



Department
for Work &
Pensions

UNCLASSIFIED

**Information, Governance & Security Directorate
Information Services for Delivery**

**INDIVIDUAL ELECTORAL REGISTRATION - CONFIRMATION
DWP DATA MATCHING METHODOLOGY**

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

- 1 -

0.0 Document Control Information

Author	Neil Taylor
Document Version Number	V1.0
Document Reference	IER - DWP Data Matching Methodology

0.1 Document Amendment History

Version	Date	Reason for Update
V1.0		Baselined for external publication

0.2 Glossary of Abbreviations

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
CCDM	Customer Centric Data Mart
CESG	Communications-Electronics Security Group
CIS	Customer Information System
CIT	Corporate Information Technology
CO	Cabinet Office
DSDNI	Department for Social Development in Northern Ireland
DWH	Data Warehouse
DWP	Department for Work and Pensions
EMS	Electoral Management System
ERO	Electoral Registration Officer
ERTP	Electoral Registration Transformation Programme
GB	Great Britain
GFTS	Generic File Transfer Service
HMG	Her Majesty's Government
HPES	Hewlett Packard Enterprise Systems
IER	Individual Electoral Registration
IGSD	Information, Governance & Security Directorate
ITG	Integration Back Bone
NINO	National Insurance Number
OS	Ordnance Survey
RAG	Red, Amber, Green
SI	Statutory Instrument
UPRN	Unique Property Reference Number
UTF	Universal Character Set Transformation Format

Contents

1.	Background	4
2.	Document Purpose	4
3.	High Level Approach	4
4.	Data Sources & limitations	5
5.	Initial File Receipt and Postcode Allocation	8
6.	Validation, Standardisation & Cleansing	8
7.	The six stages of data matching	11
8.	Fuzzy Matching	12
9.	Multiple Matches	13
	Appendix 1 – Output File/Questions	14

1. Background

- 1.1 The Political Parties and Elections Act 2009 (the 'PPE Act') made provision for reforming electoral registration. This reform included a framework for moving to Individual Registration (IER) in order to modernise the electoral registration and tackle electoral reform.
- 1.2 As part of the transition to the new system, electoral registration officers participated in a series of data matching pilot exercises from August 2012 through to the 31st March 2013, to test the potential value and accuracy of matching entries on the electoral register against DWP data in order to confirm the identity and residence of an individual, thereby enabling them to be automatically transferred to the new IER register without the need to provide additional personal identifiers.
- 1.3 These pilots built on a previous exercise, carried out in late 2011, which informed the ability of using DWP data and the potential matching algorithms available.
- 1.4 A full evaluation of the pilot exercise was carried out by both the Cabinet office and the Electoral Commission, which recommended that confirmation utilising data matching with DWP as a process was an exercise which added significant value to the transition to the IER system. The Cabinet Office evaluation can be found here: <https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration>, and the Electoral Commission one here: <http://www.electoralcommission.org.uk/voter-registration/individual-electoral-registration>.
- 1.5 Following on from this, a confirmation dry run exercise was carried out through July and August 2013, which will test all of the systems required to carry out a confirmation exercise but allowed all electoral registers to be passed through the IER system and the DWP data matching, in order to evaluate how the confirmation exercise translated across a national scale.
- 1.6 This exercise improved upon the results from the pilot exercise, in that an average of 78% of records nationally across the electoral roll could be confirmed through data matching to administrative data.

2. Document Purpose

- 2.1 The purpose of this document is to detail DWP's data matching methodology which was used for the purposes of the confirmation dry run, which will be carried forward to the live confirmation exercise in 2014.

3. High Level Approach

- 3.1 DWP's Information, Governance and Security Directorate have traditionally delivered matching products for a range of internal customers and other government departments, many with the intent to confirm the validity and accuracy of relevant personal details.

- 3.2 During the 2nd pilot exercise in 2012/13, it was agreed that the high level aim was to ingest the set of electoral data for participating Local Authorities, and attempt to match the electoral data against a set of DWP data derived from the its Customer Information System.
- 3.3 The Customer Information System (CIS) is a system used by the Department of Work and Pensions (DWP) to store basic identifying information about citizens who have been allocated a National Insurance Number or children who have been allocated a child benefit reference number.
- 3.4 The end result of the match between the electoral register and DWP's data enabled a series of questions regarding the validity of a person's identity and their address details to be answered and for that information to be returned to the IER system.
- 3.5 This in turn would enable the assignment of a RAG rating to each entry on the IER system, allowing the Electoral Registration Officers to determine the accuracy of their information in relation to DWP held data.
- 3.6 The definition of the RAG rating is defined, owned and run within the IER system, and DWP do not have a requirement to define the level to which the RAG rating should be set, only to provide information to inform such a rating.

4. Data Sources & limitations

- 4.1 The confirmation exercise uses 3 sources of data during the matching process. These being the DWP Customer Information System, the individual electoral register data, and the Ordnance Survey AddressBase data.

CIS

- 4.2 CIS is the master of customer information across DWP and interacts with a broad range of its benefit systems, as well as selected systems such as those held by HMRC to maintain an aligned picture of a person's information.
- 4.3 As well as identifying information such as name, address, date of birth, National Insurance Number, etc., CIS also keeps a limited record of benefits that an individual may have claimed over the last two to three years. It also retains deceased records indefinitely for all deaths that are notified to the system.
- 4.4 The environment the matching exercise takes place in, is not attached directly to the Master CIS system, relying instead on a routine extraction of data from it which is brought into the IGS Data Warehouse on a daily basis and then made available to the matching environment on a weekly basis each Friday evening.

- 4.5 As a result the CIS data used in the match carries with it a small amount of latency between its picture of a customer and that held by the master CIS system. The impact of this for example is that those electoral rolls which are matched on a Saturday will match against the most up to date set of CIS data available to the matching environment, whereas those matched on a Thursday are matching against the most latent data.
- 4.6 The impact of this is minimal in that only in scenarios where an individual or system notifies a change of address or name between the weekly update points, would the matching environment reflect a different view of information from that held on the master system, re-aligning itself at each Friday update.
- 4.7 CIS as a master system is dependant on the accuracy of its information regarding a person, on the systems that feed it, and more importantly the interactions a person has with those systems in informing them of changes to said details.
- 4.8 There are a number of demographic factors, which play on the likelihood and timeliness of a person interacting with a dependent system to the master CIS system, and thus can affect the accuracy of the information held and used to match against the electoral role.
- 4.9 Factors such as migration, education, and stability of employment, amongst others, may lead to a lower level of interaction between a person and the systems in question. For example, a student in higher education living away from home is likely to be associated to his or her parent's address on the CIS system, as they will during the period of education be less likely to interact with DWP benefit systems, and therefore notify a change of address. Whereas the electoral role may be more likely to reflect their educational address.
- 4.10 It is critical to understand that DWP's system are primarily aimed at the delivery of benefit, and therefore those people who interact with the department on a regular basis are most likely to reflect an accurate picture of their personal data on its systems.
- 4.11 The extract of CIS utilised in the matching exercise does not include records which DWP classify as restricted. However, the number of these records is statistically insignificant, and therefore although this will necessarily mean electoral records for these customers can never be matched, they were deemed not to affect the potential match rate to any significant degree.

Electoral Registers

- 4.12 The 380 electoral registers are supplied from Local Authorities via a small set of electoral management systems, to the IER system in a standardised structure on a scheduled basis.

- 4.13 The IER system does not attempt to cleanse the data before passing to DWP, but does ensure via validation processes that the records conform to a set structure.
- 4.14 The register records within the IER system are allocated unique keys, and are batched up into nightly files, representing the registers uploaded to the IER system on any given day. The structure of this file is detailed in appendix 1.
- 4.15 It should be noted that the electoral register is primarily an address based product, i.e. people are associated to an address whereas the DWP CIS is a customer based data source and therefore an address is recorded in association to a person.
- 4.16 As such the accuracy of the address content is fundamental at the lowest level of granularity to the electoral register. In line with this the majority of electoral registers provided a Unique Property Reference Number (UPRN) alongside the standard address detail in the set of information provided to the matching exercise.
- 4.17 With respect to personal details, the electoral registers either do not currently hold certain pieces of information, such as National Insurance Number or only hold information for certain sets of people, i.e. dates of births are only held for attainer's.
- 4.18 They are also dependent, like the DWP CIS system, on the content of a persons name being that which is provided by the customer in their interaction with the relevant processes, and therefore the accuracy and completeness of that information can vary across electoral rolls.
- 4.19 The latency of the data held on the electoral register can also be a factor in the ability to successfully resolve a match. The gap between the point at which information is received to update the electoral register and the point at which that set of data is uploaded to the IER system and then sent to DWP for matching will mean that the larger the gap the more potential change will occur to a CIS record.

Address Base

- 4.20 The third source of data used in the matching exercise is an extract from the Ordnance Survey AddressBase plus product.
- 4.21 AddressBase comprises 3 sets of data from GeoPlace and the Ordnance Survey. These 3 files contain information from Royal Mail, PAF, NLPG and OS AddressLayer2 with further information from the Valuation Office. Matched, cleansed and enhanced by GeoPlace, the AddressBase product offers a record of all current, historic and not yet built properties in England & Wales.

- 4.22 AddressBase is made up of Royal Mail PAF data matched to UPRN assigned by local authority custodians to all addresses (even those still at the planning stage).
- 4.23 AddressBase Plus takes this core dataset and enhances it further by using address data from the NLPG as the core address record. In addition to postal addresses, this file contains non-postal locations such as subdivided properties, places of worship and community centres.
- 4.24 The provision of this product in to the matching environment allowed DWP to assign the UPRN provided by the product to its set of claimant addresses allowing an alignment to the provision of the UPRN on the electoral roll.
- 4.25 However, it should be noted that in order for DWP to assign a UPRN to a CIS record, it firstly has to match the address detail to that held by the AddressBase product, which is held in a different structure to that on the DWP CIS system.
- 4.26 As a result, before a match takes place between the electoral register and CIS an external match occurs which is itself dependent on the ability to resolve an address match using data which can vary in quality, and therefore can impact the assignment of a UPRN value to a CIS record.

Data Matching Methodology used within Confirmation Dry run

The data matching methodology and process that DWP carries out to fulfil the purpose of the confirmation exercise is carried out over a number of distinct stages. These are as follows:

5. Initial File Receipt and Postcode Allocation

- 5.1 On receipt of the daily file from the IER system, DWP firstly check the structure of the file to ensure that no corruption has occurred during transfer.
- 5.2 The next step of the process retrieves a distinct list of postcodes from the provided file, and uses that list to reduce the volume of CIS records to a more manageable processing size.
- 5.3 This has the effect of excluding any CIS records from the matching exercise which do not contain a postcode in their address. The majority of these records do not contain a postcode because they relate to a foreign address, however, there can be instances where a British address held on CIS does not hold a postcode, and in this instance the person attached to that address would not be available for matching.
- 5.4 The postcode filtered CIS data is reduced further by excluding records where the age of the person is less than 16 years old.

6. Validation, Standardisation & Cleansing

- 6.1 The next stage of the process ensures that in matching electoral register records to the CIS data, the data and its structure are as aligned as possible, therefore the process firstly validates, standardises & cleanses both data sources to the same level.
- 6.2 The following is a list of the standardisation, cleansing & validation contained within the matching algorithm.

Variable/Subject Area	Type	Definition
ERO_ID	Validation	<p>During the pilot phase and the confirmation dry run, it was discovered that the unique ID attached to the electoral roll was not in fact unique, but could be present when an exact duplicate electoral record was present in the IER system.</p> <p>Therefore DWP checks the uniqueness of the records and removes duplicates when they are found.</p>
Mandatory Fields	Validation	<p>The key fields which are used in the matching algorithm - Surname, Address Line 1 & Postcode are checked for population.</p> <p>Those records which do not contain a value in these fields are rejected as a failed match.</p>
All Fields	Standardisation	To ensure that there is no impact from the case of the value held within text string matching, all text fields are set uppercase.
NAME & ADDRESS LINES	Standardisation	As various punctuation characters can be present in both the electoral roll and CIS data, all such characters are removed from all relevant text fields.
POSTCODE	Standardisation	<p>Although postcodes already have a set format, user inputted values can sometimes mean these formats are not adhered to.</p> <p>Therefore postcodes on both datasets are standardised to a 7 length postcode – I.e. S1 1AU becomes S1 1AU</p>
ADDRESS LINES 1-4	Cleansing	Administrative data systems are reliant on user input of addresses and as such they vary in the value

		<p>they can hold for common strings.</p> <p>Therefore these common strings are formatted to the following standard:</p> <p>STREET=ST AVENUE=AVE ROAD=RD COURT=CT PARK=PK CLOSE=CL GARDENS=GDNS GROVE=GR AVE=AV DRIVE=DR CRESCENT=CRES LANE=LA TERRACE=TERR PLACE=PL ST.=ST WAY=WY SQUARE=SQ</p>
ADDRESS LINES 1-4	Cleansing	<p>The location within a text string of the numeric element of an address can vary from system to system, and any string to string matching can be adversely affected by its location.</p> <p>Therefore the solution parses the numeric element of the address and holds it separately for both sources.</p>
FIRSTNAME/LASTNAME	Cleansing	<p>Variations in the spellings of common names can lead to mismatches during the matching exercise, therefore for this matching solution the variants in the name 'Mohammed' have been standardised to a single spelling.</p>
MIDDLENAME	Standardisation	<p>Due to the variants in either full length middle names or initial only, an additional field is created with a standardised middle name initial which incorporates the need to ensure that multiple middle names when displayed as initials should be for example 'J S' for John Smith rather than 'J' only</p>
NAME & ADDRESS LINES	Standardisation	<p>A concatenated string for both NAME and ADDRESS is created. In the case of NAME, 2 are created to include the variant length of the</p>

7. The six stages of data matching

- 7.1 The matching methodology is based on a six stage approach to matching per electoral roll following the standardisation and cleansing routines.
- 7.2 Each of the six stages fall into the 2 categories of data matching used in the solution: Address Blocking and Identity Matching.
- 7.3 **Address Blocking** – Stages 1, 3 & 5 detailed below, fall into this category. This blocking effectively creates a Cartesian product between the input source and the matching source. This is done in a number of passes in order to cope with the variant data quality of address information on both the input source and matching source.
- 7.4 So for example if you have a record on the input source with a UPRN value of 1, and on the matching source there are 2 records with a UPRN of 1, then a Cartesian product would occur on the output data. In this instance 2 records would be output. If there were 2 records on the input source with a UPRN value of 1, and 2 records on the matching source, then 4 records would be output.
- 7.5 **Identity Matching** – Following each address blocking stage, those records successfully joined together, are carried into an identity match stage.
- 7.6 The identity match takes both the initial personal details variables, as well as the derived ones, such as initials, soundex values, etc. and creates a series of 0/1 flags based on an input source variable and a matching source variable. I.e. SURNAME = SURNAME.
- 7.7 These flags are then used in combination in order to provide an answer to the agreed series of questions.
- 7.8 Six stage process detail:
- Stage 1 – UPRN address block – Those records within the register which have a valid UPRN, will be linked to the DWP CIS data where applicable via the relevant UPRN. This will be carried out using the principle of creating a Cartesian product as detailed above. Records successfully joined at this stage will be passed to stage 2, with those that are unsuccessful passed to stage 3.
 - Stage 2 – UPRN based Identity Match – The agreed questions will be passed against the Cartesian product, and those records that pass uniquely with an identified strong match question, will be output to a successful match pot. Those that fail the identity match, or those that did not have a strong enough match will be failed to pass through the next stage.

- Stage 3 – Address Detail block – Those records within the register which were not successfully passed through stage 2 or not joined during the stage 1 address blocking phase, are matched by the variant contents of the address fields and a Cartesian product is created for identity matching. However, the CIS records which are used in this stage of the match, are filtered to remove those NINO records which were attached to a successful match in stage 2.

This has the effect that whilst potentially reducing the strength of the address match the number of records potentially being incorrectly attached are reduced.

- Stage 4 – Address Detail based Identity Match – The agreed questions will be passed against the Cartesian product, and for those that pass uniquely with an identified strong match question, they will be output out to a successful match pot. Those that failed the identity match, or those that did not have a strong enough match will be failed to pass through the next stage.
- Stage 5 – Postcode and last name block – Those records within the register which were not successfully passed through stage 4, or failed to match via the address detail blocking stage, are matched by a combination of full POSTCODE and LASTNAME and a Cartesian product is created for identity matching. However, the CIS records which are used in this stage of the match, are filtered to remove those NINO records which were attached to a successful match in stage 2 & 4.

The effect of this stage would be that it would combat the difficulties in resolving addresses, but postcode to postcode matching means for those areas where a large number of households exist within a postcode, the Cartesian product becomes larger and there is more risk of incorrectly matching. The inclusion of last name as part of the join reduces the risk of that whilst also reducing the volume of data being processed.

- Stage 6 – Postcode and Last name based Identity Match – Unlike stage 2 and 4 only the following questions will be passed through this stage as effectively a fuzzy match at this point would be of little value. It should be noted that NINO will not be present on electoral data during CDR, therefore those questions noted below which include NINO will never be applicable.

7.9 Records successfully matched at stages 2, 4 & 6 are appended with the pot of records which failed at stage 5, as well as those records which failed at initial validation, to create an output file with the same number of records as received in the originating register.

8. Fuzzy Matching

- 8.1 The matching algorithm includes a series of questions which refer to fuzzy matching. For the baseline code we have only used SOUNDEX as a method of fuzzy matching, but soundex can be seen as Western English Biased, which means that when used in densely populated areas effected by historic and current migration it becomes a less useful tool, and can also skew results.
- 8.2 Also as a driver to combat misspelling, the fact that SOUNDEX always uses the first letter of a string as a constant, means that it does not tackle a mis-spelt name where the first letter is mis-spelt.
- 8.3 SOUNDEX also suffers from the fact that the value it returns can be the same for 2 different strings.

9. Multiple Matches

- 9.1 The use of address blocking, and the concept of Cartesian products means that when identity matching is carried out, multiple matches can be assigned to a single input source record.
- 9.2 Traditionally in this instance, the record would normally be flagged as not matched. However, for the E RTP project, the view was taken that providing information relating to multiple matching would add value for an Electoral Registration Officer.
- 9.3 The current output therefore contains a flag which indicates where a record has multiple matches, and also a note of the level of the best match. So in the instance that 1 input source record has a strong match against 1 matching source record, and a weaker match against another, then the Electoral Officer can choose to accept or reject this record on that basis.
- 9.4 It was identified during the pilot phase that more information around the multiple matching as well as the level at which a household was identified in the address blocking, would be beneficial, and therefore these now form part of the output file created in the algorithm.

Appendix 1 – Output File/Questions

The table below represents the series of columns provided by DWP following the matching of electoral data back to the IER system. Although all of the following information passed to the IER system, not all of it is subsequently made available to the Electoral Registration Officers.

DATA RECORDS				
[Note: Records on the output file should only include the columns that have actually been populated with any results from the match. The values provided in each column for each record will indicate what the data represents and then a colon separating the label from the value e.g. the field containing the Initial data validation (IDV) will hold values like "IDV:TRUE"]				
COL ORDER	COLUMN TEXT	TYPE	LENGTH	DESCRIPTION
1	ID	VARCHAR	27	ELECTORAL REGISTER ID
2	IS	VARCHAR	50	STAGE AT WHICH RECORD MATCHED
3	AMS	VARCHAR	50	ADDRESS STAGE AT WHICH RECORD MATCHED
4	IDV	VARCHAR	30	FILE FAILED INITIAL DATA VALIDATION
5	NIM	VARCHAR	30	NINO MISSING
6	DOBM	VARCHAR	30	DOB MISSING
7	CADP	VARCHAR	30	CURRENT ADDRESS LINES MISSING
8	PADP	VARCHAR	30	PREVIOUS ADDRESS LINES MISSING
9	CUPM	VARCHAR	30	CURRENT UPRN MISSING
10	PUPM	VARCHAR	30	PREVIOUS UPRN MISSING
11	ABMCUP	VARCHAR	30	CURRENT UPRN NOT FOUND ON DWP ADDRESSBASE
12	ABMPUP	VARCHAR	30	PREVIOUS UPRN NOT FOUND ON DWP ADDRESSBASE
13	CAUD	VARCHAR	30	ERO CURRENT ADDRESS DETAILS DIFFER TO UPRN LINK
14	PAUD	VARCHAR	30	ERO PREVIOUS ADDRESS DETAILS DIFFER TO UPRN LINK
15	CUAM	VARCHAR	30	UPRN CURRENT ADDRESS MATCH FOUND ON CIS
16	CNPA	VARCHAR	30	DEFAULT TO NULL
17	CHDWPA	VARCHAR	30	DEFAULT TO NULL
18	PUAM	VARCHAR	30	DEFAULT TO NULL
19	PNPA	VARCHAR	30	DEFAULT TO NULL
20	PHDWPA	VARCHAR	30	DEFAULT TO NULL

UNCLASSIFIED

21	CNUAM	VARCHAR	30	NON UPRN CURRENT ADDRESS MATCH FOUND ON CIS
22	CNNPA	VARCHAR	30	DEFAULT TO NULL
23	CNHDWPA	VARCHAR	30	DEFAULT TO NULL
24	CNUAM1	VARCHAR	30	DEFAULT TO NULL
25	CNUAM2	VARCHAR	30	DEFAULT TO NULL
26	CNUAM3	VARCHAR	30	DEFAULT TO NULL
27	CNUAM4	VARCHAR	30	DEFAULT TO NULL
28	CNUAM5	VARCHAR	30	DEFAULT TO NULL
29	PNUAM	VARCHAR	30	DEFAULT TO NULL
30	PNNPA	VARCHAR	30	DEFAULT TO NULL
31	PNHDWPA	VARCHAR	30	DEFAULT TO NULL
32	PNUAM1	VARCHAR	30	DEFAULT TO NULL
33	PNUAM2	VARCHAR	30	DEFAULT TO NULL
34	PNUAM3	VARCHAR	30	DEFAULT TO NULL
35	PNUAM4	VARCHAR	30	DEFAULT TO NULL
36	PNUAM5	VARCHAR	30	DEFAULT TO NULL
37	IMD1	VARCHAR	30	IDENTITY MATCH ACHIEVED
38	IMD2	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME, MIDDLE_NAME
39	IMD3	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
40	IMD4	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME
41	IMD5	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME INITIAL
42	IMD6	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME
43	IMD7	VARCHAR	30	NINO, DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS
44	IMD8	VARCHAR	30	NINO, DOB, LASTNAME, MIDDLE_NAME
45	IMD9	VARCHAR	30	NINO, DOB, LASTNAME, MIDDLE_NAME INITIAL
46	IMD10	VARCHAR	30	NINO, DOB, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME
47	IMD11	VARCHAR	30	NINO, DOB, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME INITIAL
48	IMD12	VARCHAR	30	NINO, DOB, LASTNAME, FUZZY FIRSTNAME
49	IMD13	VARCHAR	30	NINO, DOB, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

UNCLASSIFIED

50	IMD14	VARCHAR	30	NINO, DOB, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
51	IMD15	VARCHAR	30	NINO, DOB, FUZZY LASTNAME, FIRSTNAME
52	IMD16	VARCHAR	30	NINO, DOB, FUZZY LASTNAME, MIDDLE_NAME
53	IMD17	VARCHAR	30	NINO, DOB, FUZZY LASTNAME, MIDDLE_NAME INITIAL
54	IMD18	VARCHAR	30	NINO, DOB, FUZZY LASTNAME
55	IMD19	VARCHAR	30	NINO, DOB, LASTNAME
56	IMD20	VARCHAR	30	NINO, DOB, MIDDLE_NAME INITIAL
57	IMD21	VARCHAR	30	NINO, DOB, MIDDLE_NAME
58	IMD22	VARCHAR	30	NINO, DOB
59	IMD23	VARCHAR	30	NINO, LASTNAME, FIRSTNAME, MIDDLE_NAME
60	IMD24	VARCHAR	30	NINO, LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
61	IMD25	VARCHAR	30	NINO, LASTNAME, FIRSTNAME
62	IMD26	VARCHAR	30	NINO, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME
63	IMD27	VARCHAR	30	NINO, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME INITIAL
64	IMD28	VARCHAR	30	NINO, LASTNAME, FIRSTNAME FIRST 3 INITIALS
65	IMD29	VARCHAR	30	NINO, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME
66	IMD30	VARCHAR	30	NINO, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME INITIAL
67	IMD31	VARCHAR	30	NINO, LASTNAME, FUZZY FIRSTNAME
68	IMD32	VARCHAR	30	NINO, LASTNAME, MIDDLE_NAME INITIAL
69	IMD33	VARCHAR	30	NINO, LASTNAME, MIDDLE_NAME
70	IMD34	VARCHAR	30	NINO, LASTNAME
71	IMD35	VARCHAR	30	NINO, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME
72	IMD36	VARCHAR	30	NINO, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
73	IMD37	VARCHAR	30	NINO, FUZZY LASTNAME, FIRSTNAME
74	IMD38	VARCHAR	30	NINO, FUZZY LASTNAME, MIDDLE_NAME
75	IMD39	VARCHAR	30	NINO, FUZZY LASTNAME, MIDDLE_NAME INITIAL
76	IMD40	VARCHAR	30	NINO, FUZZY LASTNAME
77	IMD41	VARCHAR	30	DOB, LASTNAME, FIRSTNAME, MIDDLE_NAME
78	IMD42	VARCHAR	30	DOB, LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

UNCLASSIFIED

79	IMD43	VARCHAR	30	DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME
80	IMD44	VARCHAR	30	DOB, LASTNAME, FIRSTNAME
81	IMD45	VARCHAR	30	DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS
82	IMD46	VARCHAR	30	DOB, LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME INITIAL
83	IMD47	VARCHAR	30	DOB, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME
84	IMD48	VARCHAR	30	DOB, LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME INITIAL
85	IMD49	VARCHAR	30	DOB, LASTNAME, FUZZY FIRSTNAME
86	IMD50	VARCHAR	30	DOB, LASTNAME, MIDDLE_NAME
87	IMD51	VARCHAR	30	DOB, LASTNAME, MIDDLE_NAME INITIAL
88	IMD52	VARCHAR	30	DOB, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME
89	IMD53	VARCHAR	30	DOB, FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
90	IMD54	VARCHAR	30	DOB, FUZZY LASTNAME, FIRSTNAME
91	IMD55	VARCHAR	30	DOB, FUZZY LASTNAME, FIRSTNAME FIRST 3 INITIALS
92	IMD56	VARCHAR	30	DOB, FUZZY LASTNAME, MIDDLE_NAME
93	IMD57	VARCHAR	30	DOB, FUZZY LASTNAME, MIDDLE_NAME INITIAL
94	IMD58	VARCHAR	30	DOB, FUZZY LASTNAME
95	IMD59	VARCHAR	30	DOB, LASTNAME
96	IMD60	VARCHAR	30	LASTNAME, FIRSTNAME, MIDDLE_NAME
97	IMD61	VARCHAR	30	LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
98	IMD62	VARCHAR	30	LASTNAME, FIRSTNAME
99	IMD63	VARCHAR	30	LASTNAME, FIRSTNAME FIRST 3 INITIALS
100	IMD64	VARCHAR	30	LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME
101	IMD65	VARCHAR	30	LASTNAME, FUZZY FIRSTNAME, MIDDLE_NAME INITIAL
102	IMD66	VARCHAR	30	LASTNAME, FUZZY FIRSTNAME
103	IMD67	VARCHAR	30	LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME
104	IMD68	VARCHAR	30	LASTNAME, FIRSTNAME FIRST 3 INITIALS, MIDDLE_NAME INITIAL
105	IMD69	VARCHAR	30	FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME
106	IMD70	VARCHAR	30	FUZZY LASTNAME, FIRSTNAME, MIDDLE_NAME INITIAL
107	IMD71	VARCHAR	30	FUZZY LASTNAME, FIRSTNAME

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

UNCLASSIFIED

108	IMD72	VARCHAR	30	FUZZY LASTNAME, FIRSTNAME FIRST 3 INITIALS
109	IMD73	VARCHAR	30	FIRSTNAME AND LASTNAME REVERSED
110	IMD74	VARCHAR	30	SURNAME CHANGED DUE TO CHANGE IN MARITAL STATUS
111	IMD75	VARCHAR	30	MORE THAN ONE MATCH ACHIEVED
112	IMD76	VARCHAR	30	NUMBER OF CIS RECORDS MATCHED
113	NDCISM	VARCHAR	30	NO DWP CIS MATCH
114	NDWPM	VARCHAR	30	NO DWP MATCH (ORPHAN)
115	OS	VARCHAR	30	IDENTITY MATCH SCORE
116	AGE	VARCHAR	30	EXPECTED AGE
117	DOD	VARCHAR	30	DWP RECORDS DATE OF DEATH
118	UPD	VARCHAR	30	DATE OF LAST UPDATE
119	UPD1	VARCHAR	30	DATE RANGE OF LAST UPDATE
120	CDATE	VARCHAR	30	DATE OF LAST CIS UPDATE
121	RDATE	VARCHAR	30	DATE OF DWP RUN
122	QCON	VARCHAR	30	CONCATENATED STRING OF IDENTITY MATCH QUESTION LEVEL
123	CNUAM6	VARCHAR	30	POSTCODE, LAST NAME
124	IMD77	VARCHAR	30	NINO, DOB, CONCATENATED NAME MATCH
125	IMD78	VARCHAR	30	DOB, CONCATENATED NAME MATCH
126	IMD79	VARCHAR	30	CONCATENATED NAME MATCH
127	CT1	VARCHAR	30	HOW MANY ERO RECORDS PER POSTCODE
128	CT2	VARCHAR	30	HOW MANY ERO CONCATENATED NAME RECORDS PER POSTCODE
129	CT3	VARCHAR	30	HOW MANY ERO CONCATENATED NAME RECORDS PER POSTCODE STUB +1
130	CT4	VARCHAR	30	HOW MANY ERO CONCATENATED NAME RECORDS PER POSTCODE STUB
131	CT5	VARCHAR	30	HOW MANY CIS RECORDS PER ERO POSTCODE
132	CT6	VARCHAR	30	HOW MANY CIS CONCATENATED NAME RECORDS PER ERO POSTCODE
133	CT7	VARCHAR	30	HOW MANY CIS CONCATENATED NAME RECORDS PER ERO POSTCODE STUB +1
134	CT8	VARCHAR	30	HOW MANY ERO CONCATENATED NAME RECORDS PER ERO

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

UNCLASSIFIED

				POSTCODE STUB
--	--	--	--	---------------

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED

IER – DWP Data Matching Methodology – Confirmation Exercise

- 20 -