

# PASA PDWG

Produced in partnership with:

**itm**  
Solution delivered...

PASA Experts for Pensions

## Addendum: Data Matching Convention (DMC) Guidance

March 2023

## Additional Guidance on Data Matching

PASA's Data Matching Convention (DMC) Guidance was updated in August 2022 to list four areas where additional core guidance was required:

1. Guidance on matching without a National Insurance Number (NINO)
2. Guidance on Possible Match responses
3. Considerations for 'Split Administration' scenarios
4. Impact of whether personal identifiers are verified or self-asserted

In March 2023, we're releasing Guidance on the first three of these topics. Topics 1 and 2 are covered in this Guidance, while Topic 3 is covered in a separate Guidance discussing the wider issues around pensions dashboards compliance for split administration scenarios. Topic 4 will be covered later this year.

### Notes on terminology:

We use the term 'scheme' to refer to the entity which needs to comply with dashboards regulations, including making decisions on appropriate matching criteria, and we refer to 'trustees or scheme managers' as the decision makers. However, this Guidance is also aimed at FCA regulated pension providers, so our use of these terms should be interpreted in this context where applicable.

We use the term 'ISP' to refer to the technology processing the comparison of Find request data against the personal details held by the scheme. In practice, this technology may be provided by an ISP (Integrated Service Provider) or by other means, such as a scheme building its own technology solution.

## Additional Guidance 1 - Matching without a National Insurance Number (NINO)

### Why do schemes require some match criteria which don't rely on NINO?

The term 'match criteria' describes a set of data elements schemes use to compare against dashboards Find requests, for example, *Last Name/Date of Birth/NINO*. Schemes may define a number of different match criteria and for each one, they'll decide whether a positive result will trigger either a Match Made response or only a Possible Match response.

As NINO is the closest option the UK has to a unique identification number and is also usually held by pension schemes in their member records, there's a broad consensus it should feature in at least some of a scheme's match criteria.

It's expected NINO will usually be included as a data item in dashboards Find requests, and will also usually be available in scheme records - even more so in the case of providers of contract-based pensions. However, this may not always be the case and it's therefore important for schemes to consider matching criteria without having NINO as one of the data items - to maximise the opportunities to match savers to their pensions.

Scenarios, where a NINO may not be available, include:

- A. A user who doesn't provide a NINO when searching for their pensions as it's not a mandatory field within the Pensions Dashboards Programme's (PDP) data standards
- B. A user who has a UK pension scheme entitlement but isn't authorised to have a NINO. This will include some users who reside outside the UK
- C. A scheme which doesn't hold a NINO for every saver. This may include individuals who didn't have a valid NINO allocated to them at the time their employer enrolled them into the pension scheme. Also, there are historic reasons why groups of members, many of whom aren't aware of their scheme membership, may have very sparse personal data held without a NINO, such as Equivalent Pension Benefit (EPB) only members
- D. A scheme which holds NINOs for their savers but may not trust their accuracy (determining the accuracy of NINO data is very difficult due to the limitations of verification options)
- E. A NINO sent in a Find request from dashboards will have been typed by the user (self-asserted) and not verified by the digital identity service, which leaves the NINO prone to error (for example, a mistype by the user). It's also worth noting a NINO could in theory be deliberately entered as someone else's known NINO, which is why a NINO match on its own can never guarantee a response should be a Match Made.

Some of the scenarios relating to mistyped or mis-held NINOs could be addressed using 'fuzzy' comparisons, as covered in our Additional Guidance on Possible Matching below. But most of the scenarios above can only be addressed by adding additional match criteria which don't include NINO as a data item at all.

## Choosing additional sets of match criteria which don't include NINO

We suggest **devising multiple additional match criteria** which don't include NINO at all, even if trustees and scheme managers only want to use these additional match criteria for Possible Match responses. The use of multiple criteria will mean if one set of match criteria which doesn't include NINO fails to return a match, then there are further criteria which may be successful.

Key examples of match criteria which don't include NINO, which should be considered for either a Match Made or Possible Match response, are as follows:

Evaluation of Match Criteria					
Option*	Match criteria	Individual savers' experience with this option	ISP technical operation of this option	Admin implications of this option	Risk of using the option for a Match Made response
NN1	Last name Date of birth First name Current Postcode	Pensions are found for individuals whose current address and last name are up to date in the scheme's records	Simple match regime to implement and operate	Works best with frequent address tracing	Risk of false positive is minimal when considered statistically, and these data items are all digitally verified
NN2	Last name Date of birth Email address	Pensions are found for the individual where email address or mobile number are entered and known by the scheme		Requires good coverage of quality email addresses and/or mobile numbers	Email addresses and mobile numbers may not be digitally verified
NN3	Last name Date of birth Mobile number				
NN4/5	Add First name as additional data item to NN2 and NN3				
NN6	First name Last name Date of birth	Pensions are found for individuals irrespective of whether their current address is up to date in the scheme's records, and irrespective of any self-asserted data (typed by the user without digital verification) being provided		Would only (optionally) require tracing of last names, which change much less frequently than addresses.	Risk of false positive is material when considered statistically, so only suitable for a Possible Match response
NN7-12	<b>Any of the criteria above</b> but allowing matches on alternate name or additional postcodes <b>**</b>	As for the equivalent criteria above		As for the equivalent criteria above	This weakens the strength of the match as alternate names and additional postcodes are self-asserted

\*NN = No NINO

\*\* In the November 2022 update of the Data Standards, it's been confirmed Alternate Forenames will also be provided as part of a Find Request. This could also be incorporated, although initial analysis suggests schemes may be less likely to use this.

While email address and mobile number may seem unusual candidates for matching, they share common properties with NINO, such as being unique identifiers and self-asserted. They have an added benefit for matching as, unlike NINO, email addresses and mobile numbers can be verified by the scheme. Although initially email addresses and mobile numbers won't be verified by the Digital Identity Service, the PDP's data standards are set to evolve and may allow verification to be performed and validated for these data items at a later date.

### Why is only the Postcode part of the address being recommended for matching?

Current address as a whole (including Postcode) is verified by the Digital Identity Service so, trustees and scheme managers can be assured this data element will be correct. The PDP's Data Standards define the following fields for Address:

- Address Line 1
- Address Line 2
- Address Line 3
- Address Line 4
- Address Line 5
- Postcode

However, many of the address lines are still prone to variance and incorrect locations within the data fields of the administration platform. For example, one address can easily be represented in multiple ways:

	Example 1	Example 2	Example 3
Address line 1	One Apple Street	One	1 Apple St.
Address line 2		Apple Street	
Address line 3	Blackburn		
Address line 4		Blackburn	Blackburn
Address line 5			Lancashire
Postcode	BB2 1WB	BB2 1WB	BB2 1WB

While comparisons of addresses of this nature can easily be compared by human beings, getting a machine to decide if 'Lancashire' was an important detail, either requires complex technical capabilities or risks unpredictable outcomes.

Postcode provides a more reliable means of matching as the format is restricted. However other address fields can still be helpful for administrators manually resolving a Possible Match, and also validating if the postcode held is consistent with the address, and the address itself is a valid address.

### There isn't one rule for all, and rules need to be kept under review

The above recommendations provide suitable matching options for many schemes however, they may not be successful for all. For example, some schemes may have unpopulated data fields, or it may have been some time since the scheme last completed a tracing exercise. By assessing the recency, accuracy, completeness and uniqueness of data items within their scheme data, schemes can then select the most effective criteria to match.

Once initial match criteria are defined, the next big test will come when dashboards are in mass public use, either during the expected live testing phase or at the point of full public availability (the 'Dashboards Available Point'). If savers are still struggling to match with scheme data, or if a scheme is generating too many incorrect Possible Match responses, and these are having an operational impact on scheme administrators attempting to resolve them, then a scheme may need to refine their set of match criteria which don't include NINO. (See Section 2 of additional Guidance on Possible Matching below for further discussion on planning for operational impacts).

Criteria may need to be changed after mass public use as we will learn a lot through this 'live testing', including a better understanding of user behaviour in supplying self-asserted data fields, such as NINO. There will also likely be further adaptations in the dashboards' user journey and wider industry improvements which will necessitate keeping matching criteria under review.

As mentioned above, PDP's Data Standards may allow for further data items to be verified by the Identity Verification Provider in the future. Schemes, understandably, will be more wary of matches made which involve unverified data items. However, when a data item becomes digitally-verified schemes may change their match criteria to place more reliance on that data item, or decide to upgrade their response for match criteria which uses that data item to be a Match Made instead of a Possible Match. This will be covered in more detail at a later date in future Guidance.

## Additional guidance 2 – Possible Match responses

### What is a Possible Match response?

A Possible Match is defined in the Dashboards Regulations as:

*“a positive match where the elements of personal data provided by an individual for the purposes of matching only partially meet the matching criteria [set by trustees or scheme managers] such that the trustees or managers of the scheme are unable to determine (in the absence of further information) whether or not the individual is a member of the scheme”*

The response to a View request for a case flagged as a Possible Match provides only limited information, as defined in the PDP Data Standards – we refer to this as the ‘Possible Match response’. This includes the name of the pension arrangement, contact details for the pension administrator, and a pension reference that the individual can provide if they choose to make contact with the administrator.

The purpose of a Possible Match response is to enable a positive user experience in the event a confident, but not definite, match is made between a Find request and scheme records. It’s important to note supporting Possible Match responses where needed is a requirement of the legislation, and is clearly set out in [TPR’s guidance on Pensions Dashboards](#).

The legal responsibility for supporting Possible Match responses rests with trustees and scheme managers, even though it’s their administrators and other parties who will be doing the work. Trustees and scheme managers need to be comfortable any approach being taken will meet their obligations. Even where trustees and scheme managers believe the personal data they hold is already at a very high standard and will be maintained as such, support for Possible Match responses is still an essential safety net for dashboard user experience. Not least because of the uncertainties about how savers will volunteer ‘self-asserted data’.

Options for matching should cover this scenario and make recommendations which manage risk for trustees and scheme managers. This should strike the right balance between reconnecting savers to their pensions and complying with legislation, while avoiding large volumes of calls to administrators if too many ‘false matches’ are returned.

The other benefit of using Possible Match responses is pensions data quality should improve in the future, by enabling schemes to maintain engagement with members (or at least those members who want to be engaged because they’re using dashboards). For example, a former member who has changed name and address since their time in the scheme should be able to be reconnected through Possible Matching.

This opportunity for communication between the scheme and a previously ‘disengaged’ member can be used to clarify information through the scheme administrator’s verification processes. This doesn’t mean all relevant

information will end up being updated to match the data in the Find request, but it should result in improvements in the scheme's records.

### What makes a good set of match criteria for a Possible Match response?

We've defined the term 'match criteria' to describe a set of data elements schemes use to compare against dashboards Find requests, for example, *Last Name/ Date of Birth/ NINO*. Schemes can also define their own sets of match criteria to trigger a response (either Match Made or Possible Match), making their matching assessment as comprehensive as possible.

The effectiveness of the match criteria a scheme chooses to use for **Possible Match responses** will ultimately be judged against two considerations:

- A. Sufficient coverage:** Do the set of match criteria, in combination, pick up a high enough percentage of scheme members who have failed to be picked up as a Match Made, due to discrepancies in data between the Find request and the scheme's records?
- B. Sufficient focus:** Is each match criteria, in isolation, sufficiently tightly defined so they won't incorrectly provide responses for too many dashboards users, and hence invite individuals to take part in a wasted journey to contact the scheme's administrators?

If the sets of match criteria used for Possible Matching don't have **sufficient focus**, then there will be significant impacts on both users' dashboards experience and the schemes' operational costs in the wasted effort spent to resolve Possible Matches.

### Using standard match criteria for Possible Matching responses

The default way of carrying out the comparison when applying a set of match criteria is to compare each data element provided in the Find request with the equivalent data element held in the scheme's records and seek an exact match, possibly with an allowance for basic discrepancies such as blank spaces in the text, or accented characters held in names.

Match criteria which seek an exact match when comparing fields can certainly be used for Possible Match responses, and some examples were given in the previous section on matching without NINO. However, if schemes only use match criteria which seek an exact match, they may find it difficult to meet both considerations set out above of sufficient coverage and sufficient focus for effective Possible Match criteria, as shown in the examples in the table below.



### Examples of 'exact match' criteria for Possible Match responses:

Example	Match criteria	Comments on effectiveness for Possible Match responses
1	Date of birth NINO	This set of match criteria would have sufficient focus to be included as one of the Possible Match criteria, but on its own wouldn't have sufficient coverage.
2	Last name NINO	These match criteria would undoubtedly improve the coverage of the Possible Match criteria, but <i>Example 4</i> in particular provides too little focus and could over-invite individuals incorrectly, particularly for scheme members with more common last names.
3	NINO First name	
4	<i>Last name</i> <i>Date of birth</i>	

This table shows match criteria which operate only on an 'exact match' being made between the data elements may not meet the criteria (sufficient coverage and sufficient focus) for effective Possible Matching.

### Using match criteria which include 'fuzzy' comparisons

The other option is for schemes to add matching criteria where some of the comparisons of data elements are done on a 'fuzzy' basis. This will allow the match criteria to pick up on differences between the Find request and the scheme records which are clearly due to an error such as a 'typo'. For example, a mistyped NINO either by the user or in the scheme records.

The exact nature of the 'fuzzy' comparisons available to schemes will depend on the options made available by their ISP. For example, a minor typo in a scheme's record of a member's Last Name could be identified using word comparison techniques.

Match criteria using 'fuzzy' matching can be a very focused way of making a Possible Match response. For example, if NINO and Date of Birth match, but Last Name has a small difference (say 'Smith' in the Find request, but 'Smtih' on the scheme records) then it's almost certain this is the scheme member, but a Possible Match resolution would allow the scheme to confirm the match and update their records. Alternatively in this situation, some schemes may even be happy to treat this as a Match Made response instead.

This approach of 'Fuzzy' matching can be applied across different personal identifiers, each of which will require different techniques to check if two data element values are similar. Below is an example of how a 'fuzzy' match can be applied to the NINO data element using a word comparison technique to check the number of characters which are different:

Match criteria set: Last Name, Date of Birth, NINO.

Result: Last Name and Date of Birth match exactly.

Scheme data NINO: AB123456C

Saver	Saver Input	Match type	Recommended application
1	AB123456C	Exact match	Match Made
2	AB123453C	Fuzzy match: 1 character out	Possible Match
3	AB123465C	Fuzzy match: 2 characters transposed	Possible Match
4	AB123573C	Fuzzy match: 3 characters out	No Match

Each data field should have algorithms specifically designed for the context of that data field. For example, Date of birth may benefit from different algorithms (or 'Fuzzy' match types) due to it being a strict number format and the data field having three separate sections (DD, MM, YYYY).

Match criteria set: Last Name, Date of Birth, NINO.

Result: Last Name and NINO match exactly.

Scheme data Date of birth: 02/11/1972

Saver	Saver Input	Match type	Recommended application
1	02/11/1972	Exact match	Match
2	03/11/1972	Fuzzy match: 1 day out	Possible Match
3	11/02/1972	Fuzzy match: DD and MM transposed	Possible Match
4	17/11/1972	No match: > X days out	No match

Schemes should analyse their current data to understand which algorithms will be most logical to apply to their matching criteria and speak to their administrator or ISP to ensure their needs are supported.

## Examples of match criteria that use fuzzy comparisons

Examples of match criteria which incorporate fuzzy comparisons of at least one of the data elements are shown in the following table:

Evaluation of some Match Criteria that include 'fuzzy' matching					
Option*	Match criteria	Individual's experience of this option	ISP technical operation of this option	Admin implications of this option	Risk of using this option
FM1	Last name Date of birth <i>Fuzzy match</i> -NINO	Whether these criteria are used for Possible Match, or some of them even for Match Made responses, the individual's experience is better than the potential alternative of a 'No match'.	This is a complex operation which may be harder to achieve in the required service level response times when dashboards are used at scale, for example, it may require a higher level of underlying processing power.	If these are used as Possible Match responses then resolving them will add an administrative burden, but if the alternative is a 'No match' then this could lead to a greater administrative burden through enquiries and even complaints over time.	Low risk of too many false matches if used for Possible Match responses as the criteria are all sufficiently focused, but care needs to be taken if deciding to use any of them as Match Made responses.
FM2	<i>Fuzzy match</i> -First name Last name NINO				
FM3	First name <i>Fuzzy match</i> -Last name NINO				
FM4	<i>Fuzzy match</i> -First name <i>Fuzzy match</i> -Last name Date of birth NINO				
FM5	<i>Fuzzy match</i> -Last name Date of birth NINO				
FM6	Last name <i>Fuzzy match</i> -Date of birth NINO				

\*FM = 'fuzzy' Match

## Guidance on choosing Possible Match criteria

Industry research has suggested it's possible to achieve a 99% + find rate (coverage) when using multiple match criteria for both Match Made and Possible Match responses. Successful Possible Matching, which achieves a high find rate, but is also focused enough to ensure a low rate of incorrect matches, requires a combination of match criteria and techniques.

From the analysis in this Guidance, the types of match criteria which can be chosen for Possible Match responses fall into four main categories:

Type of match criteria	Example match criteria	Suggested confidence in the match	Reason to consider a Possible Match response
Exact match comparisons which don't give enough confidence to respond with a Match Made	Last name Forename Date of birth Current Postcode	Close to 100% confidence in the match	If not deemed certain enough to be a Match Made then should be suitable for a Possible Match response
Exact match comparisons, which also imply another key data item held (like Date of birth) may be incorrect	Last name Forename NINO	Full confidence in the match	If members match on this criteria but <u>don't</u> match on <i>NINO, Last name and Date of birth</i> , this implies Date of birth is different and hence a Possible Match response will help to resolve
Match criteria which include a fuzzy field comparison	Last name Forename Date of birth <i>Fuzzy match-NINO</i>	Full confidence in the match	A Possible Match response will enable the data discrepancy to be resolved. However, it may be decided to treat this as a Match Made response instead, to reduce operational overhead, and/or arguably improve the saver experience, if the view is the discrepancy was most likely as a result of an error in the saver's self-asserted data
Match criteria which include multiple fuzzy field comparisons	<i>Fuzzy match-First name</i> <i>Fuzzy match-Last name</i> Date of birth Current Postcode	Close to 100% confidence in the match	Not likely to be deemed certain enough to be a Match Made by most but should be suitable for a Possible Match response

Looking back again at how we judge the effectiveness of a set of match criteria for Possible Matching, we described previously the following two considerations:

- A. Sufficient coverage:** Does the set of match criteria, in combination, pick up a high enough percentage of scheme members who have failed to be picked up as a Match Made, due to discrepancies in data between the Find request and the scheme's records?
- B. Sufficient focus:** Is each match criteria, in isolation, sufficiently tightly defined so that they will not incorrectly provide responses for too many dashboards users, and hence invite individuals to take part in a wasted journey to contact the scheme's administrators?

Having considered the benefit of fuzzy comparisons, we would add a third consideration:

- C. Technical feasibility:** Are their chosen match criteria for Possible Match responses, including the possible use of fuzzy comparisons, supported by their chosen ISP?

For trustees and scheme managers to fully assess these considerations, we recommend undertaking an exercise to analyse the quality and depth of the scheme's records of personal details, and discussing the technical options for matching which will be supported by their chosen ISP. There may also be a need to assess the risk appetite of the scheme's data controller (usually, also the trustees or scheme manager), particularly if some of the match criteria may be promoted to be used as criteria for Match Made responses.

### Planning for operational overhead

There's no escaping the fact Possible Matching will cause an operational overhead for administrators, as they're likely to be contacted by savers for each Possible Match returned. The key question is how can schemes estimate the scale of what their overhead will be, and when the demand will start.

With regards to the 'when', there are two relevant future points when the use of dashboards is expected to commence at scale. The first is through live testing, the timing and scale of which are to be defined and agreed upon, but it is expected to build up gradually.

The second relevant point on timing is the setting of the DAP itself. The DWP has released a paper which sets out proposed criteria the Secretary of State will take into account when setting a date for the DAP, giving the industry 6 months' notice.

When it comes to the expected volume of dashboard use in general, PDP released a report: [Willingness to Pay analysis](#), which estimated 56% of UK adults will likely use dashboards between one to two times a year. Schemes can use this report to begin estimating how many of their scheme members will use pensions dashboards.

The other thing to estimate is how many scheme members will need Possible Match responses, based on the quality and depth of the personal details data held, and the set of match criteria proposed to be used for Possible Match responses. This is analysis a scheme's administrator or ISP should be able to carry out.

Having carried out this analysis, the results can then be used to implement a data improvement exercise. This will have the effect of reducing the expected number of members who would require Possible Matching. No amount of one-off data improvement will reduce this number to zero, but if you can halve your expected population requiring Possible Match responses then this is hugely beneficial for capacity planning.

For further information on carrying out data improvement ahead of pensions dashboards please review the [PASA Guidance](#).

Using all the above information, administrators can have a target in mind when allowing for the capacity to resolve Possible Match requests. Once mass public use of dashboards commences then scheme administrators may need

to adjust their Possible Match criteria and administrative capacity, based on the performance of their matching approach.

We expect it will be beneficial for schemes to adapt their matching criteria as they analyse their performance, their scheme data improves, and we learn more about how users interact with dashboards, and the data they choose to ‘self-assert’ to help them find their pensions.

### Directing savers in the event of a Possible Match

Possible Matching benefits user experience by reducing the likelihood of a user receiving no match and in turn no information. However, to ensure the user has an improved experience through receiving a Possible Match, the user journey from this point will need to be planned effectively.

Administrators will need to decide, with the agreement of the trustees or scheme managers:

- How they will receive communication from a user in receipt of a Possible Match
- What information do they require from the user to resolve a Possible Match

The dashboards Data Standards allow schemes to provide the following back to the saver in receipt of a Possible Match, which will influence how they get in contact:

Field	Description
Administrator URL	URL of the pension administrator, to allow the individual to access an administrator website.
Administrator email	Email address to contact for further information.
Administrator phone number	Full telephone number to allow the individual to contact the administrator/provider via telephone.
Administrator contact preference	Method of contact preference of the administrator.

In addition, the November 2022 version of the Data Standards has confirmed optionally schemes can also return a value for a **pension reference** field which can be quoted by the saver in any contact they make with the administrator. This may be best implemented using a case number for the specific Possible Match resolution, rather than an actual reference number in use for the scheme member concerned – who may or may not prove to be the same individual as the dashboards user.

### Possible Match resolution process

Where savers decide to follow up the Possible Match response with the administrator, then a new process will be required to handle and resolve this. It should be noted there are time limits for savers to respond to notification of a Possible Match, and ISPs and administrators have rules to follow around when Possible Match responses expire

for those savers who don't respond – this is covered in dashboards Regulations and Standards, and we won't cover this aspect in this Guidance.

Where savers initiate the Possible Match resolution process, administrators should consider technical solutions which reduce the amount of time spent manually interacting with dashboard users, such as in telephone calls.

The challenge for administrators is the Possible Match resolution process will have to start with an identity verification step, as the user who is contacting them may or may not be a member of the scheme. It's worth noting different Possible Match responses will have different levels of certainty over the member's identity as we've seen in the table above, which may impact the options available for the identity verification process.

Depending on where a user in receipt of a Possible Match fails the match criteria, a scheme should know the specific information required to satisfy a match. We recommend schemes review their current identity verification (IDV) processes with a view to handling dashboards' Possible Match resolutions. This should allow schemes to streamline their IDV, reducing the amount of time involved for both user and scheme to resolve a Possible Match.

We won't cover the technicalities of the options ISPs will provide at the end of the Possible Match resolution process in this Guidance, but trustees and scheme managers can obtain more detail on these from their chosen ISP or equivalent. As noted earlier, where the Possible Match resolution process concludes the saver is indeed the scheme member, then this may also enable the administrator to make improvements or corrections to the personal details they hold for the member.

One of the evolving areas is the options to be offered in relation to savers who are concluded to be a 'No Match' to potentially suppress a subsequent repeat of the Possible Match response. We expect to add some additional guidance in this respect in the next version of the overall DMC Guidance.

Get in touch:

[info@pasa-uk.com](mailto:info@pasa-uk.com)

[www.pasa-uk.com](http://www.pasa-uk.com)