PRECISION WATER LTD

FOCUSED ON MARKET ACCURACY

PRECISION WATER LTD PROJECT DÉJÀ VU

SPID TO VOA & UPRN MATCHING
PROJECT FINAL REPORT

Version 1.0
4 November 2022

Contents

1. Version control

Version number	Date	Author	Description
0.1	22/07/2022	Damian Sharp	Outline Draft
0.2	29/08/2022	Damian Sharp	First Draft
0.3	30/09/2022	Damian Sharp	Further draft
0.4	28/10/2022	Damian Sharp	Near final draft
1.0	04/11/2022	Damian Sharp	Final Report

2. Executive Summary

<u>Headlines</u>

Project Déjà Vu successfully demonstrated that the model used in Scotland to match SPIDs to non-domestic rates references can be adapted to the English market and deliver match rates of **72-86%** to both VOA and UPRN without service verification visits.

The project identified and provided evidence for simple market updates that would take that match rate to **at least 90%** if wholesalers act on recommendations.

The methodology from the pilot project:

- provides high-confidence matches
- is **scalable** and can be used for larger areas including entire wholesaler operating areas or even the whole of the English non-household market
- was demonstrated to work across the different types of geographical area in England from town centre to isolated rural premises
- proved successful both in areas where water and sewerage were supplied by the same wholesaler and where there was a water-only wholesaler
- was effective whether the relevant wholesalers had done extensive matching before market opening or not.

<u>Scope</u>

The project matched all SPIDs in two separate areas, each of around 10,000 SPIDs.

The first area was between Walsall and Wolverhampton (served by South Staffordshire Water and Severn Trent Water). The second area was around Huddersfield and Halifax (served by Yorkshire Water).

<u>Results</u>
After review by Project Déjà Vu, the matching status was:

	Phase 1	Phase 2	Total
Core SPIDs	10,557	11,817	22,374
Matched to UARN	61.6%	72.4%	67%
Verified absence code[1]	10.8%	10.3%	10.5%
Valid CMOS status	**72.4%**	**82.7%**	**78%**
SPIDs for deregistration	5.0%	2.1%	3.5%

The key features of Phase 1 were **identifying new matches** (over half of matches were new) and **merging multiple SPIDs for the same premises** (including a significant number of SPIDs that needed to be paired).

---

[1] An absence code of "other" was not counted as verified

The key features of Phase 2 were **updating matches to reflect changes in use** (3% churn per year) and **identifying additional matches** (but to a much lesser extent than Phase 1 – only 15% of matches were new).

3. Summary Project Brief

Project Déjà Vu is a pilot project to demonstrate whether the methodology successfully developed in Scotland to match SPIDs to the non-domestic rating list can be adapted to the different market rules and data sets of the English competitive water market.

In Scotland, the Central Market Agency undertook a project to match SPIDs to the references held by the Scottish Assessors Association (equivalent of the Valuation Office Agency in England). Known as the SAA project, this work successfully established matches or valid absence reasons for over 80% of supply points and identified corrective actions for the remaining SPIDs.

An adapted methodology was trialled and developed during the project as new issues were identified. Particular adaptations were needed to take account of the different approach of Ofwat to eligibility and bulk supplies.

Matches were sought for the Valuation Office Agency (VOA) reference number and for the Unique Property Reference Number.

The overall approach was to follow a structured review methodology of:

- Matching algorithm to suggest potential matches
- Manual review of all suggested matches
- Manual review of all unmatched SPIDs to identify
  - additional matches
  - premises where no VOA reference and/or UPRN can be expected
  - SPIDs to be deregistered
- Manual review of all unmatched VOA references and UPRNs to identify
  - additional matches
  - premises very unlikely to have water or sewerage services
- Review site visit information for
  - quality control
  - help in resolving complex sites

The success of the methodology was judged primarily by the percentage of SPIDs that have, with a high degree of confidence:

- identified VOA reference(s)
- identified UPRN
- valid exception codes

A secondary criterion for evaluating the methodology was the percentage of deregistrations. Project Déjà Vu had two potentially successful outcomes:

(1) The percentage of SPIDs that don´t meet either the primary or secondary success criteria is under 25% - compared with the current market position of 30-50% depending on the wholesaler; or

(2) The methodology is shown not to produce significant improvements over existing market data

4. Business Drivers

Matching SPIDs with VOA reference numbers (referred to here as Unique Assessor Reference Numbers or UARNs[2]) and Unique Property Reference Numbers (UPRNs) has multiple benefits for market participants.

A high-percentage match
- reduces market friction
- reduces operating costs for wholesalers and retailers
- allows more efficient identification of gap sites
- allows pro-active response to changes in property use
- can underpin charging reform

Reduced market friction

Supply point address data quality varies greatly in the market with incomplete or inaccurate address data causing problems for wholesalers and retailers alike. At the most basic level, retailers and wholesalers can find themselves at cross purposes over which premises are referred to in an operational form.

Matching SPIDs to external references leads to reduction in:
- abortive visits from bilateral requests and meter reading
- creation of duplicate Supply Points
- incorrect deregistration of Supply Points

These reductions prevent unnecessary disputes between wholesalers and retailers and speed up resolution of genuine issues that impact customers.

Reduced operating costs

In addition to wholesalers and retailers benefiting from reduction in market friction, there are other areas where workload falls, and operating costs are reduced.

Accurate address data including UARN and UPRN references make it easier to establish whether premises are occupied or vacant and trace owners and occupiers.

Retailers spend substantial amounts verifying whether premises are vacant, and revenue is lost because this process is more difficult than it needs to be.

Similarly, debt control for owners / occupiers who can pay but hope to avoid payment is hampered by poor address and occupancy data. Improving address data will both improve collection and reduce the costs of debt management.

---

[2] Project Déjà Vu matched SPIDs to the UARN. The UARN is not held in CMOS but the UARN can be mapped directly to the VOA BA reference number which is data item D2037.

More efficient identification of gap sites

Current gap site incentive schemes have a high rejection rate due, in part, to retailers applying for sites that turn out already to have a SPID with incomplete or inaccurate address data.

Increasing the match rate to UARN and UPRN and following that up with address updates will mean that retailers and wholesalers will spend less time on premises already in the market and will be able to concentrate on genuine gap sites where the only questions are about whether the premises are served and whether they are eligible premises.

Pro-active response to changes in property use

Matching and maintaining UARN and UPRN records makes it much easier for wholesaler and retailer alike to keep track of changes in property use through links to external databases and the planning system.

This can include picking up splits or merges in properties as well as demolitions and conversions to household premises.

UARNs and UPRNs can also help retailers track move in, move out activity by customers allowing them to keep the vacancy status up to date in the market and improve debt management.

Charging reform

In Scotland, high-quality matching for a high percentage of SPIDs was an essential prerequisite for two significant changes in charging practice.

Scottish Water introduced vacant charging in 2017 so that all premises contribute to their share of the costs of common services. This is particularly true of surface water drainage where vacancy has no impact on reducing the services that the wholesaler must provide.

The link to UARNs also allowed a transition to using the current rateable value (RV) for services charged based on the RV. This improves transparency for customers and simplifies billing for retailers.

These particular changes may not be appropriate for England but matching to UARN and UPRN provides additional options to simplify charges and spread the burden over all premises that enjoy the benefit of services.

Cumulative impact

Address matching will not eliminate any of the issues described in this section, but it will reduce the problems caused by each one and help wholesalers and retailers focus on those premises with genuine data issues.

This can be the start of a virtuous circle of data improvement where each step forward makes the next step easier.

Prior to the project starting, around 45% of core SPIDs did not have a linked UPRN and 45% of core SPIDs did not have a linked VOA BA reference.

Further analysis reveals that of the core SPIDs that did have VOA BA references, 20% of these VOA BA references are not currently in the VOA rating list. Additionally, analysis of UPRNs show that of the 55% of core SPIDs that do have a linked UPRN, 1% of these UPRNs are not valid UPRNs and 4% of them are linked to two of more core SPIDs.

5. Selection of pilot areas

Project Déjà Vu was funded to look at two pilot areas of about 10,000 SPIDs each. The areas chosen were between Wolverhampton and Walsall for the first phase and around Huddersfield and Halifax for the second phase.

These pilot areas were selected to provide coverage of different types of SPID set up that are common in the English market.

Once the desired criteria were established, specific postcode areas were selected to keep distances down for our site visit delivery partner, LRS, thus allowing more site visits to be undertaken within the project budget.

Wholesaler selection

The criteria for wholesalers were to find groups of SPIDs with all the following characteristics:

- Water Only Company (WOC) with no sewerage
- WOC with Water and Sewerage Company (WASC) sewerage
- WASC water and sewerage
- Sewerage only

To give greater coverage no wholesaler would appear in both pilot areas. Also, no wholesaler owned by the same parent company as a phase one wholesaler would appear in the phase two pilot area.

Mix of premises

Postcode areas needed to have a mix of premises and not be disproportionately dominated by any one or two types of premises.

Urban and rural considerations

Within two pilot areas it is not possible to replicate all the different types of settlement from large city centre to rural areas with no large settlements. However, the pilot areas were chosen to include:

- Town centre
- Town outskirts
- Semi-rural areas
- Rural areas close to large settlements

Contiguity and whole outcode areas

In selecting pilot areas, we took account of lessons learned from Scotland where partial postcodes and postcode errors were most easily caught and dealt with by tackling all SPIDs in an area rather than a series of smaller samples.

Our selection required that:

- postcode outcode areas were either included or excluded in full, eg every SPID and UPRN with postcode WV5 xxx is included
- chosen postcode outcode areas must be contiguous (ie, share a common border) and enclaves were not allowed

Further details of the pilot areas can be found at Appendix A.

6. Data loaded for analysis

The team loaded three datasets into Precision Water's Matchbox system:

- Valuation Office Agency data
- Unique Property Reference Number data from AddressBase
- Market Data Set files from CMOS

Valuation Office data

Project Déjà Vu used an extract of the non-domestically rated premises for the relevant postcode areas.

In addition to all live Unique Assessor's Reference Numbers (UARNs) this extract includes UARNs that have been deleted recently. This is very valuable in tracking change of use and churn in UARN data.

During the project the VOA released updates at roughly e-week intervals. The initial match for Phase 1 was done against the VOA files with identifier Epoch 28. Before Phase 2 began the VOA files for Epoch 31 were uploaded and included new UARNs as well as deleting some premises from the non-domestic rates list.

Unique Property Reference Number (UPRN) data

Precision Water secured an extract of UPRN data from AddressBase Premium. To keep costs within the project budget, the extract had to be limited to premises that might reasonably be eligible for inclusion in the English non-household market.

UPRN data were uploaded for commercial premises and other potentially eligible premises (eg places of worship, community halls, care homes etc).

The UPRN data also included alternative addresses for premises and historic records showing recent changes in UPRN structure.

The UPRN data also includes a cross-reference to the UARN where the premises are on the non-domestic rates list.

Market Data Set

Six MDS files (WSPID, SSPID, WSSCO, SSSCO, METER and READS) were uploaded to Matchbox as well as the New and Partial SPIDs (NAPS) file.[3]
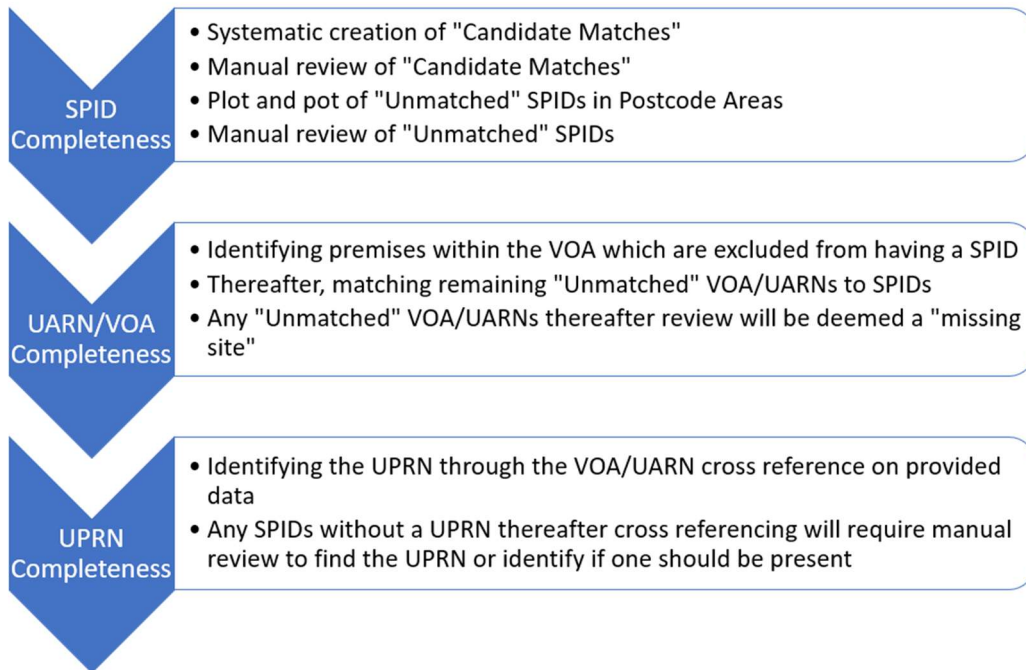
The MDS files were uploaded initially for 7 January 2022 and the files were updated with the 10 May 2022 versions ahead of the start of Phase 2 of the project.

---

[3] The Matchbox system is capable of handling all MDS files but concentrated on these 6 files for Project Déjà Vu

7. Summary Methodology

This section contains a summary of the methodology used to match SPIDs with UARNs and UPRNs. The detailed methodology contains intellectual property of the Central Market Agency and Precision Water Ltd and is not reproduced here.

The project followed a systematic approach as set out in the diagram below.

SPID Completeness
- Systematic creation of "Candidate Matches"
- Manual review of "Candidate Matches"
- Plot and pot of "Unmatched" SPIDs in Postcode Areas
- Manual review of "Unmatched" SPIDs

UARN/VOA Completeness
- Identifying premises within the VOA which are excluded from having a SPID
- Thereafter, matching remaining "Unmatched" VOA/UARNs to SPIDs
- Any "Unmatched" VOA/UARNs thereafter review will be deemed a "missing site"

UPRN Completeness
- Identifying the UPRN through the VOA/UARN cross reference on provided data
- Any SPIDs without a UPRN thereafter cross referencing will require manual review to find the UPRN or identify if one should be present

a) Creating candidate matches

We used a series of algorithms to compare addresses from the MDS and the VOA dataset and create candidate matches. Those matches were then given a confidence rating derived from the extent of the match.

These algorithms considered the extent of matching characters between the two addresses (once manipulated into a common format).

The algorithms also accounted for the precision of the addresses matched and whether there were other very similar addresses in the VOA dataset. For example "15 High Street" and "15 High St" are a very strong match when there's no other premises at 15 High Street but less strong if the VOA contains "First Floor, 15 High Street" and "Second Floor, 15 High Street".

b) Structured review process

All postcode areas were reviewed using a common, structured review process to ensure consistency and complete coverage.

The same steps were used to review all candidate matches and seek to create new matches where those could be identified.

The process followed three distinct phases:

- SPID completeness – ensuring every SPID was reviewed
- VOA completeness – ensuring every UARN was reviewed
- UPRN completeness – ensuring every UPRN was reviewed

Initial review of all SPIDs, UARNs and UPRNs in a postcode outcode area were conducted by the same individual. This allowed the reviewer to gain a complete understanding of the outcode area and improved consistency of decision making.

Where the initial reviewer was unable to resolve matches or wanted a second opinion on premises, they would escalate the SPID for review by one of the senior team.

c) SPID Completeness

The objectives of the SPID completeness phase were to ensure that:

- all candidate matches were reviewed and
- all unmatched SPIDs were reviewed to make additional matches or determine that no match could be expected, as appropriate

At the end of this process as many SPIDs as possible were matched to the VOA register. This allowed us to identify both the UARN and BA reference number where a SPID link was identified.

Within this phase there was a structured approach to reviewing candidate matches then reviewing unmatched SPIDs.

All candidate matches identified through the algorithms are reviewed manually by an agent, where the plotting and actioning of candidate matches being categorised in their respective groups (SPID-UARN):

- one-to-one
- many-to-one
- one-to-many
- many-to many

Thereafter, the actioning of all unmatched SPIDs (SPIDs without a candidate match) occurs. These are grouped up into their respective postcode outcode groups and batched/worked in full in these postcode areas. This allows for greater knowledge of the area by an individual area, where a group of complex sites, addresses, etc. can be actioned together.

Whenever new matches were made, candidates matches were broken or SPIDs were identified as having no match, a reason code was recorded in Matchbox and the reviewer had the opportunity to make additional comments and record evidence for the decision.

1 to 1 Candidate Matches

All matches were reviewed to determine whether they should be confirmed or broken.

Other potential matches for that SPID were reviewed to determine whether a better match existed or whether the SPID might serve more than one premises on the VOA list.

Many-1 Candidate Matches

These are situations where more than one SPID may serve the same premises.

First, each match was reviewed to determine whether it was a sufficiently strong match to the premises. Weaker matches were broken.

After that initial review, if there was still more than one SPID serving the premises then the SPIDs were reviewed to determine whether they represented a:

- **Merge** – to correctly record all the services to that UARN, the services would need to all be transferred to a single SPID and the other SPIDs deregistered. This could happen due to reconfiguration of a building, including physical merges of two premises or due to incorrect site set up in the market
- **Duplicate** – all the services on that SPID, if any, are already on other SPIDs serving those premises. Duplicate SPIDs were recommended for deregistration.

1-Many Candidate Matches

These are situations where a single SPID appears to serve more than one UARN.

First, each match was reviewed to determine whether it was a sufficiently strong match to the premises. Weaker matches were broken.

After that initial review, if the SPID still appeared to serve more than one UARN then the UARNs were reviewed to determine whether they represented a:

- **Bulk meter** – multiple premises served by a single meter with all services accounted for on that one SPID. Bulk meter SPIDs were reviewed to check that it was credible that they met the definition of single eligible premises within Ofwat's latest guidance[4].
- **Split** – physical changes have been made to the premises that mean that they no longer meet the definition of single eligible premises and multiple SPIDs are required. In some cases, this might require metering arrangements to change or building agreements to be reached among customers.

Many-Many Candidate Matches

---

[4] "Eligibility Guidance on whether business customers in England and Wales are eligible to switch their retailer – DRAFT", February 2022, section 2.2, second and third paragraphs

These are situations where the results of the initial matching algorithm showed that SPIDs could be matched to more than one UARN and UARNs could be matched to more than one SPID.

Further human review was needed to try to simplify these groups. The ideal result would be to resolve them into a series of strong one-to-one matches.

In some cases, further information such as customer names, meter locations, site visit results and/or internet searches allowed all of the matches to be resolved.

In other cases, some simplification was possible but final one-to-one matching was not possible.

Unmatched SPIDs

All unmatched SPIDs to establish whether:

- one or more matching UARNs could be identified
- there was evidence that the SPID should be deregistered (eg demolished or converted to non-household premises)
- the SPID served premises that would not be expected to appear on the VOA register (eg places of worship, care homes, farms)

d) UARN/VOA Completeness

The objective of UARN/VOA completeness is to ensure that all UARN/VOA references can be linked to a SPID. Where this is not the case, and the premises for the VOA/UARN reference is eligible, these will be deemed as "missing sites".

Missing sites may be sites that are:

- not connected to any wholesaler's network (not served)
- being billed as household sites and may need their eligibility reviewed
- gap sites that should have new SPIDs registered

It was outside the scope of Project Déjà Vu to do the further checks necessary to determine which of the missing sites are gap sites.

Identifying Excluded Premises

The first step in UARN/VOA completeness is to exclude premises that are highly unlikely to have water or sewerage services, eg advertising hoardings and ATMs.

A review of the "Primary Description Text" was undertaken for each UARN in the VOA dataset. This excluded types of premises that are highly unlikely to have services or which are highly likely to be permeable with no water supply.

Unmatched VOA/UARN Review

Once these premises are excluded, each remaining unmatched UARN is reviewed to identify whether it can be matched to an existing SPID, including a review of whether it would form part of a bulk supply.

During this process, UARNs are:

- matched to an existing SPID; or
- excluded because there is evidence that they have no relevant services; or
- designated as missing sites

Once SPID and UARN/VOA completeness was completed, all SPIDs and UARNs had been reviewed and assigned a status, with the majority also being linked (SPID to UARN).

e) UPRN Completeness

The AddressBase Premium data contain a cross-reference between UARN and UPRN. These were used to create links from SPID to UPRN wherever a UARN link had been identified.

UPRNs which covered UARNs that had been excluded during the UARN/VOA completeness stage were also excluded from being expected to have a matching SPID.

Remaining UPRNs were reviewed to see if a SPID could be identified as matching the UPRN. The biggest groups of these matches were for non-household premises that are exempt from non-domestic rates including places of worship, care homes and farms.

Where a new SPID to UPRN match is identified, the AddressBase Premium cross-reference was consulted to see if this also allowed creation of an additional SPID to VOA match. This could happen where the VOA and UPRN addresses are different and/or the additional GIS information in AddressBase Premium allowed a match to be made.

f) Escalation

Whenever the initial reviewer was unsure about the candidate matches or wanted advice on how best to record matches, the UARN and/or SPID was marked for escalation in Matchbox.

This triggered a review by one of the senior team who might also consult other members of the team. The results of these reviews were fed back to all the team to spread lessons learned and ensure consistency of approach among the team.

If the senior team were unsure about whether a match was valid or about how to record the situation at a premises, they consulted Damian Sharp who acted as the final arbiter of any ambiguous matches.

8. Quality assurance

The project's approach to quality assurance was key to the high levels of confidence in matches that were achieved.

There were 3 elements to the quality assurance approach:

- Internal QA within the team
- Review by the Central Market Agency
- Site visits

The Quality Assurance function provided a strong level of confidence that the project deliverables, the results of the pilot matching exercise and the recommendations that flowed from them, were fit for their intended purpose.

a) Internal QA

Throughout the project we operated a rigorous internal quality assurance process where matching decisions (including breaking matches) were reviewed by another member of the team.

The frequency of review was determined by the match status achieved after review. For example, one in every four matches where an address update was proposed was reviewed by another member of the team. The table below sets out the range of statuses and the frequency of sampling for quality assurance.

Each match status that was selected for QA was recorded in an Excel QA tracker, where the tracker included the Core SPID, Agent who actioned the match/selection, Pass/Fail selection and where a Fail was present, relevant comments.

Where the QA process identified that a different status was more appropriate than the one initially selected, that was fed back to the agent who made the original selection. The match status was then updated after the original agent had had the chance to challenge the QA comments if they wished.

Where themes emerged from the QA, these were discussed in team meetings to ensure that lessons learned were shared among the whole team.

The internal QA process led to the creation of some additional statuses where that helped provide more accurate results.

Status	Sample frequency
Address Update	1 in 4
Agricultural	All
Bulk Meter	All
By Elimination	All
Confirmed Match	1 in 10
No VOA	All

Status	Sample frequency
Religious	All
Care Home	All
Duplicate	All
Merge	All
Multi	All
Split	All
Temporary Building Supply	All
Demolished	All
Domestic	All
Deregister	All

Internal QA had a different in approach in both phases:

Phase 1 – A more focused all-in-one approach once all steps of the operations had completed
Phase 2 – A daily review of actions to promote proactiveness and ensure agent accuracy regularly

Quality Assurance was based primarily around process analysis. During the lifetime of the project, the QA function reviewed processes and procedures to identify errors, omissions, or ambiguities, and to recommend enhancements. The review covered documentation, IT systems and project team responses, with a view to assessing project governance, methodology and implementation. The objective of this element of QA is to prevent defects from occurring.

In addition, the Quality Assurance function included an element of Quality Control, by checking draft outputs for defects. This will involve a review of a sample of matches, including an independent corroboration of results and a witnessing of the application of the matching methodology by the project team. The objective of this element of QA is to confirm that defects have not occurred.

b) Review by the Central Market Agency

The Central Market Agency (CMA) reviewed the governance of the project and the delivery plan to satisfy itself that the project's methodology and controls were consistent with delivering the planned results.

CMA staff also sampled matching results and conducted independent assessments of those matches and discussing any questions or different conclusions with the project team. This provided confidence that the methodology would work independent of the person applying the methodology and helped the project team clarify some points of detail in the methodology by explaining them to the CMA.

The CMA produced a separate assurance statement for the project covering project governance, methodology and implementation.

c) Site Visits

In addition to helping resolve complex sites, the site visits were targeted at providing assurance for matches made by the original matching algorithm and those made by agents following their review of potential matches.

Site visits did not verify services but verified the existence (or otherwise) of the premises, occupant (where possible), brief details of surrounding premises, and provided both photographs and GIS location data for the premises visited.
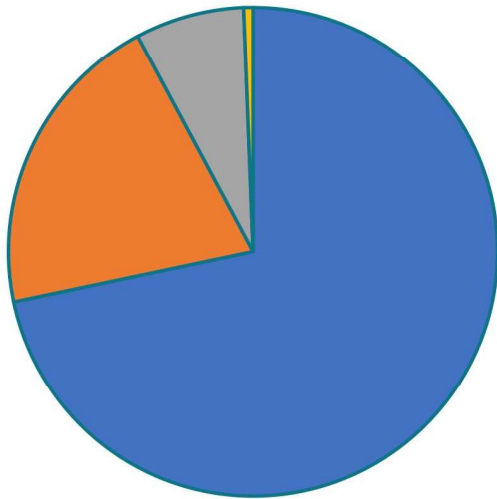
Site visit results were received weekly in a standard format with accompanying photographs.

Every site visit result was reviewed to assess whether it confirmed the existing match made, clarified any uncertainty about matches or contained evidence that an existing match should be broken.

General SPID Stats

P1 - SPID Stats

Category	%
Valid	71.7
Wholesaler Action Required	22.1
Removeable	5.6
Internal Investigation	0.6

- Valid
- Wholesaler Action Required
- Removeable
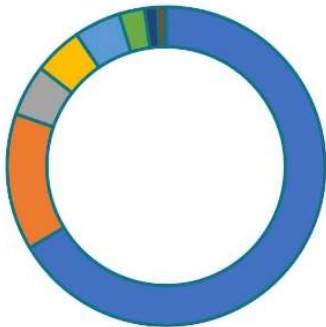- Internal Investigation

P2 - SPID Stats
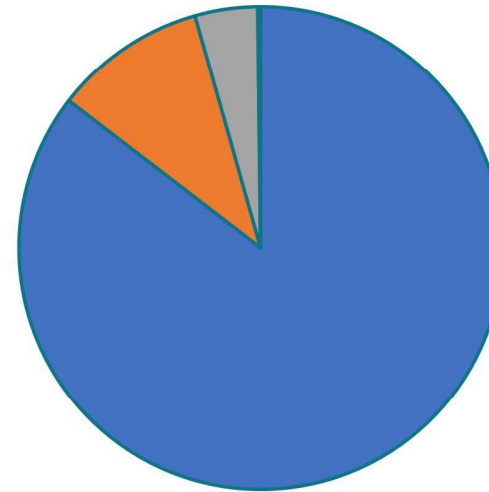
Category	%
Valid	85.9
Wholesaler Action Required	11.8
Removeable	2.1
Internal Investigation	0.2

- Valid
- Wholesaler Action Required
- Removable
- Internal Investigation

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

Phase 1 – Additional SPID Stats

P1 - SPID Drill Down - Valid

Category	%
Confirmed Match	48
Address Update	10.1
No VOA	3.8
Agricultural	3.6
Bulk Meter	3.4
Religious	1.9
Care Home	1
TBS	0.5
By Elimination	0.1

- Confirmed Match
- Address update
- No VOA
- Agricultural
- Bulk Meter
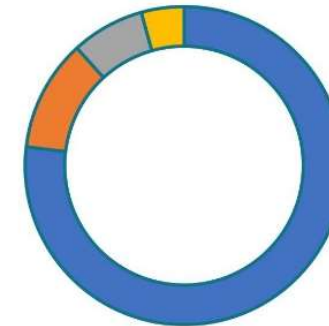- Religious
- Care Home
- TBS
- By elimination

P1 - SPID Drill Down – Removeable
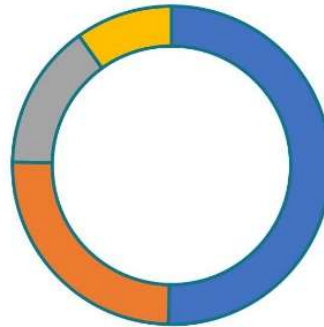
Category	%
Deregister	2.5
Domestic	1.3
Demolished	0.7
Duplicate	0.5

- Deregister
- Domestic
- Demolished
- Duplicate

P1 - SPID Drill Down – Wholesaler Action

Category	%
Merge	17
Multi	2.5
Info Required	1.6
Split	1

- Merge
- Multi
- Info required
- Split

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

Phase 2 – Additional SPID Stats

P2 - SPID Drill Down - Valid
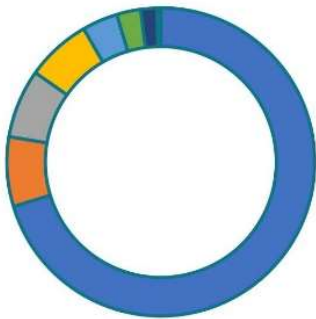
Category	%
Confirmed Match	60.3
Agricultural	6.3
Bulk Meter	6.2
Address Update	5.8
No VOA	3.2
Religious	2.2
Care Home	1.5
TBS	0.3
By Elimination	0.1

- Confirmed Match
- Agricultural
- Bulk Meter
- Address update
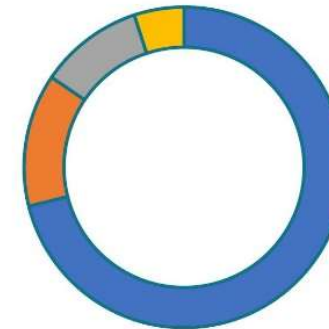- No VOA
- Religious
- Care Home
- TBS
- By elimination

P2 - SPID Drill Down – Removeable
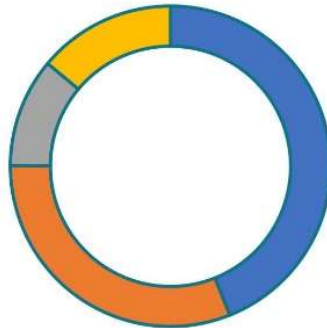
Category	%
Domestic	0.8
Deregister	0.7
Demolished	0.3
Duplicate	0.3

- Domestic
- Deregister
- Demolished
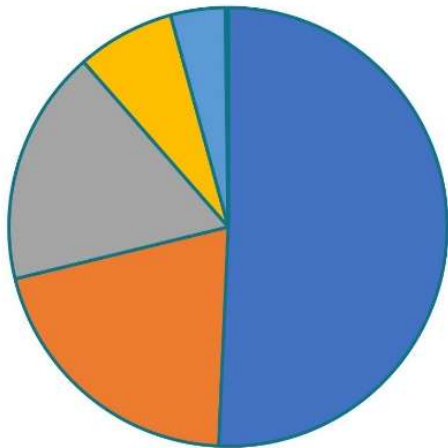- Duplicate

P2 - SPID Drill Down – Wholesaler Action

Category	%
Merge	8.4
Info Required	1.6
Multi	1.2
Split	0.6

- Merge
- Info required
- Multi
- Split

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

General VOA/UARN Stats

P1 - VOA/UARN Stats

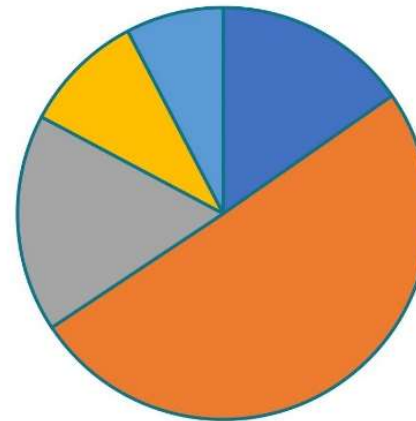| Category	%
UARN found	50.7
No update	20.5
No UARN expected	17.4
Matches at least 1 (of multiple)	7.2
Update required	4
No UARN found	0.2

- UARN found
- No update
- No UARN expected
- Matches at least 1 (of multiple)
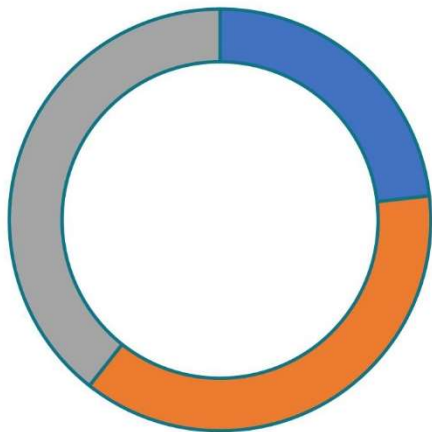- Update required
- No new UARN found

P2 - VOA/UARN Stats

Category	%
No update	50.3
No UARN expected	17.1
UARN found	15.3
Matches at least 1 (of multiple)	9.5
Update required	7.7

- UARN found
- No update
- No UARN expected
- Matches at least 1 (of multiple)
- Update required

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

Additional VOA Stats

P1 – Drill Down – VOA

Category	Count of VOA
Default Excluded	1209
Missing Sites	1956
VOA Deleted	2055

- Default Excluded
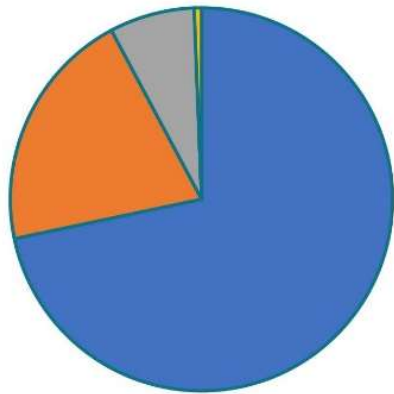- Missing Sites
- VOA Deleted

P2 - Drill Down - VOA
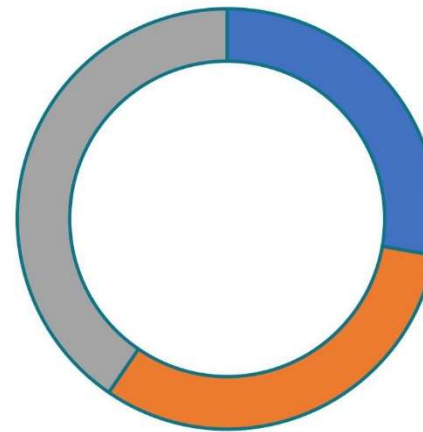
Category	Count of VOA
Default Excluded	1971
Missing Sites	2254
VOA Deleted	2871

- Default Excluded
- Missing Sites
- VOA Deleted

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

General UPRN Stats

P1 - UPRN Stats

Category	%
Valid	71.7
Wholesaler Action Required	22.1
Removeable	5.6
Internal Investigation	0.6

- Valid
- Wholesaler Action Required
- Removeable
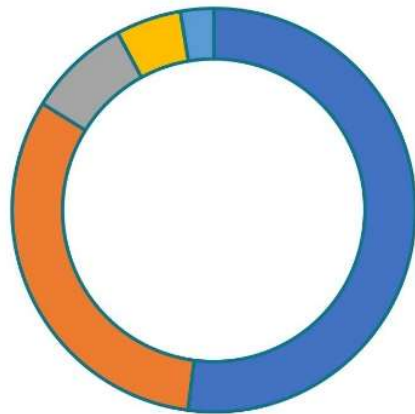- Internal Investigation

P2 - UPRN Stats
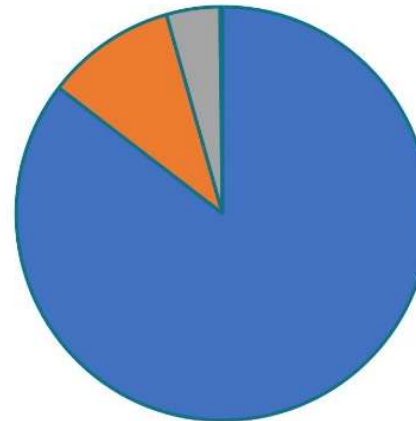
Category	%
Valid	85.9
Wholesaler Action Required	11.8
Removeable	2.1
Internal Investigation	0.2

- Valid
- Wholsaler Action Required
- Removable
- Internal Investigation

PRECISION WATER LTD
FOCUSED ON MARKET ACCURACY

Additional UPRN Stats

P1 - UPRN Drill Down - Valid

Category	%
No update	37.3
UPRN found	22.8
Update required	5.9
No UPRN found	3.8
Matches at least 1 (of multiple)	1.9

■ No update

■ UPRN found

■ Update required

■ No UPRN found

■ Matches at least 1 (of multiple)

P2 - UPRN Drill Down - Valid
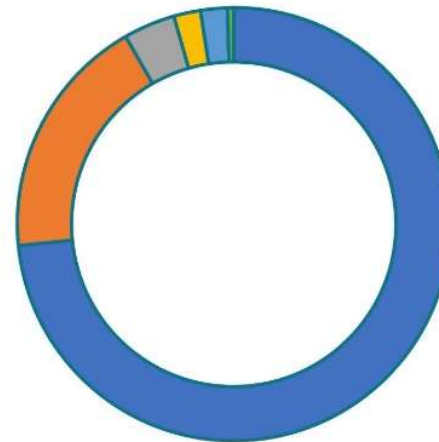
Category	%
No update	62.7
Update required	15.6
UPRN found	3.3
Matches at least 1 (of multiple)	1.7
No UPRN found	1.7
No UPRN Required	0.4

■ No update

■ Update required

■ UPRN found

■ Matches at least 1 (of multiple)

■ No UPRN found

■ No UPRN Required

10.   Site challenges

In carrying out Project Déjà Vu, we had to tackle challenges in the data sets and market rules to produce consistent results. Our approach was to establish a policy on how to deal with the different issues that arose and document the evidence used to make individual decisions in the light of those policies.

Challenges arose from market structure issues (eg multiple premises served by a single SPID, and mixed-use premises) and from market data quality issues (eg address quality and format, and customer name quality).

Market structure issues

The rules of the English market do not allow a one-one match for every SPID and every non-household premises. In particular, unlike in Scotland, Ofwat's guidance on eligibility allows multiple non-household premises to be served by the same supply point if there is a private distribution network or evidence of a joint billing agreement.

The eligibility of mixed-use premises is also a particularly difficult challenge with different Wholesalers taking different approaches.

To resolve market structure issues consistently within Project Déjà Vu we developed policies for the treatment of:

- splits v bulk meters
- mixed-use premises
- eligibility

These policies are set out below.

It is our view that the market would benefit from greater consistency across England in how these issues are tackled: not least because customers with premises across England will become more and more frustrated if consistent practices are not applied.

a) Splits v Bulk Meters

In both Phases of the project, we found evidence that many supply points (and their meters) were serving multiple premises that the VOA had rated separately.

This is legitimate where there is a private distribution network or evidence of a joint billing agreement.

Our policy was only to recommend that such a SPID should be split into individual SPIDs for each VOA premises where there was evidence that a building had been physically altered to divide it into clearly separate premises, eg a building divided into three units each with their own separate entrance directly onto the street.

We assumed in other cases that the Wholesaler had confirmed that there was a private distribution network or that the Retailer had evidence of a joint billing agreement, provided

either by the customer(s) or by the Wholesaler when the SPIDs were transferred into the market in April 2017.

There is a further challenge in how to record legitimate bulk supply SPIDs in the market. It is our view that MOSL must give clear guidance on what should be entered in the VOA reference number and UPRN fields as Wholesalers have understandably taken very different approaches in the absence of such guidance.

MOSL should also consider additional data items that would allow the recording of bulk supplies and the premises they serve, eg through the creation of a "bulk supply" flag or the option to record primary and additional VOA references.

This approach would help reduce the number of abortive gap site applications and facilitate future splits if customers insist on having their own SPIDs and/or premises are physically restructured.

b) Mixed-use premises

From previous work elsewhere, we are aware that the eligibility of mixed-use premises is an issue for all Wholesalers and that several different approaches are used to determine whether mixed-use premises have a principal use as a home. We note that Ofwat continues to resist providing definitive guidance on how to determine this.

For consistency, we have assumed that all mixed-use premises that had SPIDs had correctly been identified as eligible for inclusion in the market.

We have included mixed-use premises without a matching SPID in the list of "missing sites". The missing sites list includes all potentially eligible premises without a matching SPID. This is not the same as a "gap sites" list. No checks have been carried out of whether the premises are served nor of whether it is likely that the principal use is not as a home.

c) Eligibility

For a number of existing SPIDS, we needed to review the SPID against Ofwat's eligibility guidance – typically because of a change of use in those premises.

Where there was clear evidence that the premises had been converted to household premises or had been demolished, then these SPIDs were recommended for removal from the market.

There were other SPIDs where it was not clear whether the supplies served eligible premises or not. Examples of these include common supplies for buildings – taps, bin stores, cleaning cupboards, garages. The Ofwat guidance is that such common supplies are eligible for the market if, and only if, the building in question is eligible for the market.

It was not always possible to confirm the link between the SPID and the master building, especially if the master building was, in fact, likely to be considered as household premises.

These common supplies form a separate category that we recommend should be reviewed by the wholesaler and retailer to confirm eligibility.

<u>Market Data Quality Issues</u>

These are issues that arise from inconsistent or incomplete data held within the MDS. The two areas where this most impacted Project Déjà Vu were:

- Address
- Customer Name / Banner Name

d) Address

Within address data there were two distinct issues – poor quality / incomplete SPID addresses and SPID addresses held in a different format from the VOA.

<u>Poor quality / incomplete address</u>

Wherever we were able to identify a match despite incomplete or poor SPID address data, we have recommended an address update to the National Address Gazetteer address which is compliant with BS7666.

There were a small number of SPIDs where the premises address and other data were too poor to allow a confident match to the VOA records and these were marked as being for wholesaler investigation as they might have further asset records that would allow the premises to be clearly identified.

<u>Address format</u>

The format of addresses could cause issues, especially in multi-occupancy sites and industrial estates.

There were particular issues where units might be known by different numbers and combinations of numbers and letters in the MDS, the VOA and on-site. For example, Unit 24 in the MDS might be described as Unit B/8 in the VOA and as something else again on the signboard / store directory.

Where this was identified, we chose the UPRN address as the primary address.

11.   Supplies that are not addressable

Unique Property Reference Numbers (UPRNs) are always, and Unique Assessor Reference Numbers (UARNs) are mostly, dependent on the premises in question being "addressable". Addressable premises have an official address allocated by the local authority. In most cases they will also have a postcode allocated by Royal Mail.

However, there are supplies, especially water supplies, that are for non-household premises where those premises are not addressable.

The most frequent examples are:

- **Troughs** – the address in the market is usually linked to a farm. However, fields served by troughs can change customer by being leased to a different farmer without any change to the market address. We did not consider it appropriate to match the trough to the current farm because that would create an additional maintenance burden on the Wholesaler.
- **Statues & Fountains** – some statues and fountains have UPRNs and there is a category within the National Address Gazetteer for them, but many will not have UPRNs unless they are considered as "monuments"
- **Standpipes** – these are not recorded in the National Address Gazetteer
- **Signal Boxes** – in most cases signal boxes are recorded as part of the railway undertaking as a single VOA entry and as a cumulo UPRN
- **Railway Stations –** some railway stations have their own UPRN and VOA reference as addressable objects but others are included as part of the railway undertaking.

Where non-addressable objects have services then we have identified a suitable, specific absence code to confirm that no VOA reference and/or UPRN should be expected.

12.   Appendix A – Pilot area details

Phase 1

The phase 1 area was chosen from the WS (Walsall) and WV (Wolverhampton) areas.

The postcode outcodes chosen were:

- WS2 – Bentley, Walsall
- WS3 – Bloxwich, Walsall
- WS10 – Sandwell
- WS11 – Cannock Chase
- WS12 – Cannock Chase
- WV6 – Tettenhall, Wolverhampton
- WV7 – Albrighton, Wolverhampton
- WV8 – Codsall, Wolverhampton
- WV9 – Pendeford, Wolverhampton
- WV10 – Featherstone, Wolverhampton
- WV11 – Wednesfield, Wolverhampton
- WV12 – Short Heath, Wolverhampton
- WV13 – Shepwell Green, Wolverhampton

These postcodes cover the area between (but exclude) Walsall town centre and Wolverhampton city centre.

Number of SPIDs	10,557
Number of UARNs	15,974
Number of UPRNs	28,879
Wholesalers	South Staffordshire Water (WOC)
	Severn Trent Water (WASC)
Urban / rural mix	Largely town outskirts
	Some semi-rural areas
	No major town centre or fully rural areas
Contiguous area?	Yes

Phase 2

The second phase area was chosen from West Yorkshire. The HD, HX, LS, WF and BD postcode areas were considered and the HD and HX areas were chosen as being the best fit for number of SPIDs, mix of premises and providing an urban / rural mix that complemented the phase 1 selection.
The HD postcode area covers Huddersfield and surrounding areas including Brighouse, Holmfirth and Skelmanthorpe.

The HX postcode area covers Halifax and surrounding areas including Elland and Hebden Bridge.

Together these postcode areas span the M62 covering the large towns of Halifax and Huddersfield and their extensive hinterland.

The second pilot area is a lot less compact than the first and includes much more agricultural activity and unfarmed countryside.

Number of SPIDs	11,817
Number of UARNs	21,734
Number of UPRNs	40,773
Wholesalers	Yorkshire Water
Urban / rural mix	Major town centres Town outskirts Small towns Semi-rural areas Rural areas including national park land
Contiguous area?	Yes

13. Appendix B – Results for Phase 1 Area (South Staffordshire / Severn Trent)

The detailed results for the Phase 1 Area were made available to South Staffordshire Water, Severn Trent Water and MOSL.

14.   Appendix C – Results for Phase 2 Area (Yorkshire)

The detailed results for the Phase 2 Area were made available to Yorkshire Water and MOSL.

15. Appendix D – Case Study

Leeds Road Retail Park, Huddersfield

In this case study, we are reviewing the following SPIDs:

3200457252 – Unit J3, Leeds Road Retail Park, Leeds Road, Huddersfield, HD1 6PF

3200457287 – Leeds Road Retail ParkHD1 6PF, Unit K, Leeds Road, Huddersfield, HD1 6PF

3200457244 – Unit J1, Leeds Road Retail Park, Leeds Road, Huddersfield, HD1 6PF

the following VOA references:

11414667000 – UNIT J, LEEDS ROAD RETAIL PARK, LEEDS ROAD, HUDDERSFIELD, HD1 6PF

11465114000 – STARBUCKS UNIT K, LEEDS ROAD RETAIL PARK, LEEDS ROAD, HUDDERSFIELD, HD1 6PF

and the following UPRNs:

10094115536 – Starbucks Unit J1 Leeds Road Retail Park Leeds Road Huddersfield HD1 6PF

10094115537 – Subway Unit J2 Leeds Road Retail Park Leeds Road Huddersfield HD1 6PF

10094115538 – Greggs Unit J3 Leeds Road Retail Park Leeds Road Huddersfield HD1 6PF

A desk analysis of the premises shows the following in StreetView (as at September 2021):

A desk analysis of the UPRN addresses confirms that the businesses listed against each address is accurate.

As you can see we have 3 SPIDs (Unit J3 with customer Greggs PLC, Unit K with no customer name and Unit J1 with customer name Soul Coffee House (East) Ltd – also known as Starbucks).

In this scenario we only have 2 VOA references – One for a Unit J and no firm name and one for a Unit K with a firm name of Starbucks.

We are able to link SPID 3200457244 to VOA 11465114000 and UPRN 10094115536. This is established using the customer name link on both the SPID, VOA & UPRN while the SPID and UPRN have identical addresses. The status used in this scenario is a "Confirmed Match" as we wouldn't recommend updating the address to reflect the VOA address, given we know that the VOA address is inaccurate with the exclusion of the firms name.

We are subsequently left with VOA 11414667000 for Unit J and SPIDs 3200457252 and 3200457287 for Units J3 and K respectively. We now know that SPID 3200457252 address is accurate and can be linked to UPRN 10094115538. We now know through both process of elimination and a desk analysis customer name match that SPID 3200457287 matches to UPRN 10094115537. As a result of this we have came to the following conclusion when matching in system:

SPID 3200457252 matched to VOA 11414667000 with status confirmed match and linked to UPRN 10094115538

SPID 3200457287 not matched to any VOA reference with status "No VOA" while also providing an address update for the address to be "Unit J2 Leeds Road Retail Park Leeds Road Huddersfield HD1 6PF" and a UPRN link of 10094115537.

In conclusion, without using a blend of desk analysis, VOA and UPRN data there would be no plausible resolution and this site would be deemed potentially complex or recommendations would be made solely based from SPID and VOA addresses which would not have provided the best recommendations to resolve the escalation.

16.   Appendix E – List of matching categories

Status	Description
Address Update	Where the provided SPID requires an addition/removal or full address cleanse to increase accuracy
Agricultural	Where the provided SPID is (or is part of) an agricultural supply. This can include farms, farm buildings, allotments, troughs, fields, etc.
Bulk Meter	Where the provided SPID is a bulk meter. This is a recommendation based on factors such as meter size, consumption, VOA/UARN addresses.
Default Excluded	This is used to exclude VOA's where they are deemed not suitable to have a supply (or SPID match). This may be due to the supply being deemed non-permeable or majority domestic.
By Elimination	Where a SPID to VOA/UARN link has been identified when all other links have been made, leaving a link through elimination of others.
Confirmed Match	Where a successful one-to-one link between SPID and VOA/UARN has been identified.
No VOA	Where a SPID to VOA/UARN link cannot be identified. Thereafter it is expected that a UPRN link can be identified manually or a reason for the premises to have neither link (i.e., statue, outside tap, communal residential supplies, etc.).
Religious	Where the provided SPID is (or is part of) a religious premises.
Default	A system/internal status when SPIDs are entered into Precision Matchbox
Potential Match	A system status for use with Candidate Matches
Care Home	Where the provided SPID is (or is part of) a care premises.
Duplicate	Where services on a provided SPID are a duplicate of another.
Escalate	To be used in complex cases or where an agent is unable to completely verify the status of a reference.
Info Required	To be used by an authorised member reviewing escalations. This is a sub-status of escalation advising that further info is required by the wholesaler/retailer to establish the SPID premises.
To Be Updated	To be used by an authorised member reviewing escalations.
Merge	Where the recommendation is to merge multiple SPIDs into one.
Multi	Where there are multiple SPIDs linked to multiple VOA/UARNs and the matches cannot be broken down sensibly.
Potential Gap	Where the provided VOA/UARN cannot be linked to a SPID. This is also known as a "missing site".
Split	Like "Bulk Meter", however the recommendation is to split 1 SPID into multiple. This is recommended when it is obvious that multiple addresses now exist (i.e., 14-16 High Street is now 2 shops, 14 & 16 respectively.)
TBS	Where the provided SPID is an active Temporary Building Supply
Demolished	Where the provided SPID has demolished and there is sufficient evidence to prove this.
Domestic	Where the provided SPID is fully domestic and there is a Council Tax reference to prove this.
Deregister	Where the provided SPID should not exist in the market, however, is not a completed demolishment or the premises does not yet have a Council Tax reference. This can include extremely poor addresses, active change of uses or completed TBS.
VOA Deleted	A system status to highlight VOA/UARN references which have now been removed from the VOA register.

From the above we can identify the following colours of statuses from the below key:

	A confirmed status
	A system status
	A candidate match status
	A recommendation/further action required status
	A removed/remove status