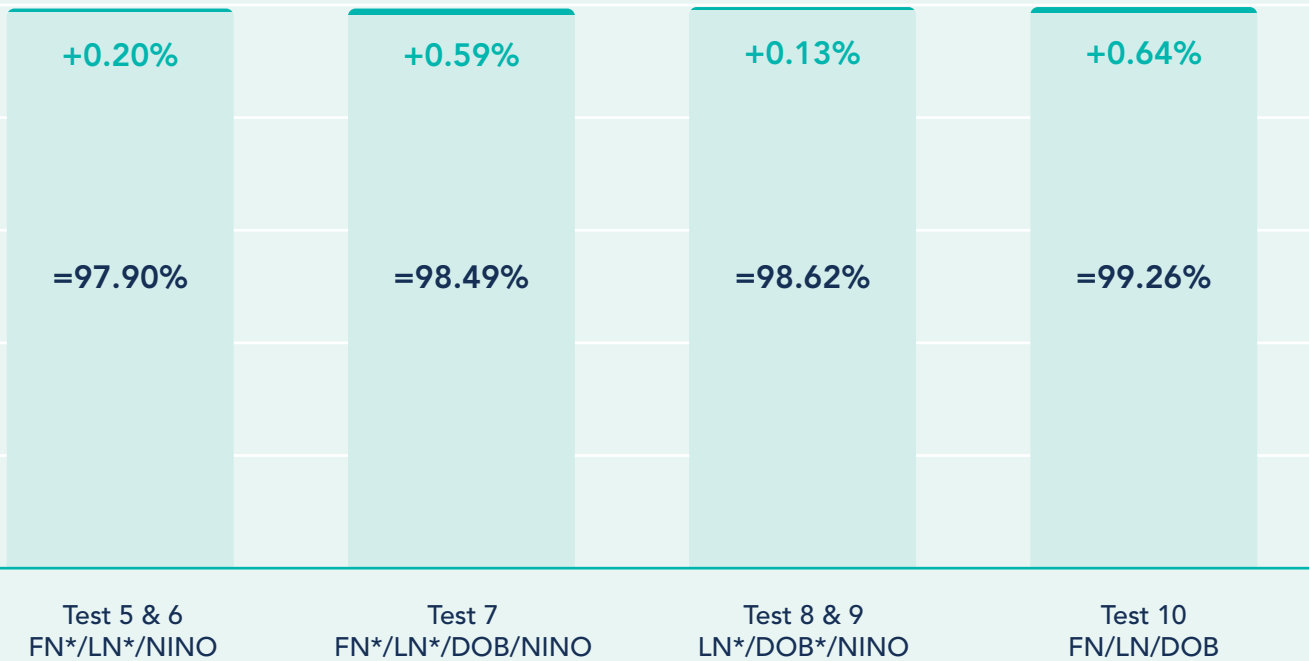
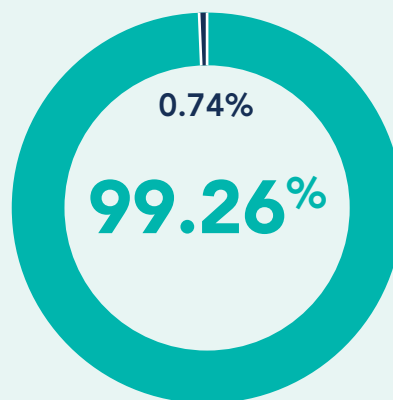


Figure 3:

Through a combination of different match criteria we found it was possible to increase the find rate, against our research dataset, to

99.26%



■ Unmatched after Test 10 ■ Matched by Tests 1 to 10

Every pension matters

The final 0.74% of unmatched records are still a problem. Across the whole UK pensions universe this could equate to **c.740,000 pensions not being found**. This is a lot better than 10 million pensions not being found, but these 740,000 pensions could be very important to the people they belong to.

In the report, we discuss why this final very small minority of records aren't being matched, what might be done about this, and what potential further research work might be done.

* In Tests 5 to 9, we searched for name / DOB matches which were nearly the same, e.g. with a Levenshtein Distance (LD) of 1, or 2 for FNs/LNs longer than 6 characters, or just 1 digit different on DOB. In Test 7, we also "swapped" FN & LN, i.e. we compared the FN held by the scheme with the LN in the Find Request, and vice versa. Read the report for a fuller description of these tests.

Summary

3) Matching decisions for trustees

What does this mean for trustees' decisions on matching?

The first thing to reiterate is that these research findings are not seeking to, and indeed they cannot, tell trustees which match criteria they should use. We're not saying the particular match criteria we researched are the ones schemes should necessarily use, nor in the sequence in which we researched them (and see Annex II for some thoughts on possible alternative sequencing of match criteria).

What we're seeking to do is illuminate the matching challenge, using real data, at scale.

Working with their data provider(s), trustees will need to do their own scheme-specific thinking to decide on the different match criteria to use on their particular scheme - see What should schemes be doing right now? below. We think it's likely they will want to consider using a combination of match criteria, which we call a "match policy".

Match policies: Combinations of match criteria for matches made and possible matches

Now, knowing this single match criteria on LN / DOB / NINO will only find 9 out of 10 of their schemes' pensions, an optimal solution for trustees may be to decide on a more sophisticated match policy incorporating a range of different match criteria. And across that range of match criteria, the trustees will need to decide which criteria they wish to signify a "match made", and which they wish to signify a "possible match".

For example, for a record matching our Test 2 (FN / DOB / NINO), but failing Test 1, we can deduce Last Name doesn't match. So the trustees could decide to use this as an opportunity to say this is a possible match: this means the member would be asked (via the dashboard they're using) to contact the scheme through normal channels, and evidence their up-to-date Last Name so the scheme's records can be updated accordingly. Subsequent Find Requests would then match on LN / DOB / NINO (i.e. match criteria 1) which is a good outcome for all.

By signifying a possible match instead of a match made trustees will also prevent any risk of presenting the wrong pension record to a member. This is sometimes referred to as a "false positive", which would likely lead to a personal data breach.

Or, the trustees could decide they're comfortable saying these FN / DOB / NINO matches are actually matches made, increasing their find rate (in our example by 3.67% to 94.47%).

So trustees need to weigh up maximising their find rate against improving their data accuracy (assuming these members do actually contact the scheme to get their Last Names updated).

A further consideration is the volume of member queries possible matches could potentially create. Looking beyond this report it's important that adequate solutions are developed to handle such traffic, because presenting no match instead of a possible match will have a negative impact on member experience: a factor pensions dashboards aim to improve.

An indicative match policy is illustrated in Figure 4 for the 10 match criteria we tested (noting these aren't recommended match criteria, nor a recommended sequence of match criteria).

Figure 4:

Trustees' chosen match criteria sequence	Trustees' chosen match criteria specification	Trustees' chosen match criteria Response
1	Exact match on all of LN / DOB / NINO	
2	Exact match on all of FN / DOB / NINO	
3	Exact match on all of FN / LN / NINO	
4	Exact match on all of FN / LN / DOB / PC	
5	FN (LD1/2), LN (Exact) / NINO (Exact)	
6	FN (Exact) / LN (LD1/2) / NINO (Exact)	
7	Transposed FN & LN / DOB (Exact) / NINO (Exact)	
8	LN (LD1/2) / DOB (Exact) / NINO (Exact)	
9	LN (Exact), DOB (1 digit diff) / NINO (Exact)	
10	FN (Exact) / LN (Exact) / DOB (Exact)	

4) What should schemes be doing right now on matching, and what's next?

In our research, we were aiming to shed light on what can be a dry and abstract subject. So we've tried to make the report as easy as possible to read as it's essential all trustees engage with this critical topic, so dashboards work well for consumers. But this research isn't the end of the debate on matching - far from it. We took a research approach which aimed to illuminate the challenges and suggest ideas for solutions. Based on our research findings, we've suggested below various future actions: for pension trustees and providers, for the Pensions Dashboards Programme, and for the industry as a whole.

Actions for:	Actions	When
Pension trustees / providers	<p>Scheme-specific conversations and investigations:</p> <p>Have detailed conversations with your chosen ISP(s) about what match, and possible match, criteria might be most appropriate for your particular scheme.</p> <p>This may involve thoroughly investigating the personal data you hold for your deferred and active members, and also tracing / verifying this data with external agencies. If you still have poor personal data records once there are high numbers of dashboards users, this could result in large volumes of burdensome work for your administrators as users come through the possible match resolution process.</p> <p>You may decide you wish your ISP(s) to adopt a suite of different match criteria in combination (we call this a "match policy"). For each match criteria you must decide whether you wish to return a match made or a possible match response, balancing maximising matches made with improving data accuracy and avoiding incorrect matches (by using possible matches).</p>	From Autumn 2022 onwards
Pensions Dashboards Programme (PDP)	<p>Scale consumer testing:</p> <p>Numerous questions must be user tested to further guide trustees' decision making on their matching policy. These can then be tested at scale once schemes are connected from the initial deadline of August 2023. Questions include:</p> <ul style="list-style-type: none"> • What proportion of dashboard users input their NINO? • What proportion input it correctly? • What proportion input other self-asserted data elements such as Previous Last Name? • How will consumers react to the decisions on matching taken by their different pension schemes & providers? • How will users react to possible matches? • How can the UX design of the end-to-end "possible match / data correction journey" be optimised to deliver ongoing improvements in schemes' data accuracy? 	From beta testing phase, but at scale from Autumn 2023 (once Cohort 1(a) of schemes and providers have connected to the pensions dashboards ecosystem)
Pensions industry as a whole	<p>Further learnings, research and refinements:</p> <p>As more and more schemes carry out the scheme-specific investigations above, further learnings will become available. Also, insights will flow from the consumer testing on matching described above. These additional learnings could highlight the need for further industry research, building on the concepts in this initial report. See Section 3 Recommendations and next steps for a potential list of areas where further industry research may be required.</p>	2023

1. Introduction and Background

1.1 Why is digital matching important for the success of dashboards?

The success of pensions dashboards in the UK from the mid-2020s relies on many things.

Arguably the most important will be the match criteria pension trustees and providers decide to use.

Pensions dashboards can only display a user's pensions if trustees (or their suppliers acting on their behalf) are able to match the individual using a dashboard to the pension records held by the scheme.

This is why the Government's Pensions Dashboards Programme (PDP) said two years ago:

“ Being able to digitally match individuals confidently to all their pension entitlements is at the heart of the whole pensions dashboards endeavour. ”

And the PDP-commissioned independent research report from PwC, published in October 2020, said:

“ The ability of different data providers to match individuals to their pension entitlements is critical to the success of the PDP [and] the true extent of the challenges faced by [real] matching data may only become apparent when data providers test data with the pensions dashboards ecosystem. ”

By doing this research, we wanted to kick-start this real data testing by simulating, at scale, Find Requests from dashboards and seeing how well they would match against real data.

1.2 What are trustees' new duties, what help is there, and why is research needed?

New duties

Regulation 22(1) of the indicative draft Pensions Dashboards Regulations 2022 (final regulations expected in Autumn 2022) requires trustees to:

- "decide on criteria to use for matching"
- "keep a record for at least six years", of their match criteria decisions, and
- "have regard to any guidance on matching" issued by DWP or TPR.

DWP's January 2022 consultation on the draft Regulations (Chapter 2, Paragraph 9) set the expectation that schemes must:

- "take reasonable, diligent steps" to ensure successful matching.

The Information Commissioner's Office (ICO) response to that DWP consultation stated:

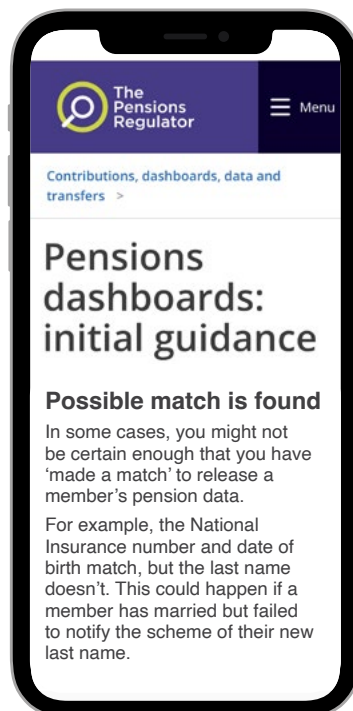
- "data providers may need to undertake (or update) a Data Protection Impact Assessment (DPIA) before match processing is carried out ... giving consideration, when setting their match criteria, to the minimal yet adequate data sufficient to accurately identify matches ... these "minimal" and "adequate" aspects should be carefully balanced to ensure processing is fair and proportionate".

Different types of match

The draft regulations envisage pension schemes arriving at two different types of match:

- Match made, where a pension held by a scheme definitely belongs to the individual using a dashboard, and
- Possible match, where a pension held by a scheme might belong to a dashboard user but the trustees either do not have sufficient confidence in the matching of data elements to be able to say it is a match made, or it is clear that some of the user's data conflicts with the trustees' records and they want to take the opportunity to correct this.

TPR's initial Guidance, published in June 2022, explains more on possible matches:



Possible matches will play a key role in achieving successful matching and enabling data accuracy to improve over time as more and more people use dashboards.

What is successful matching?

Successful matching is matching which finds pensions, balancing and minimising two risks:

- **finding wrong pensions**, i.e. where a scheme incorrectly matches a dashboard user to somebody else's pension which they hold - this would be a data breach in the case of a "match made", and
- **not finding pensions** which should be found, i.e. where a scheme holds a pension for a dashboard user but they fail to make a match meaning the pension isn't found.

The concept of possible matches in DWP's Regulations is a tool designed to help limit the number of pensions not found and also, as a bi-product, help improve data accuracy over time.

1. Introduction and Background

What help is there for trustees?

- In December 2021, updated in August 2022, the Pensions Administration Standards Association (PASA) published helpful introductory **Data Matching Convention (DMC) Guidance**
- In June 2022, The Pensions Regulator (TPR) issued **initial guidance on matching**, with a subsequent consultation on TPR's dashboards compliance and enforcement policy, including on schemes' matching, expected in Autumn 2022
- In July 2022, the Information Commissioner's Office (ICO) said, in its **response to DWP's consultation**, that matching for dashboards means schemes should consider completing (or updating) and publishing a Data Protection Impact Assessment (DPIA), with a focus on data minimisation when setting match criteria, i.e. what is the minimum but adequate amount of personal data that data providers need to fulfil matching?

But it's important to understand that all guidance is necessarily generic and cannot have blanket applicability for all schemes. None of PDP, TPR, ICO, PASA, or any other body, will ever be able to tell a particular scheme what match criteria to use, because it needs to be scheme-specific. Ultimately, only the scheme can devise their most efficient matching policy because they have the best insight into their data accuracy and verification processes.

In particular, criteria for possible matches need to be scheme-specific, depending on a scheme's type and profile and how its trustees wish to use possible matching to help find pensions and improve data accuracy over time. For example, large master trusts, taking on more and more employers over time, could really benefit from making good use of possible matching to improve their ongoing personal data accuracy.

What help do trustees need?

To comply with the new legislative duties above therefore means trustees will need:

- **scheme-specific expert advice and support** on what match criteria to use, and
- **pensions dashboards data provider(s)** whose digital matching capabilities and solutions can meet their scheme-specific matching requirements.

Why is there a need for research?

Matching is a seemingly straightforward, but actually very involved, topic.

So to help trustees, and the wider industry, understand and prepare for dashboards matching we have undertaken some primary research into the real performance of data matching on some real sample large pension schemes. We were aiming to:

- illuminate the detailed analytical rigour needed to support the expert advice on matching which all trustees will need before they can decide on their match criteria
- illustrate the real impacts of trustees' choices of different match criteria on the success of their matching, and therefore what matching solutions they may need from ISPs
- increase understanding across the whole industry, by sharing the research findings widely, given the critical nature of matching to the overall success of dashboards.

1.3 Understanding the research

We've written this report assuming readers are familiar with the overall design of, and terminology used within, the Government's pensions dashboards ecosystem.

If you're new to dashboards, read Annex I first. There we've summarised the key concepts and terminology you need to understand to get the most value from this report. Annex I is based on the latest published information from DWP, TPR and PDP and we use these terms consistently throughout this report.

We've used the term "trustees" throughout the report to refer to the party who has the legal duty to comply with dashboards legislation and is the data controller, although this could instead be taken to be a public service pension board or an FCA-regulated pension provider.

1.4 What was the focus of the research?

Given the new duties on trustees set out above, before deciding what match criteria to instruct their suppliers to use, trustees will want to:

- **assess the expected performance of different match policies** (i.e. different sets of match criteria): this will enable schemes to compare different match criteria to understand if some will make more matches, and possible matches, than others, thus minimising their risk of not finding pensions which should be found
- **assess how various advanced matching techniques and / or data improvement activities** could improve matching performance
- **gain confidence that their initial chosen match criteria will not have an unacceptable risk of finding wrong pensions**, particularly in relation to possible match criteria, so they are in a good position to learn from experience and refine criteria further where necessary over time.

Our research proposes a systematic approach to these key assessments in a way which all trustees will need to consider before they can set their own match policies.

1.5 What's in this research report?

Section 2 summarises the research process we adopted and sets out the key research findings, with some further details and discussion in Annex II. Based on our research findings, Section 3 lists ITM's key recommendations for next steps to be taken by a) pension trustees and providers, b) the Pensions Dashboards Programme in the next testing phases, and c) the pensions industry as a whole.

2. Research process and findings

ITM has been analysing pension scheme data for 20 years - this isn't the first personal data matching research we've done.

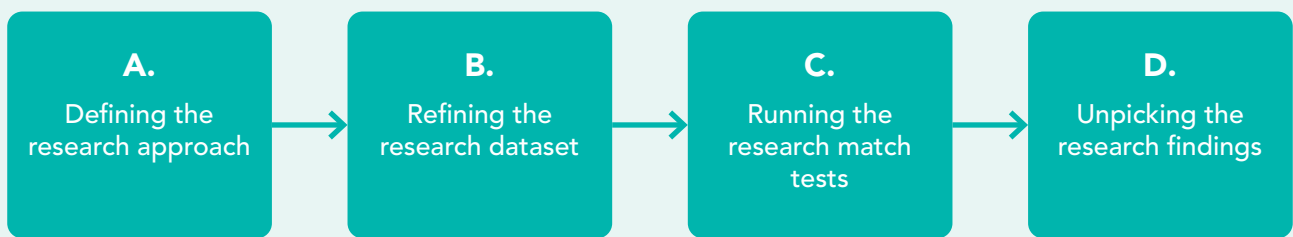
For example, there are many similarities between dashboards data matching and the matching needed for the automatic consolidation of small pension pots.

So this 2022 research report can in some ways be seen as building on the previous ITM / Altus 2021 thought leadership research and analysis for the PLSA on data matching for small pots.

In that 2021 report, we made various educated assumptions about the accuracy of personal

data held by pension schemes. We said then that we would expand our work to investigate how the true accuracy of certain data items, such as Last Name and First Name, impacts data matching. This is what we've done in 2022, by verifying and tracing real data, at scale.

Below is a summary of the steps we took:



2.A. Defining the research approach

2.A.1 Approach to the research dataset: Our key ambition was to test matching at scale. So we chose the real active and deferred memberships of some large defined contribution (DC) schemes and some smaller defined benefit (DB) schemes, covering more than a quarter of a million members.

In other words, we were aiming to understand how successful various match criteria would be if these sample schemes' whole memberships logged in to use a dashboard.

You might question whether a scheme's entire membership actually needs to be "findable" from the scheme's connection date as, in reality, it's extremely unlikely a scheme's entire membership will login to dashboards and trigger Find Requests. However, any member could log in, so trustees need to be comfortable their chosen match criteria can find pensions right across their membership.

Which schemes?

We focused on large DC schemes used for automatic enrolment (AE), including some master trusts, because these are the schemes with the earliest deadlines to connect to the dashboards ecosystem, about a year from now in August / September 2023.

But we also wanted to research how well matching might work on some real DB schemes. The research results shown are the aggregate results across all the schemes we researched.

Not everybody will use dashboards, but anybody could. So as soon as you're connected, you need to be able to find any pension in your scheme.

Our approach to the research data was to take a copy of all the personal details held by the sample schemes and then use this data copy to simulate Find Requests (which data providers will receive from the central digital architecture of the dashboards ecosystem).

Of course, in the real world, Find Requests received for these users will have up-to-date details, not the (potentially out-of-date or inaccurate) personal details held by the schemes in our copy data. So we verified and traced the data with other sources to "enhance" our copied data to better simulate the real world Find Requests which will be received in reality.

Because data changes all the time, the enhanced data in our simulated Find Requests (sFR's) still might not have been truly up-to-date in all cases. So in reality there could actually be more differences between the sFR data and the original scheme data than our research findings suggest. Schemes need to work hard continually to keep their personal data as up-to-date as possible at all times.

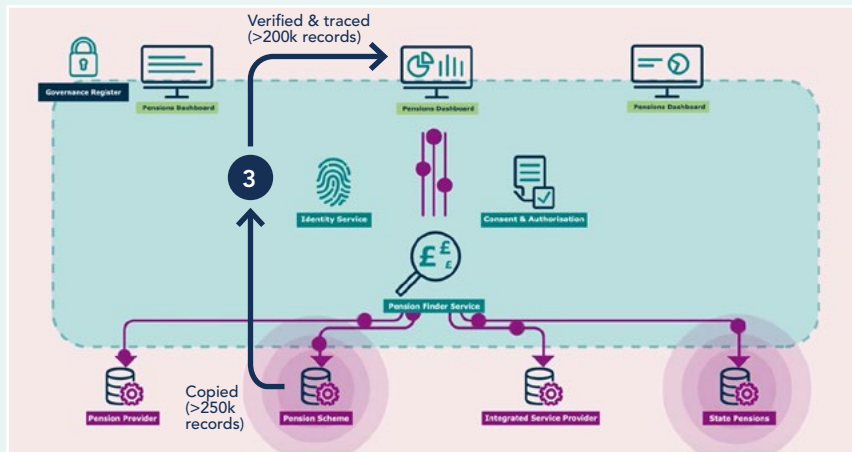
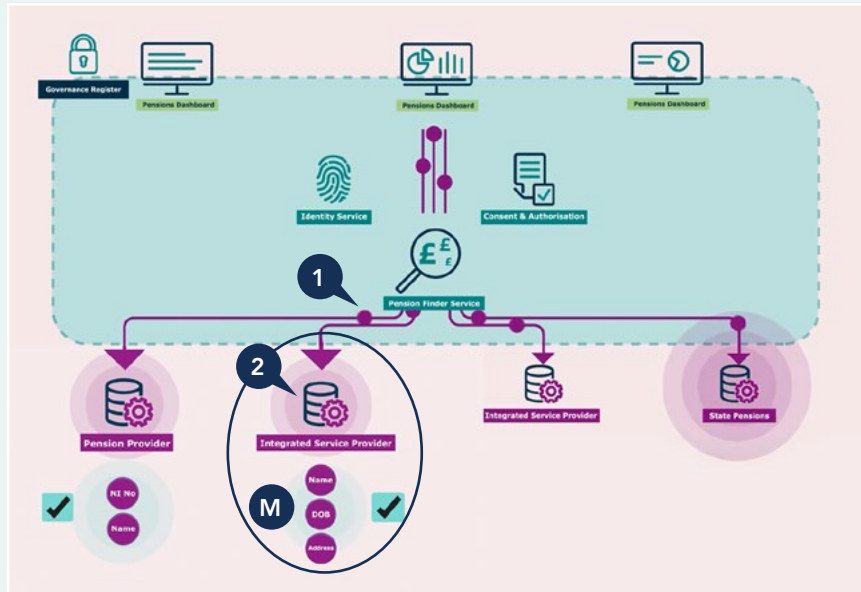
How we simulated matching at scale for our research

How matching works:

Personal data flows from a dashboard, through the central digital architecture, to data providers, in a Find Request (1 in the diagram) - each purple dot in the diagram is a Find Request.

Each connected data provider then compares all the Scheme data (2) they hold against the personal data in each incoming Find Request - this is a process called matching (M).

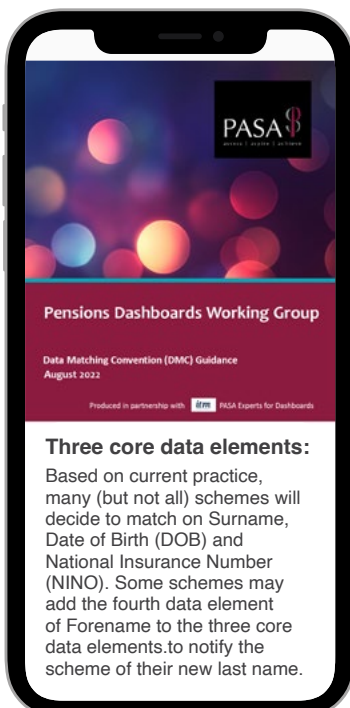
Images courtesy of the [Pensions Dashboards Programme](#)



ITM research approach:

For our sample schemes, we copied the Scheme data already held, and then enhanced it using verification and tracing with other data sources (3).

We then used this enhanced data as a batch of c.200,000 simulated Find Requests (sFR's) and compared these back against the original Scheme data to assess different match criteria.



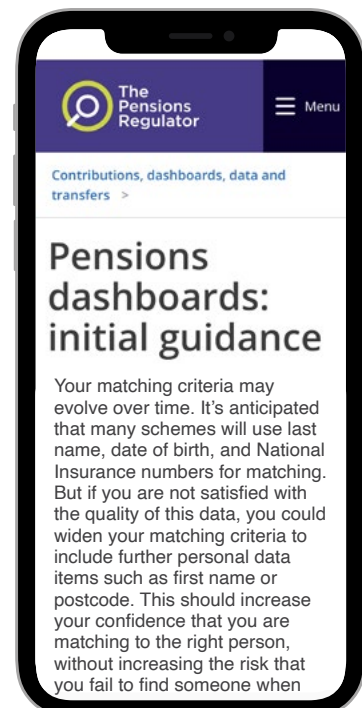
2.A.2 Approach to the research match tests:

Both the PASA DMC Guidance and the TPR initial dashboards Guidance on matching anticipate the starting point for many schemes will be to think about matching on Last Name, DOB and NINO.

So these three data elements were a key initial part of our Test match criteria.

We then tried to match using different Test match criteria / data elements to see how the matching performance changed.

Based on our findings, we believe schemes may wish to use more than one set of match criteria, in combination. We call this combination of different match criteria a "match policy".



2. Research process and findings

2.B. Refining the research dataset

Below are the steps we took to refine the sample schemes' personal data in order to simulate Find Requests for our matching research.

The key point is that we verified and traced the schemes' currently held personal data with other sources in order to enhance this data so it better simulated the real world Find Requests which will be received by data providers.

Step	Description
B.0	We identified some sample schemes to use as part of our research, using their real personal data held.
B.1	Our first step was to take a copy of the personal details held for the active and deferred memberships of our selected pension schemes. These included some large defined contribution (DC) pension schemes used for automatic enrolment (AE), including master trusts with multiple employers, and some small defined benefit (DB) schemes. In total, this amounted to more than 250,000 records.
B.2	We then verified the copy data against another data source to check that all the member records had correct addresses, ensuring the addresses of the copy data matched the Royal Mail PAF format (this helped improve the addresses in our simulated Find Requests which we wanted to use in one of our Test match criteria).
B.3	We then reduced our copy data down to just the verified records, for example removing members living overseas, and members with UK addresses that could not be resolved. These then became the traceable records.
B.4	We compared, or traced, the records from step B.3 with a further data source, highlighting inaccuracies (in the schemes' original data) in some members' names, dates of birth (DOBs) and addresses. It should be noted that this could not include NINO, as this is not available from tracing data sources. This was the key step in our research approach as it enabled us to simulate up-to-date Find Requests, at scale, and then test how well these would match against the original scheme data.
B.5	We then updated our verified copy data (from step B.3) with the more accurate / more up-to-date names, DOBs and addresses, as it will be this more accurate information which will be "passed down" in Find Requests from the central digital architecture. See Important note on tracing below about schemes doing this themselves as part of their data management and improvement plans. We now had two similar, but crucially different, sets of data: B.5 Enhanced up-to-date data to make our simulated Find Requests (sFR's) as realistic as possible B.0 The original, and in numerous cases, out-of-date scheme data to be compared against our sFR's.
B.6	Our final data preparation step was to create the list of the simulated Find Requests (sFR's) where we were certain the person in the sFR data (B.5) was the same person in the original scheme data (B.0). This gave us a solid base of records where we would expect a match to be made: in other words, if the original scheme data (B.0) was accurate and up-to-date in all cases, then we knew we should expect 100% of records to be matched. After this final data preparation step we had approximately 200,000 records where we had simulated Find Requests - and this number represented our 100% target. Our ambition, by running different Test match criteria (between the B.5 and B.0 data), was to see how nearly each match criteria could approach the 100% match target. This is covered in Step C below.

Important note on tracing

Of course, tracing members is not new. Schemes have been doing this for years, especially in respect of individual member events causing pensions to crystallise, such as retirement. Some schemes also undertake bulk tracing exercises on a regular basis.

One challenge, however, can come in updating members' records with more up-to-date information, such as later Last Names or Addresses, which are found through tracing.

In our simulation (step B.5), we updated our copy data (B.3) with the more accurate traced information, and made the reasonable assumption that these more accurate details are what would be received by schemes in pensions dashboards Find Requests.

So one question is – to what extent can schemes update their records with what appears to be more up-to-date information obtained from tracing, to improve their data accuracy ready for pensions dashboards?

The answer from tracing exercises varies, but from ITM's experience a reasonable expectation is that when a group of known disengaged deferred members are submitted for tracing, a scheme would end up updating records for c.40% of them, and confirming currently held details for c.20% of them, giving a c.60% return.

The level of member involvement required before records are updated can depend on the strength of the tracing result, for example credit reference agency activity in the last three months indicating a change of address is a very strong result.

The remaining 40% who wouldn't get updated (or confirmed) reflect the fact that obtaining member consent to update records with more up-to-date information can be very challenging amongst a highly disengaged membership, especially disengaged deferred members. It also reflects the fact that where tracing results differ from scheme records, there will usually be some cases where it's the scheme records which will be shown to be correct.

Anything like a 60% return is still very valuable preparation for dashboards, and schemes can and should continue with bulk tracing exercises to improve the accuracy of the personal data they hold. In July 2022, PASA published useful [Dashboard Accuracy Data Guidance](#) on this specific topic.

But as explained above, there are limits to the data accuracy improvements which can actually be implemented through tracing.

Trustees are therefore going to need sophisticated approaches to maximise matching (covered in Step C).

But the good news is, once dashboards are launched, engaged dashboard users may be keen to correct their information with schemes when they receive "possible match" responses, as well as potentially consenting to their schemes holding their email address and mobile phone details, dramatically improving ongoing engagement.

At a later date, the central Consent Service could potentially be enhanced to enable the dashboard user to consent to these data updates happening with schemes automatically, but more research may be needed to make that a reality, welcome as it might be.

In the above ways, dashboards can be seen as a wonderful new tool for ongoing data management, where there can be ongoing self-improvement of data following the launch of dashboards to the general public from the Dashboards Available Point (DAP) onwards.

2. Research process and findings

2.C. Running the research match tests

Across c.200,000 member records, we attempted to match simulated Find Request (sFR) data with the original scheme data, using various different Test match criteria.

Remember, from the preparatory work we did in Step B above, these were all records where we would expect a 100% match success (if the scheme data was 100% accurate).

Below is a summary of the Test match criteria we ran, with the research result for each test shown. Further details and discussion for a few of the Test match criteria we used, and the relevant findings, are shown in Annex II.

It's important to understand our sequence of Test match criteria below isn't necessarily what we would recommend trustees adopt as their match policy (especially as matching decisions should be scheme-specific). Rather, we structured the sequence of tests in a way which we felt would best illustrate how different match criteria might perform.

What did we find?

On the schemes we researched, with a starting total of over 250,000 member records, the proportions of simulated Find Requests which matched against the schemes' original records were:

	New Matches	Total Cumulative Matches
Test match criteria 1: Exact match on Last Name and DOB and NINO	90.80%	90.80%

Findings narrative (see Annex II for more): It's likely many trustees will be comfortable instructing their data provider to send a match made response back to the dashboards ecosystem where there is an exact match on LN / DOB / NINO. This is the core Matching Option 1 in the [**PASA Data Matching Convention \(DMC\) Guidance**](#).

This "9 out of 10 pensions found" base isn't a bad starting point. However, for the reasons and assumptions explained in the Summary section above, this can perhaps be seen as a "best case" outcome for exact matching on LN / DOB / NINO. In reality, maybe only 8 (or even 7) out of 10 pensions might be found with LN / DOB / NINO.

But in any case, 9 out of 10 isn't good enough, as finding every pension matters. This is the core reason we believe trustees will wish to specify a match policy which incorporates a number of different match criteria aiming to nudge up their find rate as close as possible to 100%, as illustrated by our Test match criteria 2 to 10.

New Matches Total Cumulative Matches

Test match criteria 2:

Exact match on First Name and DOB and NINO
(on the unmatched records from Test 1)

3.67%

94.47%

Findings narrative: FN / DOB / NINO is unlikely to be a match criteria schemes would use on its own. In our research, on its own, it achieved an 89.65% match rate (less than Test 1's 90.80% match rate on LN / DOB / NINO).

However, when used in combination with LN / DOB / NINO (i.e. Test 1), then FN / DOB / NINO can help "nudge up" the find rate from 90% up towards 100%: in our research, by 3.67% to make a cumulative 94.47% matches.

For these additional 3.67% of our simulated Find Requests which matched on FN / DOB / NINO (against the 9.20% unmatched population from Test 1), we can deduce that, because they didn't match in Test 1, it is Last Name which doesn't match. (In other words, because DOB and NINO matched in Test 2, it must have been a non-matching LN that caused Test 1 to fail.)

So the trustees would know they do not have an up-to-date or correctly spelt Last Name to match against on any of these 3.67% of their deferred and active records, should these individuals choose to use a dashboard. Trustees could decide to use this as an opportunity to instruct their ISP to send a possible match response back to the dashboards ecosystem. This means the (possible) member would be asked (via the dashboard they're using) to contact the scheme through normal pension scheme channels, and evidence their up-to-date Last Name so the scheme's records can be updated accordingly. Subsequent Find Requests would then match on LN / DOB / NINO (i.e. match criteria 1) which would be a good outcome for all - the member and the scheme. Alternatively, trustees could decide that they're comfortable saying these FN / DOB / NINO matches are actually matches made, increasing their find rate in our example by 3.67% to 94.47%. So trustees need to weigh up maximising their find rate against improving their data accuracy (assuming that members in such cases do actually contact the scheme to get their Last Names updated). This is true for all the following Test match criteria.

Test match criteria 3:

Exact match on First Name and Last Name and NINO
(on the unmatched records from Test 2)

2.58%

97.05%

Findings narrative: In terms of find rate, we're now getting into the high 90s which is better.

For an exact match on FN / LN / NINO, which failed both LN / DOB / NINO (Test 1) and FN / DOB / NINO (Test 2), we can deduce that it's DOB which doesn't match. It's quite significant that the DOB doesn't match on 5 out of every 200 of our sample records, even when noting the caveat in our comment earlier that sometimes scheme records prove to be correct when compared with tracing data. This isn't just important as a personal data identifier, but DOB can also significantly affect scheme benefits. Trustees might therefore want to treat these as a possible match, encouraging the individual (via the dashboard they're using) to get in touch with the scheme directly with evidence of their correct DOB so the scheme's records can be updated accordingly, or alternatively the scheme-held DOB can be verified as being correct.

2. Research process and findings

	New Matches	Total Cumulative Matches
--	-------------	--------------------------

Test match criteria 4:

Exact match on Last Name and First Name and DOB and Postcode (on the unmatched records from Test 3)

0.65%	97.70%
-------	--------

Findings narrative (see Annex II for more): NINO cannot be a mandatory user input on dashboards because not all users will have NINOs (although users will be encouraged to provide it where they have one). Also schemes often have members for whom they don't hold valid NINOs because employers have never provided them.

Remember that LN / FN / DOB / Postcode (PC) are all attributes in the Find Request data which will have been verified by the central Identity Service of the dashboards ecosystem (as explained in Annex I). Schemes can therefore have a good level of confidence in these four data elements being passed to them in a Find Request. Using LN / FN / DOB / PC, we increased the find rate by two thirds of a further percentage point - every % matters!

Trustees might want to call this a possible match, requesting the user to input their NINO to a dashboard next time, so they match under match criteria 1 (LN / DOB / NINO), or alternatively to correct the invalid NINO held by the scheme (which potentially caused them to fail Tests 1 to 3).

However, trustees might be uncomfortable even saying this is a possible match, given no match has been possible on NINO, but if they know that they don't hold a valid NINO for a member then this may be the only way of complying with their obligations to allow that member to find their pension.

Test match criteria 5 & 6:

Same as Test 3 but LD1/2 on FN (Test 5) or LN (Test 6) (on the unmatched records from Test 4)

0.20%	97.90%
-------	--------

Findings narrative: In Test match criteria 5 and 6 we were repeating the Test 3 match on FN / LN / NINO except that we applied a more sophisticated approach to matching on FN (in Test 5) and on LN (in Test 6).

The greater sophistication was not to search for an exact match on FN (in Test 5) or on LN (in Test 6). Instead we searched for a FN / LN held in the scheme data which was "almost the same" as the FN / LN in our simulated Find Request.

We used a number of more sophisticated techniques to match on names which were almost the same.

For example, we looked for names which had a Levenshtein Distance (LD) of 1 between them (or 2 for names longer than six characters). The Levenshtein Distance (LD) between two strings of characters, or words, is the minimum number of single-character edits (including substitutions, insertions or deletions) required to change one word into the other. It is named after the Soviet mathematician Vladimir Levenshtein.

So this would make a match where there is just 1 (or 2 in longer names) character(s) different between the name held by the scheme and the name received in our simulated Find Request. For example, Sonia Davis would be matched with Sonia Davies, or with Sonja Davis, remembering of course that there was also a match on NINO.

As you can see, these two Test match criteria 5 and 6 combined did not make a huge many more matches, i.e. just 0.20%. But in a scheme with 100,000 deferreds and actives, that's still 200 more pensions which would be found, and finding every pension matters.

Trustees might wish to signify this as a possible match in order to get whichever name (First or Last) is incorrectly held by the scheme corrected by the member with their consent.

New Matches Total Cumulative Matches

Test match criteria 7:

Transposed LN & FN and exact match on DOB and NINO (on the unmatched records from Tests 5 & 6)

0.59% 98.49%

Findings narrative: In Test match criteria 7, a further aspect of greater matching sophistication we tested was on transposed First and Last Names, as well as matching exactly on DOB and NINO.

So for example this might be where the scheme holds the member's FN & LN as Davis Sonia, whereas in the Find Request the FN & LN are correctly Sonia Davis.

This was a surprisingly common occurrence in our sample schemes' data, enabling us to make 0.59% more matches. That would mean 590 more pensions being findable in a 100,000 member scheme, which is significant. Trustees might feel this could signify as a match made, or they may wish to say it is a possible match, enabling the First and Last Names held by the scheme to be correctly swapped, with the member's consent.

Test match criteria 8 & 9:

Same as 1 but with Levenshtein Distance 1/2 on LN (8) / DOB (9) i.e. Levenshtein Distance 1/2 on LN (Test 8) or 1 digit difference on DOB (Test 9) and exact match on NINO (on the unmatched records from Test 7)

0.13% 98.62%

Findings narrative (see Annex II for more): At 98.49% after Test 7, we were still just shy of the magic 99%, so we wanted to keep going with additional test match criteria.

In Test match criteria 8 we repeated Test 1, i.e. LN / DOB / NINO, but using the same Levenshtein Distance (LD) test of 1 (or 2 for names longer than six characters) that we had used in Test 6 above.

In Test match criteria 9 we wanted to test some more sophisticated matching on DOB. From our extensive experience dealing with pensions data over many years, we know that DOB is often only slightly wrong. So we searched for exact matches on LN and NINO (like Test 1) but looked for matches on DOB which were only slightly different, such as a single digit being different in the DD, the MM, or the YYYY part, e.g.

Scheme data

Find Request data

01/05/1972

01/06/1972

Given that LN and NINO matched exactly in these cases, it seems very plausible that the original data feed from the employer was mistyped by the employer (when the member first joined the scheme), and the member has never spotted the error (i.e. their incorrect DOB) shown, for example, on annual benefit statements.

In combination, Tests 8 and 9 made another 0.13% of matches. As with Tests 5 and 6, trustees might wish to signify these as possible matches in order to get whichever of Last Name or DOB is incorrectly held by the scheme corrected by the member with their consent.

Test match criteria 10:

Exact match on FN and LN and DOB (on the unmatched records from Tests 8 & 9)

0.64% 99.26%

Findings narrative: We still hadn't reached 99%. So we wanted to try further "loosening" the match criteria to just FN / LN / DOB, i.e. the same as Test 4 above but without the match on Postcode. As you can see, this made a useful 0.64% further possible matches.

But if trustees are uncomfortable saying the Test 4 matches were possible matches (given no match had been possible on NINO), then they will definitely be uncomfortable with the Test 10 matches, where the Postcode test is also taken away.

However, surely there's a possibility that a user with exactly the same First Name, Last Name and Date of Birth could be our member? This is a good example of where trustees may feel they are unacceptably increasing the risk of finding wrong pensions, sometimes referred to as "false positives". We discuss the risks of finding wrong pensions below.

2. Research process and findings

What about the final 0.74% unmatched records?

Knowing you can match 99.26% of all your scheme’s pensions might feel good. 0.74% of your scheme’s records may not be very many records, but, across the whole UK pensions universe this could equate to c.740,000 pensions not being found.

These 740,000 pensions could be very important to the people they belong to. So finding every pension matters.

As a scheme, what you really want is for your match policy to be the Domestos of matching, i.e. “finding [at least] 99.9% of all known pensions”. Dashboards are about making people’s pensions visible,

correcting schemes’ personal data as appropriate through consent-based resolutions of possible matches. As pre-war US Supreme Court Justice Louis Brandeis said, and the **Transparency Task Force reminds us, “sunlight is the best disinfectant”**.

But how far should trustees / data providers go?

DWP has clearly set its expectation (in Chapter 2, paragraph 9 of its January 2022 consultation on the draft Dashboards Regulations) that schemes must ensure they “take reasonable, diligent steps to search for matches”. In due course, further Guidance on matching from DWP or TPR may support and expand on this expectation.

But you can see from the later Test match criteria we used that we were finding fewer and fewer matches. In this context, what are reasonable and diligent steps?

There may also be a question of system performance to be investigated if all data providers are running a very large number of multiple match criteria.

We wanted to explore what could be learnt from the final 0.74% of records in our research about how far schemes might need to reasonably go to try to match these records. What was it about this final very small minority of records which is causing them not to be matched by any of our Test match criteria from 1 through to 10?

Description of the final unmatched records, and what trustees could do about them

DOB and NINO matches	<p>For just over half this final small group, there was a match on NINO, and very often on DOB too, but various data issues with FN / LN / PC (and DOB in some cases) meant these records didn’t match under any of our 10 Test match criteria, not even with our LD1/2 and transposed sophisticated matching on LN / FN and on DOB.</p> <p>Trustees might therefore wish to consider how comfortable they would be returning a possible match where DOB and NINO match but FN / LN / PC don’t.</p>
NINO no match and near match	<p>For the other half, i.e. where there was no match on NINO, often some combination of FN / LN / DOB / PC did match, or nearly match, but not all four of these elements exactly matched, otherwise they would have passed our Test match criteria 4, or our Test match criteria 10 without PC.</p> <p>Trustees might be uncomfortable not matching on NINO, so more work on why these weren’t matching could be done in due course.</p> <p>Near matches on NINO, e.g. due to a user making a slight error when inputting their NINO such as transposing two digits, is something else which could be tested in the next phase of PDP’s beta testing. As noted above, we could not include more accurate NINOs in our simulated Find Requests in this research because it is not available from tracing data sources.</p>
Further loosening match criteria	<p>“Loosening” the possible match criteria even further, however, to (for example):</p> <p>Exact match on FN / LN plus a Near match (1 digit different) on DOB</p> <p>potentially opens up the risk of having quite a few results flowing into the possible match resolution process, sometimes where this match criteria has found the wrong pension, which trustees may not be comfortable with. We explore this below.</p>
Data corrections	<p>Ultimately, finding ways to get FN / LN / DOB / PC corrected on the schemes’ records would, of course, enable all of these records to be matched via Test match criteria 4 or 10.</p>

Risk of finding wrong pensions

As well as assessing the performance of various match criteria on real data, and exploring how various sophisticated matching techniques might help, we also wanted to begin the debate on how you can gauge the extent to which some match criteria could have an unacceptable risk of finding wrong pensions.

Finding wrong pensions is actually two distinct problems. The one most commonly talked about is the "data breach" scenario where a match made response is given following a Find Request, but the user making the Find Request is not actually the scheme member - clearly a very undesirable outcome.

The second problem is where possible match responses given following Find Requests "over reach" and end up inviting too many users who are not scheme members to enter into the possible match resolution process.

For possible match resolutions, some ultimately wrong results (i.e. it's not actually the user's pension) are probably acceptable for scheme administrators and pensions dashboard users - but too many could be nothing short of calamitous, creating potentially huge amounts of nugatory extra work.

To deal with both problems we need to have some way to go about assessing how likely a match criteria is to give an incorrect response. But how can we actually do this?

One answer is "gut feel". For example, our Test match criteria 1 to 9 - see chart in the Summary (Page 08/09) - all have the feel of safe criteria where the risk of finding the wrong person seems to be negligible.

Eight of these nine match criteria include matching on NI Number (NINO), which in combination with the other fields naturally strengthens the robustness of the match. The one match criteria in these nine which doesn't include NINO - Test 4 on FN / LN / DOB / PC - seems robust, particularly as these are the four data elements which have been centrally verified before being sent to schemes for matching (noting however the twins scenario discussed in the next section).

But once we move to the slightly looser criteria of Test 10 (FN / LN / DOB) that might be considered for possible match responses, particularly in the absence of NINO matches, then gut feel tells us we're straying into territory where the risk of finding wrong pensions is going to increase. But we don't know by how much and hence whether it would be an acceptable risk - so we need some ways of assessing this!

As such looser criteria are only likely to be used for possible match responses, then another way of assessing how many incorrect responses are made to Find Requests is to wait and see what happens when dashboards are used by high volumes of consumers.

But unfortunately that could easily be after the horse has bolted, with poor outcomes for dashboard users and large queues of possible match resolutions for your administrator to handle - much of which could be wasted effort and resolved as "it's not the user's pension".

So we need to devise methods to help us assess this up front, so schemes can set match criteria - particularly for possible match responses - which won't fail them when dashboards are in wide use. We've carried out some analysis on a potential method, introduced below.

2. Research process and findings

Devising methods to predict the risk of finding wrong pensions

While Test match criteria 4 and 10 (exact match on FN / LN / DOB / PC and FN / LN / DOB respectively) should be robust enough for consideration, some trustees may have concerns about it finding the wrong pensions.

One specific, although minority, scenario to be considered if the match on FN was just on an initial as opposed to the full First Name, is multi-birth siblings, e.g. twins, triplets, etc. Whilst these would have the same DOB, and originally the same LN (and they still might), the full First Names would be different, but just the initial character of FN could be the same.

More generally, if a Find Request matches, for example, against a scheme record on FN / LN / DOB, or even FN / LN / DOB with 1 digit difference (we didn't test this), how likely is it there is someone else out there of that exact name, and born on that day - and how many others might there be?

One way to attempt to address these questions is to use available statistical data sources to construct a model to estimate how likely it is there are individuals in the dashboards user population who share the same FN / LN / DOB (and maybe PC) data as your member. Statistical data sources that could be used for this kind of analysis include:

- Common first and last names, and their prevalence in the UK
- Historical data on the numbers of individuals born on any particular date in the UK
- Data showing the number of households and individuals attached to a post code.

In addition the model could go further and try to take account of the profile of expected dashboards "early adopters", for example the expectation that users who are closer to retirement are most likely to be early users. This would make sense to do as the aim is to predict early usage - once dashboards are being used at volume then there will be plenty more data to make predictions into later years!

This will never be an exact science, but we've started to explore this using models that estimate the prevalence of FN / LN / DOB (and PC) in the population at large.

The following example helps to demonstrate what we've been modelling:

For example, looking at popular first names and last names, the most common names of all are around 15 times more prevalent than the 150th most common names, even though ranking 150th still means you have a pretty popular name! For example, there may be 15 people with last name Smith for every one with last name Riley, and 15 people with first name David for every one with first name Alex. Once you get to more unusual names then there could be hundreds or thousands of people named Smith or David for every one person with the unusual name.

This example matters because an FN / LN / DOB match criteria may be absolutely fine for possible matching for most of your scheme members, but there may be some scheme members whose first names and last names are so prevalent that there is a high risk those members' records could be "possibly matched" a number of times incorrectly with dashboards users who happen to share their FN / LN / DOB, resulting in an incorrect response each time.

The DOB part of FN / LN / DOB is very different however - the prevalence of different dates of birth shows far less variation apart from cyclical patterns within the calendar year, and trends over time.

We're continuing to look at this area, and how analysis like this can be used to help with advice on match criteria to use for possible matching, and also to support the ongoing process of possible matching once dashboards are in use.

2.D. Unpicking the research findings

Searching for an exact match on LN / DOB / NINO is a good start, but it won't find all your pensions. Matching Previous LN (if it's input by the dashboard user) against the scheme's LN could be helpful, but again, this won't find pensions where there is a problem with DOB or NINO, or some issue with LN other than a straightforward change on, say, marriage.

Knowing that attempting to match on LN / DOB / NINO, or Previous LN / DOB / NINO, won't find all your pensions, we believe trustees will almost certainly wish to decide on a more sophisticated match policy incorporating a combination of different match criteria to increase their find rate as much as possible towards 100%.

Furthermore, sophisticated techniques, such as Levenshtein Distance, can also help nudge up the find rate further.

Where a match is made but it isn't exact, trustees may wish to signify this in their match policy as a possible match. This will enable the dashboard user to get in touch directly with the scheme and consent, through existing scheme channels, to any data corrections that may be required, so that next time they use a dashboard, their pension will be matched exactly. This will also minimise the risk of presenting a user with the wrong pension.

A further consideration is the volume of member queries possible matches could potentially create. We believe there are practical solutions that could reduce this administrative impact on schemes, such as providing a unique reference to the dashboard user when a possible match is made. This will allow the user to supply the reference to the pension administrator, who in turn can use the reference to match with the relevant possible match from the dashboards data provider.

By simplifying the administrative process through such examples it will allow the industry to use possible matches to their full potential, improving scheme pension data and improving member experiences.

For each of the different individual match criteria (in a sequence of match criteria), trustees must decide whether they wish to return a match made or a possible match response. This is illustrated below (noting that these aren't recommended match criteria, nor a recommended sequence of match criteria).

Key message: A multi-match criteria match policy, incorporating sophisticated techniques within some of the match criteria delivers:

- a) better matching results, i.e. a greater level of compliance with new duties on trustees and
- b) "self-improvement" of data accuracy through member-led resolution of possible matches.

Figure 4:

Trustees' chosen match criteria sequence	Trustees' chosen match criteria specification	Trustees' chosen match criteria Response
1	Exact match on all of LN / DOB / NINO	<div style="background-color: #008080; color: white; padding: 10px; text-align: center;"> <p>Match made</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <p>Possible match</p> </div>
2	Exact match on all of FN / DOB / NINO	
3	Exact match on all of FN / LN / NINO	
4	Exact match on all of FN / LN / DOB / PC	
5	FN (LD1/2), LN (Exact) / NINO (Exact)	
6	FN (Exact) / LN (LD1/2) / NINO (Exact)	
7	Transposed FN & LN / DOB (Exact) / NINO (Exact)	
8	LN (LD1/2) / DOB (Exact) / NINO (Exact)	
9	LN (Exact), DOB (1 digit diff) / NINO (Exact)	
10	FN (Exact) / LN (Exact) / DOB (Exact)	

3. Recommendations and next steps

We hope the findings of this matching research, on the real personal data held by some sample large defined contribution and some smaller defined benefit pension schemes, will help trustees and pension providers in their deliberations on matching, regardless of which ISP they choose to connect their scheme to the dashboards ecosystem.

Our research has highlighted that this clearly isn't the end of the debate on matching.

We took a research approach which aimed to illuminate the challenges and suggest ideas for solutions. But schemes need to do scheme-specific analysis to decide what their own match policies should be.

Building on our research findings, the table below proposes key future actions for:

- pension trustees and providers,
- the Pensions Dashboards Programme, and
- the pensions industry as a whole.

Actions for:	Actions	When
Pension trustees and providers	<p>Scheme-specific conversations, investigations & DPIAs:</p> <p>Have detailed conversations with your chosen ISP(s) about what match criteria, and possible match criteria might be most appropriate for your particular scheme.</p> <p>This may involve thoroughly investigating the personal data you hold for your deferred and active members, and also tracing / verifying this data with external agencies.</p> <p>You may decide you wish your ISP(s) to adopt a suite of different match criteria in combination (we call this a "match policy").</p> <p>For each match criteria you must decide whether you wish to return a match made or a possible match response, balancing maximising matches made with the opportunity to make data corrections (via the member-led consent-based resolution of possible matches).</p> <p>Consider where a specific match criteria might invite too many members to follow the possible match process, such as the examples earlier in this report for match criteria FN / LN / DOB, and decide if you are still happy to trial it in your possible match policy and keep its performance under review.</p> <p>The Information Commissioner's Office (ICO) has said that because schemes are matching for dashboards they should consider completing (or updating) and publishing a Data Protection Impact Assessment (DPIA).</p> <p>For the setting of their match policies, the ICO has said schemes' DPIAs should focus on data minimisation, i.e. what is the minimum but adequate amount of personal data that schemes need to fulfil their matching duties?</p>	From Autumn 2022 onwards

Actions for:	Actions	When
Pensions Dashboards Programme (PDP)	<p>Possible match awareness:</p> <p>PDP has already published information on the importance of schemes' matching processes, including possible matching. As we enter the phase where trustees need to consider their matching policy, these matching resources need to be promoted, increasing trustees awareness and improving their understanding of the intricacies and benefits of possible matching.</p> <p>Scale consumer testing:</p> <p>Numerous questions must be user tested in the beta phase, and then tested at scale once schemes are connected from the initial deadline of August 2023. These questions include:</p> <ul style="list-style-type: none"> • What proportion of dashboard users input their NINO? • What proportion input it correctly? (for example, how many will transpose two digits on input?) • What proportion of users input other self-asserted data elements, such as Previous Last Name? • How will consumers react to the decisions on matching taken by their different pension schemes & providers? • How will users react to possible matches? • How can the UX design of the end-to-end "possible match / data correction journey" be optimised to deliver ongoing improvements in schemes' data accuracy? <p>The findings and learnings from all the above, and other, testing, will enable ISPs and schemes to further enhance their matching policies in light of user experience.</p>	From beta testing phase, but at scale from Autumn 2023 (once Cohort 1(a) of schemes and providers have connected to the pensions dashboards ecosystem)
Pensions industry as a whole	<p>Further learnings, research and refinements:</p> <p>As more and more schemes carry out the scheme-specific investigations above, further learnings will accrue. And further insights will also flow from the beta testing on matching described above. These additional learnings are likely to highlight the need for further industry research, building on the concepts in this initial report. Here is a list of potential areas where we already know further industry research may be required (not an exhaustive list):</p> <ul style="list-style-type: none"> • Similar research on schemes of other types, for example schemes which have split administration across different data providers, including where external AVC providers are connecting to dashboards directly • Relative impacts of verified identity attributes and self-asserted data elements in the Find Request data, in particular Previous Last Name • Analysis of more complex Last Name mismatches • Further understanding around the criticality of NINO and what to do where there is no NINO • Multiple records, i.e. where an individual has more than one pension entitlement within a scheme • Further understanding around the risks of finding wrong pensions (also known as "false positives") , particularly for possible match policies where match criteria are by definition going to be "looser" • Mitigation of potential fraud around matching • Multi-birth siblings, e.g. twins, triplets, etc. 	2023

Annex I

Key concepts and terminology used consistently in this research report

Everybody should watch PDP's first two excellent introductory videos:

- [An introduction to the pensions dashboards ecosystem](#) (128 seconds)
- [An introduction to data standards](#) (93 seconds).

These basic videos include various terms which are defined in:

- PDP's [Glossary maintained on the PDP website](#)
- DWP's [Schedule 1 Interpretation in the indicative draft Pensions Dashboards Regs](#) (final regulations expected in Autumn 2022)
- TPR's [Initial pensions dashboards guidance for trustees](#).

Of all the terms in the PDP Glossary / Regulations Schedule 1, the key ones you really need to understand to get the most from this matching research report are shown below, together with terms that we have created for this project or as part of our ISP development.

We use the terms in the below table consistently throughout this report.

Terms	Abbreviation	Source and / or Notes
Date of Birth	DOB	
Find Data , contained in a Find Request , comprising <ul style="list-style-type: none"> - Verified Identity Attributes, and - Self-Asserted Data Elements 	FD / FR	Reg 2 Sched 1, and Find data also has a PDP website page all of its own Reg 2 Sched 1, defined as First Name, Surname (Last Name), DOB, Current address Reg 2 Sched 1: NINO, Previous names and address, Email address, Mobile phone number
First Name	FN	
Last Name	LN	
Matching Process (to be reported to TPR)		Reg 27(2)(b) although Reg 2 Sched 1 just calls this "matching" Note that a "Matching Process" also includes seeking to resolve Possible Matches
Match Criteria		Reg 22(1) prescribes that "trustees must decide on criteria to use for matching"
Match Policy , a suite of different Match Criteria		Match Policy is an ITM proposed term to define a combination of different match criteria covering how a particular scheme intends to both make matches and possible matches
National Insurance Number	NINO	
Positive Match , which can be either a <ul style="list-style-type: none"> - Match Made, or a - Possible Match 		Reg 2 Sched 1, although the PDP Glossary simply calls this a "match" "Match Made" only appears in the Regs in the plural (Reg 27(2)(c)(i) to be reported to TPR) Reg 2 Sched 1: "a positive match means a match which is either made or a possible match"
Postcode	PC	
Simulated Find Request	sFR	Created for this research project

The PDP videos above include a graphical representation of the dashboards ecosystem. You may also find it helpful to refer to the simplified ecosystem diagram which appeared in the [PDP Guest Blog](#) from Moneyhub in July 2022:



With reference to the “layers” in the above diagram, the overall matching process is as follows:

- A layer 1 individual, using whichever dashboard they wish to in layer 2, inputs their personal details to that dashboard, including optional “self asserted” data elements (like NINO, Previous name and Previous address), then
- the Identity Service in the layer 4 Central Digital Architecture (CDA) verifies the user’s identity (the very few identity attributes which are verified are First Name, Last Name, DOB and Current address), then
- assuming the user consents to this (via Consent & Authorisation Service in the layer 4 CDA), their personal details, i.e. verified identity attributes plus any optional self-asserted data elements input by the user, are passed, by the layer 4 Pension Finder Service, to all connected layer 5 data providers (this is the Find Data being passed to all data providers in a Find Request)
- all layer 5 data providers then undertake matching, using the match policies specified by their layer 6 trustee clients (there are likely to be different match policies for different schemes), returning a match made or a possible match to the CDA as appropriate.

Annex II

Further details about our research findings

Below is further detailed information about our research findings for some of the Test match criteria.

Test match criteria 1 - Exact match on LN / DOB / NINO

Members who failed to match with this Test match criteria would have failed because one or more LN and / or DOB and / or NINO did not match between our “enhanced” simulated Find Request data and the original scheme data.

Test 3 uncovered cases where DOB didn't match, and Test 4 looked at matching without using NINO.

So, focusing just on issues with Last Name not matching, a very common scenario will be where the member's LN has changed, e.g. on marriage, but the member has failed to notify the scheme, so the scheme still holds the (now) out-of-date LN.

The PASA DMC Guidance considers an Enhanced Matching Option 1 which attempts to match the Last Name held by the scheme with the Previous Last Name included in the Find Request data (if the dashboard user chose to voluntarily input it to the dashboard they're using).

We did not research this as a Test match criteria as it was not straightforward to emulate this optional user behaviour in our simulated Find Requests.

We could have made a blunt assumption that all the changed Last Names in our “enhanced” data were because the member's Last Name had changed, and hence they might choose to input their Previous Last Names to a dashboard, which would then have matched with the Last Name held by the scheme. But it wasn't clear exactly which were genuine changes of this type.

Some Last Name differences were slight, and our Levenshtein Distance tests aimed to pick those up. Others were more significantly different, but still looked like they could be the same name with maybe an input error by the employer when the name data was first passed across from the employer.

Multi-part names were also an issue. For example, sometimes the scheme would hold just one Last Name, but the Find Request would include a double-barrelled Last Name, half of which was the same as the LN held by the scheme. Three part name differences also occurred, for example van der Valk in the sFR whereas the scheme just held Valk.

More research work could be done in this area - see Recommendations and next steps.

Test match criteria 1a - Exact match on FN / LN / DOB / NINO

Where LN / DOB / NINO all match exactly (i.e. Test 1), schemes can have strong confidence this is a match made.

But if an attempt is also made to match exactly on FN, the find rate can only decrease. In our research, it reduced from 90.80% (Test 1 above) down to 85.98%.

This research finding aligns with the assumption made in **Annex C (page 35) of the PASA DMC Guidance** and also reflects **responses to the PDP Call for Input on data standards** (paras 81-85, page 15), i.e. that adding in FN to an already exact LN / DOB / NINO match only reduces match success, and therefore may not be a match criteria schemes will find helpful.

Test match criteria 4 – Exact match on FN / LN / DOB / Postcode (PC)

Some trustees might be uncomfortable not using NINO to match. We don't yet know how often users (who do have a NINO) will choose to input it to dashboards, or how accurately they will input it. We also know there will be users of dashboards who do not have a NINO at all.

So if NINO isn't provided by the user, or is held incorrectly on the scheme records, Test match criteria 4, i.e. FN / LN / DOB / PC, could be useful. It's worth remembering that these will all be verified identity attributes in the Find Request data, which can be argued to give them greater weight.

With names, we can be more sophisticated, e.g. using Levenshtein Distance to spot a name which is almost certainly the same except for 1 or 2 characters. But with NINOs, it's not so easy to spot issues like this and assume that two very similar NINOs should be the same.

We think match criteria 4 could be very important for schemes to maximise their possible matches. More research work on NINO mismatches could also be done, most likely once a body of experience has built up from scale dashboards testing.

Test match criteria 8 – Exact match on NINO and DOB plus LD1/2 on LN(8)

On its own, i.e. not in combination with the other Tests, Test match criteria 8 was our best performing Test match criteria.

Test match criteria 1, i.e. exact match on LN / DOB / NINO, made 90.80% matches.


Test 8 is a more sophisticated version of Test 1, still searching for an exact match on DOB and NINO, but searching on LN for a Levenshtein Distance of 1 (or 2 for LNs over six characters long). This made 94.06% matches.

Trustees might therefore wish to start with this match criteria. However, an advantage of splitting out Test 1 and Test 8 is that matches under Test 1 can be returned as matches made, whereas those under 8 can be returned as possible matches, enabling the Last Name held by the scheme to be corrected as appropriate.

About the research team



Ross Milligan
Senior Analyst, ITM

 [ross-milligan](#)



Chris Gill
Product Manager, Altus

 [christopher-gill](#)

Contact

rorytucker@itm.co.uk
pensionfusion.com
itm.co.uk




Maurice Titley
Chief Innovation Officer, ITM

 [maurice-titley](#)



Rory Tucker
Technical Research Analyst, ITM

 [rory-tucker](#)



Richard Smith
Independent Pensions
Dashboards Consultant

 [richard-smith](#)

Pension Fusion offers you a secure, reliable and hassle-free connection to Pensions Dashboards – brought to you by recognised experts in pension and financial services.



Pension
Fusion

Altus
An  equisoft company

itm
Solution delivered...



pensionfusion.com



ITM

22 Tudor Street,
London,
EC4Y 0AY

T: 020 7648 9990

E: info@itm.co.uk

Altus

Lower Ground Floor,
Queen Square House,
Queen Square Place,
Bath,
BA1 2LL

T: 01225 438 000

E: enquiries@altus.co.uk

Altus
An  equisoft company

itm
Solution delivered...