



Piloting approaches to capturing customer experience in the water sector for PR24; D-MeX

August 2024

Prepared by: Accent, 3 Orchard Place, London SW1H 0BF

Contact: Julian Hollo-Tas

E-mail: Julian.hollo-tas@accent-mr.com

Telephone: 020 8742 2211

File name: D-MeX PR24 Pilot Final Report.docx



CONTENTS

1	EXECUTIVE SUMMARY	i
1.1	Introduction	i
1.2	Developer Measure of Experience (D-MeX)	i
1.3	D-MeX, Survey Component	i
2	INTRODUCTION	5
2.1	Background	5
2.2	Objectives	6
2.3	Scope of this Report	6
3	Developer Services Measure of Experience (D-MeX)	7
3.1	Background	7
3.2	Objectives	7
3.3	Existing Data Review	8
3.4	Increasing the prevalence of SLPs, NAVs, and large developers in the survey data.	12
3.5	Additional research questions	14
3.6	Fieldwork	14
4	Appendices	22
4.1	Appendix A - Questionnaire	23

1 EXECUTIVE SUMMARY

1.1 Introduction

Ofwat commissioned Accent to test approaches to delivering customer satisfaction surveys for the C-MeX (Customer Experience and Customer Service components (CES and CSS)), D-MeX (qualitative/survey component only), and Business Customer Experience in Wales surveys (BCIWS), in advance of their continuation into the PR24 period.

The initial stage of the research comprised of a desk review of methodologies, conducted from a macro perspective, offering a bird's eye view of the suite of relevant methodologies potentially available for deployment by Ofwat.

Subsequent phases of the pilot programme involved reviewing the existing outputs of the various surveys (CES & CSS, D-MeX, Business Customer Experience in Wales), and testing (via physical pilots) alternative (and/or enhanced) versions of them.

This report covers the D-MeX survey pilot. Reports covering the literature review, and the reviews and pilots of the C-MeX and BCIWS survey methodologies, are published separately.

1.2 Developer Measure of Experience (D-MeX)

D-MeX is a mechanism to incentivise water companies to provide an excellent customer experience for developer services customers. These customers include small and large property developers, self-lay providers, new appointees, and some residential customers that have new mains connections installed.

Each company receives a D-MeX score based on two components:

- a survey component (previously known as the qualitative component) – a score measuring the performance of the company in a satisfaction survey of developer services customers; and
- a quantitative component – a score measuring the performance of the company across selected Water UK metrics.

Both components contribute equally to the overall D-MeX score for each company.

The scope of this pilot is restricted to the survey component of the measure.

1.3 D-MeX, Survey Component

The survey component of D-MeX is a 0-100 score based on a 0-10 scale satisfaction question. All interviewing is conducted via Computer Assisted Telephone Interviewing (CATI).

Each water company has an annual sample size set according to the number of sample records available (essentially 20% of the volume of contacts, with a minimum of 100 and a maximum of 1,000 interviews).

Research Objectives

The primary objectives of this pilot included:

- Reviewing existing data to obtain learnings from the current methodological deployment.
- Assessing the feasibility of increasing the prevalence of SLPs, NAVs, and large developers, in the data.
- Addressing a number of additional research questions
 - Are there benefits to having different surveys for different customer groups, for example having a different survey for one-off connection customers vs those which transact multiple times with multiple different water companies?
 - Is it feasible to survey customers about their experience with more than one water company in one survey?
- Running a physical pilot to test methodological changes suggested for the next price review period

Existing method/data analysis

For the existing method/data review, particular focus was given to the sample sizes and robustness of data, the sustainability of the surveying, operations learnings from the current deployment, and how to extract more from the data.

Sample sizes and robustness of data

Where sample volumes permitted the maximum number of interviews of 1,000 (large companies only), the 95% confidence interval of each company's score was around 1.75. Where sample availability restricted interviewing to around the minimum of 100 interviews, it was roughly 3. Overall, the mean 95% CI is 2.15. The average gap between each company in the D-MeX (qualitative) ranking table is 1.3 points.

These sample sizes (and margins of error) are appropriate and fit for purpose.

The impact of increasing or decreasing the sample size on the level of confidence that can be attributed to the score was also explored, and details can be found in the specific section of this report; but it should be noted that for smaller water companies' sample is already maximised so little scope exists for increased robustness of data.

Sustainability

A key threat to the sustainability of the measure was identified; the steadily increasing prevalence of customers (particularly those that feature regularly on the sample) requesting that they not be contacted regarding the research in the future ("Do Not Recontact" – DNR).

It was noted that this effect is amplified due to the practical difficulties associated with effectively de-duplicating the sample (which comes from 17 different companies, and where even the smallest inconsistency in spelling or formatting of a name can be enough to prevent effective cross referencing).

How to extract more from the data

The opportunity for advanced analytics (such as Key Driver or CHAID analysis) aimed at enabling the understanding of areas suitable for performance enhancement, with a view to improving D-MeX scores, already exists for Ofwat and water companies (using the existing bank of customer service elements).

Other operational observations

It was concluded that due to the requirement to maximise small sample sizes, CATI remains the most appropriate approach at present.

Sampling monthly (the current regime) was assessed as striking a good balance between the administrative burden of generating, processing the sample, of covering a wide enough period for deduplicating to be as effective as possible, and of approaching participants in a timely fashion post-contact.

The main area of difficulty for participants is identifying the correct transaction. The difficulty is directly related to the availability on the sample of information related to each transaction – which is inconsistent across water companies.

Quotas generally work well, with small company samples maximised, and large company samples capped.

The questionnaire itself works well, and at an average duration of 6.5 minutes could be described as being relatively short (although anecdotal feedback that the duration is too long has been passed back to Ofwat in the past and cited as a reason for non-completion).

Increasing the prevalence of large developers, SLPs, and NAVs

When considering this question it should be noted that the apparent under-representation of SLPs and NAVs in the survey data (8% of the sample, but 0.8% of the interviews) is a result of many transactions being legitimately and necessarily deduplicated down to few rather than under-sampling or low response rates.

There is therefore effectively no more available sample to prioritise, that could be used to physically interview more SLPs/NAVs.

Sufficiently few SLPs and NAVs are interviewed (roughly 2 per quarter per company, for large companies), that weighting their responses up is an approach that should be treated with caution.

Addressing the question of the prevalence of large developers suffers from a rather more fundamental issue; the lack of an agreed definition on what constitutes a large developer. Until this is resolved, assessing both the current situation and also what remedial action may be required, is not possible.

Fieldwork

To answer the two further research questions (benefits to having different surveys for different customer groups, and feasibility of conducting a single interview related to more than one transaction), whilst exploring the potential to increase the prevalence of certain customer groups in the data fieldwork was designed and deployed.

Because of the previously mentioned difficulties associated with defining a “Large” developer, this fieldwork was restricted to SLPs and NAVs.

The agreed sample frame was developer services contacts for all water companies, from November 2023. Of the 10,753 records in total, 1,637 were identified as being related to SLPs/NAVs. Records were deduplicated and merged to produce 236 unique records (204 of which were usable once DNRs had been removed).

41 interviews were achieved, a response rate of 20.1%.

Participants were initially asked to give a CSAT score for one water company based on one specific interaction (the current D-MeX approach) and were then asked to what degree they did score based solely on that interaction (vs. scoring more generally). 19 of the 41 said that they had scored generally, whilst 10 stated that they scores based on the specific interview (with 12 scoring based on a mix of specific and general considerations).

Where the participant had further interactions *with the same company* (38 of the 41), they were asked to give CSAT scores for up to 4 further interactions. Interestingly, when asked the same question about whether they had scored based on the specific interactions or more generally, now almost 50% (16 of the 38) scores specifically, and only 26% generally (with the remaining c25% citing a mix). This implies that in fact the exercise of scoring multiple interactions itself improves how specific the collected scores are.

Where participants had further interactions *with other companies* (12 of the 41), they were asked to give a single CSAT rating for each additional company, based on a single specific interaction. Whilst 12 is too low a sample size to draw statistically reliable conclusions, all were able to provide scores (3 driven by the specific case, 7 by general feelings, and 2 by a mix). These participants were also asked how easy they had found it to score different developer services teams; 7 stated it was easy or very easy, 4 stated it was neither easy nor difficult, whilst only 1 said they found it difficult.

These results lead to the conclusion that there *is* likely to be a benefit to having different surveys for different groups, that participants (NAVs and SLPs at least) *are* able to give feedback not only related to more than one company but also to more than one interaction with any given company, and that this leads to a potential solution to the problem of increasing the prevalence of SLPs, NAVs, and large developers (subject to an agreed definition being established¹), in the survey data.

¹ Pilot participants were also asked if they felt they were a large developer or not, and why, and the responses given were as varied as those collected from water companies and other stakeholders.

2 INTRODUCTION

2.1 Background

Ofwat is the economic regulator of the water and wastewater sectors in England and Wales. Ofwat work with a broad range of stakeholders including the UK Government, the Welsh Government, water companies, consumer organisations and other regulators.

At PR19 Ofwat introduced the customer measure of experience (C-MeX) and the developer services measure of experience (D-MeX) with the aim of improving the experience of residential customers and developer services customers, including house builders, self-lay providers, and new appointees.

At PR19 Ofwat also introduced corresponding bespoke performance commitments for Dŵr Cymru and Hafren Dyfrdwy, focused on the experience of business customers in Wales, the vast majority of which cannot choose their retail services provider (the business customer experience in Wales measure).

The purpose of these measures is to incentivise water companies to provide an excellent customer experience in England and Wales for residential, developer services and business customers respectively.

PR24 will retain C-MeX, D-MeX and Business Customer Experience in Wales. Ofwat agreed to review the effectiveness of these mechanisms before consulting on their proposals for the next price review period (from April 2025 to March 2030) in the PR24 draft determinations and publishing the final decision in the PR24 final determinations in December 2024.

As part of this review, Ofwat have engaged with stakeholders on the overall design of the mechanisms through a range of workshops and meetings, including water companies, customer groups and representatives of property developers and competitors in the new connections market.

Based on this engagement a number of key proposals have emerged which include broadening the operational incidents component in C-MeX, increasing the weight of the survey component in D-MeX and adding a contact component for business customer experience in Wales.

Following this, for PR24, Ofwat are reviewing surveying methods to:

- reflect significant changes in customer behaviour;
- ensure robust and reliable results;
- take account of stakeholder feedback and other research projects;
- future proof the approach up to 2030.

Ofwat commissioned Accent to conduct this research, that will inform the survey designs for the three measures of experience for the PR24 period.

2.2 Objectives

The research objectives were to:

- provide advice and make recommendations on the best approach to capturing customer views for PR24;
- pilot a range of methods for customer surveys that can be delivered in a robust and fair way, and which can be sustained through to 2030;
- test understanding of the questionnaires with the relevant customer base; and
- make recommendations for the delivery of the surveys for the next price review period, covering methodology, frequency, categorisation, different customer groups and other areas of interest.

2.3 Scope of this Report

The scope of this pilot is restricted to the survey component of the D-MeX measure (previously known as the qualitative component).

Reports covering the literature review, and the pilots of the C-MeX and Business Customer Experience in Wales surveys, are published separately.

3 Developer Services Measure of Experience (D-MeX)

3.1 Background

D-MeX is the Developer services measure of experience and was introduced at PR19 with the aim of improving the experience of developer services customers (including house builders, self-lay providers (SLPs) and new appointees (NAVs)).

D-MeX is a financial and reputational incentive mechanism designed to incentivise water companies to provide excellent customer experience for their developer services. The mechanism uses customer satisfaction surveys to give each company a score.

Each company gets an annual D-MeX score based on two components:

- Survey component, previously known as the qualitative component (50%) – based on 0-10 ratings from a monthly customer satisfaction survey; and
- Quantitative component (50%) – based on how a company performs against a set of Water UK metrics.

D-MeX is being continued in PR24, for the period 2025-30. The scope of this pilot is restricted to the survey component of the D-MeX measure.

3.2 Objectives

The primary objectives of this pilot included:

- Reviewing existing data to assess features of the current methodology, including:
 - Sample sizes and robustness of the data
 - Sustainability
 - Operational Deployment
 - Suitability of the questionnaire as a survey instrument
 - How to extract more from the data
 - Cost effectiveness
- Assessing the feasibility of either actually, or effectively, increasing the prevalence of SLPs, NAVs, and large developers, in the data.
 - Defining large developers, SLPs, and NAVs
 - Reviewing existing data for opportunities in this area
- Addressing a number of additional research questions
 - Are there benefits to having different surveys for different customer groups, for example having a different survey for one-off connection customers vs those which transact multiple times with multiple different water companies?

- Is it feasible to survey customers about their experience with more than one water company in one survey?
- Running a physical pilot to test methodological changes suggested for the next price review period

3.3 Existing Data Review

Sample Sizes and the robustness of the data being produced

The current sample size for D-MeX set at 20% of eligible contacts, with a minimum of 100 and a maximum of 1,000 interviews, per company, per year.

The data collected for Year 3 of the current D-MeX measure (April 2022 to March 2023, the most recent completed year at time of print) was utilised for analysis.

Standard errors of the mean Customer Satisfaction (CSAT) scores were calculated, and 95% confidence intervals were added. Bases are also added, and in many cases (smaller companies) this represents the maximum possible number of interviews achievable in the year.

Table 1. Year 3 D-MeX Mean CSAT scores with confidence intervals

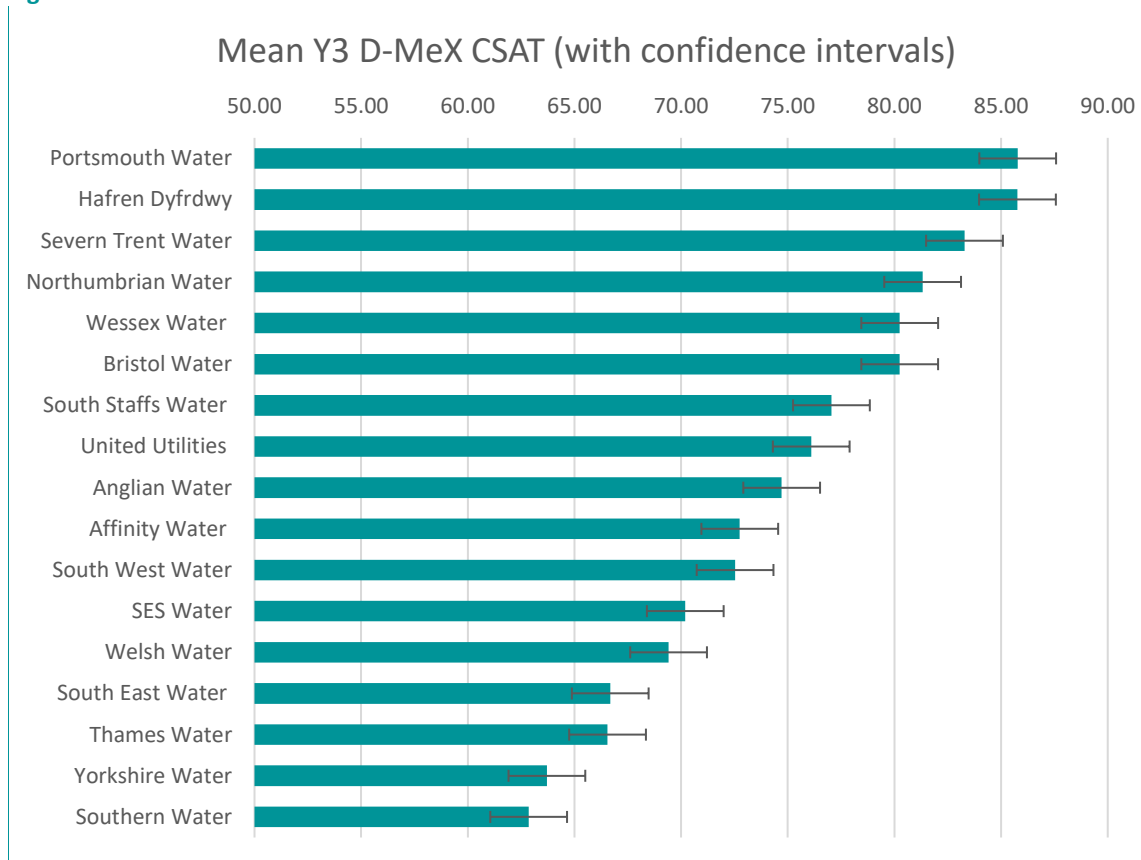
Water company	Mean CSAT Score	95% Confidence Interval; standard error of mean	Lower Confidence Limit	Upper Confidence Limit	Base
Portsmouth Water	85.78	2.90	82.88	88.68	109
Hafren Dyfrdwy	85.77	3.33	82.44	89.10	104
Severn Trent Water	83.29	1.29	82.00	84.57	924
Northumbrian Water	81.32	1.35	79.97	82.68	902
Wessex Water	80.25	1.64	78.61	81.89	604
Bristol Water	80.25	2.10	78.15	82.35	404
South Staffs Water	77.05	2.88	74.16	79.93	275
United Utilities	76.10	1.66	74.44	77.76	963
Anglian Water	74.72	1.64	73.07	76.36	1008
Affinity Water	72.75	1.98	70.77	74.73	653
South West Water	72.53	1.67	70.86	74.20	999
SES Water	70.20	3.91	66.29	74.10	205
Welsh Water	69.41	1.74	67.67	71.15	989
South East Water	66.68	2.28	64.40	68.96	591
Thames Water	66.55	1.78	64.78	68.33	1008
Yorkshire Water	63.71	1.93	61.78	65.64	1009
Southern Water	62.85	2.06	60.79	64.91	747
Average		2.13			674

Confidence intervals were on average 2.13. This means that we can have 95% confidence that the actual mean is the value stated plus or minus 2.13 (on average). These are acceptable margins for research of this type.

Due to the wide variation in sample sizes (a factor of both the selection methodology and also the number of eligible contacts each water company received), the range of confidence intervals varied widely. From as low as c1.3 for some of the larger companies, to as high as almost 4 for the smallest.

The mean scores and corresponding confidence intervals can be graphically displayed as in Figure 1.

Figure 1. Year 3 D-MeX Mean CSAT scores with error bars



Exploring this further, we can consider the range of D-MeX scores that exists (as shown in Table 1). We observe a top score of 85.78, and a bottom score of 62.85; a range of 22.93 points. Divided by 17 companies, this gives us an average gap between companies of 1.35 points.

Speaking broadly, an increase in sample size of 200 for any given company, reduces their confidence interval (CI) by roughly 0.3 and their ranking position variability by around 0.15.

As previously stated, these confidence intervals are calculated on unweighted bases of a maximum of 1,000 interviews per company. As sample sizes change, so do the confidence intervals (the greater the sample size, the narrower the band). The widest confidence intervals (Hafren Dyfrdwy and Portsmouth Water) are still at an acceptable level.

Sustainability

While the methodology has continued for four years unabated, there are questions around its ongoing sustainability. During this period, there has been a steady increase in “Do not contact” (DNC) instruction coming from companies that feature regularly on the sample (i.e. large developers and SLPs/NAV), which compromises the sustainability of the survey.

The current approach of excluding participants interviewed in the relevant exclusion period can be ineffective due to practical difficulties with effectively deploying it. For example, when the same end customer come from different water companies with differently spelt/formatted company names, different contact phone numbers, different email addresses, and different (or differently spelt) contact names, it is not feasible to effectively cross reference either manually or automatically.

It is vital that this factor is not ignored in any drive for increasing sample sizes of any customer sub-groups such as large developers, SLPs, or NAVs. Already c20% of sample records are currently lost to DNC requests, and once in the list the end customer stays there forever.

Operational deployment

Methodology

Practically speaking little choice exists regarding the methodology, since the purpose is in many cases to maximise a small sample size. Even including the positive effects of reminder emails, email surveys typically have a far lower response rate than CATI (all else being equal), so were email contact data supplied (which it currently is not), a change of method in this direction is likely to reduce rather than increase response rates (even with the introduction of an incentive potentially). Other modes (such as SMS or CAPI) are inappropriate for various reasons.

Frequency of survey

The nature of the telephone component of the research means that it lends itself very well to being used as part of an ongoing tracker (with fieldwork in deployment constantly, albeit working towards quarterly quotas).

Conducting the sampling monthly appears to be a good balance between the administrative burden (for both water companies and the agent) associated with generating, processing, and deploying the sample, covering a sufficiently long period to allow deduplication to be effective, and not allowing too long between the initial contact and the interview approach.

Quotas/representation

The current sample selection and quota setting rules are as follows:

“The sample size for most companies will be based on 20% of those customers that transact with the company in a month. For each company, where 20% of the sample is expected to be more than 1,000 in a year, the sample will be capped at 1,000; where 20% of the sample is expected to be less than 100 in a year, the sample will be increased to at least 100 in a year.”

The practical implementation of this is:

- That the largest companies have a quota set of 1,000 interviews per year, divided down to 83 per month, regardless of the number of contacts they have received
- That the smallest companies in theory have their sample sizes maximised (to at least 100 interviews)
- The mid-sized companies have their sample size constrained to 20% of the number of transactions

The result is that there is something of a vacuum created in the mid-sized companies, whereby they do not hit 1,000 interviews, and the sample is not maximised. More interviews could be achieved with these mid-sized companies; however, we would remind of the cautionary note given in the “sustainability” section of this report.

Sample

The primary difficulty with the survey as is it, is finding the correct participant and focussing them in on the correct transaction. The ease with which this can be done depends on the sample quality. If there is no contact name on the sample (or the name of someone who did not handle the transaction), it can be very difficult.

Equally, if the sample is missing data that could be used to help the participant focus on the correct transaction, the task of interviewing is more difficult. As we cover later in this section there are also several “optional” sample fields that water companies do not routinely fill in, but would be beneficial to receive, to ensure that the initial approach is most effective.

Cost Effectiveness

Because of the relatively high levels of engagement amongst developers generally, response rates are good. Even though a relatively costly methodology is used (CATI, as opposed to email) to maximise the sample, as businesses are typically more difficult to interview than consumers, the CATI interviews achieved are good value for money.

Suitability of the questionnaire as a survey instrument

Aside from the issues previously covered (with locating the participant and focussing on the correct interaction), the current survey generally works well. Terminology is well understood, recall adequate, and the survey sufficiently short (on average, around 6.5 minutes) to prevent any issues with dropouts/partial completion. There is insufficient evidence available to assess whether it is sufficiently long to discourage future participation or not. There is scope to improve though, including ensuring that the information pertinent to identifying the relevant transaction is presented to the participant at time of collecting the key CSAT score.

How to extract more from the data

The questionnaire in use for the D-MeX survey collects a customer satisfaction score and additional data including what the company did well, what they could do better, what the top three things the participant is looking for in terms of service delivery, and what one thing would make the developer services team easier to deal with.

Further, there is a bank of 11 questions rating various service attributes (ranging from ease of contact, and meeting deadlines, to the quality of information on the website).

The large bank of questions does lend itself to advanced analytics such as CHAID (Chi-square Automatic Interaction Detector). Currently, this advanced water company level analysis is not within the scope of the D-MeX survey outputs, but survey data is provided to water companies to enable them to conduct their own detailed analysis on key drivers of customer satisfaction.

3.4 Increasing the prevalence of SLPs, NAVs, and large developers in the survey data.

One of the emerging themes from Ofwat’s consultation on the measures of experience performance commitments at PR24² for D-MeX is the proposal to increase the weight of competitors (Self-lay providers (SLPs), and New Appointees and Variations (NAVs)) and large developers within the survey.

Particular attention was given to answering the primary question of whether this is feasible or not.

With the emphasis of the pilot firmly on increasing the prevalence of SLPs, NAVs, and large developers, the first stage was to define the three entities.

The second stage of the pilot was to review the data built up over the last 4 years of conducting the D-MeX survey fieldwork, to assess whether a measure as simple as increasing the quota of these participant types, or altering the sampling approach, could meet the objective of increasing their weight in the survey.

The third stage of the pilot was to assess the feasibility of increasing the amount of information obtained from each interview with a participant of this type (through redesigning the questionnaire) – something that was tested in a physical pilot of fieldwork.

Defining SLPs, NAVs, and Large Developers

Accent reviewed the responses to the Ofwat consultation on defining “large” developers. These responses revealed a range of perspectives, suggesting that further work needed to be done before reaching a definition for use in the pilot.

Accent also reviewed the content of the provided samples, for a potential solution. Whilst there are sample fields that stood out as being possibilities for identifying large developers (e.g., total number of phases, plots, number of mains connections, customer type, number of transactions), they are either so inconsistently populated by water companies as to make them unusable for this purpose, or the data that is contained within them is of little definitive use.

Table 2. Percentage of records containing the designated information (example data set: November 2023):

Water Company	Total number of phases	Plots	Customer type	Number of mains connections	Number of transactions
Affinity Water	0%	85%	100%	0%	0%
Anglian Water	0%	100%	100%	0%	0%
Bournemouth Water	0%	100%	100%	0%	0%
Bristol Water	0%	90%	98%	2%	100%
Essex & Suffolk Water	0%	95%	100%	0%	0%
Hafren Dyfrdwy	32%	100%	100%	0%	0%
Northumbrian Water	0%	95%	100%	0%	0%

²<https://www.ofwat.gov.uk/consultation/consultation-measures-of-experience-performance-commitments-at-pr24/>

Portsmouth Water	0%	90%	0%	0%	0%
SES Water	0%	99%	60%	53%	53%
South East Water	0%	31%	100%	0%	0%
South West Water	0%	86%	100%	0%	0%
Southern Water	0%	0%	0%	0%	0%
South Staffs Water	0%	93%	0%	0%	0%
Severn Trent Water	15%	99%	98%	0%	0%
Thames Water	0%	100%	74%	0%	0%
United Utilities Water	0%	72%	100%	14%	100%
Dwr Cymru Welsh Water	0%	100%	100%	19%	100%
Wessex Water	0%	19%	100%	0%	0%
Yorkshire Water	0%	100%	100%	0%	0%

Whilst “Plots” is populated as a field in the majority of cases, the wide range of values contained (from 1 to over 15,000) combined with the difficulty in setting a level which would be universally accepted as being appropriate, it was decided not to use this field. “Customer Type” is also a well populated field, but none of the designations (which included such data as “End customer”, “Agent”, “Intermediary”, “Consultant” etc.) related to the size of developer so were useful for the intended purpose. Numbers of phases, mains connections, or transactions were disregarded due to scarcity of useful data (and again difficulties with defining what constitutes a large developer). It should be noted that populating these sample fields is optional. Therefore, it was concluded that there is insufficient data contained within the sample files to allow a sample driven definition to be tested.

For these reasons, we were neither able to reach a definition of ‘large developers’, nor test measures aimed at targeting this group, through this pilot.

The definition of SLPs and NAVs is much clearer, both conceptually (in terms of an agreed definition existing) and practically (in terms of identifying them on the supplied sample) speaking.

The historic sample gave two distinct directions for defining SLPs and NAVs.

- Using SLP and NAV specific metrics available in the “Metric” column of the data files
- The information in the “Customer Type” column of the sample file (which did contain both SLP and NAV designations alongside the others previously mentioned).

It was decided that, for the purposes of the pilot, if a sample record matched on either of these variables it was treated to mean that they would be categorised as such in the pilot.

Review of Existing Data (focus on large developers, SLP, NAV)

The current methodology for D-MeX involves water companies passing over monthly records relating to all interactions their developer services teams have had with customers in the preceding month. These records are deduplicated and prepared for fieldwork. Where sufficient records exist to ensure the minimum sample size for the month a random selection of records is taken for surveying. Where insufficient records exist, all sample are used for fieldwork. All fieldwork is conducted by CATI.

The effects of the current random method of sample selection were reviewed, to identify the prevalence of SLP or NAV sample records that had not historically been part of the random selection. The objective was to assess whether adding a prioritisation to the sample selection routines could increase the prevalence of SLPs and NAVs (and later, should consensus on the definition be agreed and reflected on the sample, large developers) in the data. It was found that had all SLP and NAV sample been selected, and none left out due to being surplus to the standard sampling requirements, that the volumes of SLP and NAV records in the final sample would have been roughly 6.5% higher than it had actually been.

It is clear then, that the inclusion of these additional records will have very little noticeable impact on increasing the weight of SLPs and NAVs from the current level of 0.8% of D-MeX interviews (the impact likely to be measured in the hundredths of a percent).

Another option for increasing the weight of SLPs and NAVs in the data, is to apply weights at the data processing stage, as D-MeX scores are calculated. However, with SLPs and NAVs making up under 1% of interviews, and sample sizes capped at 250 per quarter per company, it's clear that on average very few NAV and SLP interviews (c2 per company per quarter) exist, and there are risks associated with weighting up very small sample sizes.

In summary, it can be observed that the major difficulties preventing the increase in prevalence of large developers, SLPs, and NAVs, are the difficulty with identifying the first of those groups, and the paucity of sample of the latter pair. The current feature of the methodology that does work very well, is the clear identification of SLPs and NAVs on the sample (meaning we can target them, and effectively track outcomes).

3.5 Additional research questions

Two additional research questions had been asked, which fall within the scope of this pilot;

- Are there benefits to having different surveys for different customer groups, for example having a different survey for one-off connection customers vs those which transact multiple times with multiple different water companies?
- Is it feasible to survey customers about their experience with more than one water company in one survey?

It was agreed that the way to answer these questions was through fieldwork.

3.6 Fieldwork

Methodology

For the pilot the same methodology was used as for the current D-MeX survey, namely CATI. To avoid any interruption or interference with the live D-MeX measure it was agreed that the November 2023 D-MeX sample (already with Accent, and fieldwork related to which was already finished) would be the 'freshest' and its use would not cause any interference with the operational measure. By using existing sample files, Accent had direct access and no further requests or action from the water companies was required.

Sample Frame

As described previously, the D-MeX sample files supplied by water companies in November 2023 to Accent (via Accent's portal) were used and processed.

The sample was processed in the following way:

- D-MeX records from all water companies submitted to Accent in November 2023 were combined to generate a dataset of 10,753 records.
- Out of these records we were able to identify 276 as NAVs and 1,361 as SLPs, resulting in 1,637 records that could be included in the pilot.
- Duplicate records were matched and merged both **within** and **across** different water companies to identify individuals that could be in position to answer about multiple contacts.
- Sample matching resulted in a condensed datafile of 236 unique records that were included in the pilot. 32 were removed as Do Not Recontact, leaving 204 to use.

Approach/questionnaire design

It was originally envisaged at inception stage, that the questionnaire would be redesigned to allow data related to multiple transactions to be collected, that this questionnaire would be cognitively tested with up to 10 participants, and then piloted with up to 100 participants.

Once it had been established that targeting large developers was not feasible, and that NAV and SLP sample distilled down to only 204 records, the approach was changed.

Due to the reduced sample size, a far more qualitative approach was taken with participants than had been initially envisaged. At the start of the interview, participants were informed that this pilot was being undertaken to consider how best to engage with organisations such as theirs, that it was an opportunity for their views to be taken into account, and that the survey length would depend on the number of interactions, and it could last anywhere from 5 to 25 minutes.

The questionnaire also contained several diagnostic type questions where participants were prompted for suggestions of what they think could be done to make their interview experience easier or if there are elements that make taking part in the survey more difficult. A question asking these organisations whether they considered themselves to be "large" or not, was included to potentially add some perspective to the dialogue about how best to categorise developers generally in this respect.

Fieldwork

The fieldwork was carried out between 26th February and 11th March 2024 through Accent's Telephone Unit.

A total number of 41 interviews were achieved, a 20.1% response rate.

Findings

The principal focus of the survey was to assess participants' ability to (in this order):

- score any given water company based on one specific interaction with a water company
- score that same company on further interactions
- score additional water companies based on one specific interaction with them

How able were participants to score a specific interaction (vs. scoring more generally)?

In the first instance, participants were asked to consider a specific interaction with a specific water company and asked to score their satisfaction with it. This mirrors the way scoring of interactions happens in the current D-MeX survey.

We then looked at how participants being asked about one interaction gave their score:

Q: Would you say the responses you have just given were driven by the specific interaction we asked you to focus on, or by your general experience of/with the developer services team at Water Company?

1. Entirely driven by this specific interaction
2. Mostly driven by this specific interaction
3. Driven by both equally
4. Mostly driven by general experiences
5. Entirely driven by general experience

10/41 participants said they scored the water company entirely or mostly based on the specific interaction, 12/41 were driven by both equally, and 19/41 entirely or mostly based on the general experience.

Although the evidence does not exist to support the assertion (because this question has never been asked on the live D-MeX measure), it seems reasonable to assume that the figures found here (c25% scoring based on the specific interaction, c50% based on general experiences, and c25% being driven by both equally), reflect answering behaviour within the live D-MeX survey.

We recognised that the interactions we were asking participants to score happened in November and participants were surveyed in February (a fairly large gap, that may have reduced recall). Therefore, all participants that scored entirely or mostly based on the general experience were asked a follow up question about if recency would have helped them assign a score in a more event driven way. 5/19 said asking about a more recent interaction would have helped them score the water company on a more interaction driven basis.

Were participants able to score different interactions separately/individually with an individual water company?

Secondly, we looked at participant's ability to score a water company when they had multiple interactions with them. 16/38 participants stated that they scored the water company entirely or mostly based on the specific interactions, 12/38 were driven by both a specific interaction and general experience equally, and 10/38 scores entirely or mostly based on the general experience.

It is notable that the percentage of participants scoring based on general experiences dropped from c50% to c25% at this point, that participants being led by a combination of specific and general experiences remained steady at 25%, and the proportion of participants stating that they had scored based on the specific interaction doubled from 25% to 50%. This implies that not only is it feasible for participants to consider multiple transactions at once, in fact it enhances their ability and/or willingness to score each transaction on its own merits (far more so than when asking about only one transaction, as is currently the scenario for D-MeX interviews).

Length of time since these transactions took place was again addressed - only one more participant said they would have scored the water company on a more interaction specific level had they been asked closer to the interaction, reinforcing the point that focussing on multiple transactions aided recall.

Were participants able to score different water company developer services teams within one interview?

Thirdly, we looked at participants' ability to score interactions with multiple companies, and the ease with which they were able to distinguish between different developer services teams.

All participants that had eligible transactions across more than one water company were willing and able to provide scores for additional water companies.

The numbers of interviews in which this scenario was tested is very low (too low for statistically reliable inferences to be drawn). Nevertheless, qualitatively, and indicatively the results are interesting and worth having considered.

3 of the 12 participants were able to score the water company entirely or mostly based on the specific interaction, 2 were driven by both a specific interaction and general experience equally, and 7 were driven entirely or mostly based on the general experience.

This suggests that whilst the scores given for multiple water companies within one interview appears to be more general than when multiple interactions with the same water company are scored, they are still given broadly in similar terms to the responses currently being collected in the D-MeX measure.

In addition, the relevant participants were also asked the following question about their ease in distinguishing between different developer teams at water companies:

Q: How well were you able to consider the various developer services teams, and score them individually?

1. Very easy to score each team individually
2. Easy to score each team individually
3. Neither easy nor difficult to score each team individually
4. Difficult to score each team individually
5. Very difficult to score each team individually

7 of the 12 found distinguishing between teams easy or very easy, 4 found it neither hard nor difficult and only 1 participant found it difficult.

The impact of metric or site considerations on ease of distinguishing interactions

As a further level of analysis, we considered whether participants answering about the same metric and/or the same site were able to score water companies down to the specific interaction or whether they were likely to score based on general experience with the relevant developer services team.

25 participants were asked to score a water company about the same metrics at least twice. The number of times any given participant was asked about the same metric varied from two to six times. Here we found that 12/25 relevant participants scored the water company entirely or mostly based on the specific interaction; 9/25 did so based on both specific and general interaction and a further 4/25 on a mainly general experience. This reflects the split of the sample as a whole fairly well (implying that scoring multiple interactions relating to the same metric did not dull participant ability to differentiate between them).

27 participants were asked to answer about different interactions related to the same site/address at least twice. The number of times any given participant was asked about the same site/address varied from two to four times. Here we found 10/27 participants scored the water company entirely or mostly based on the specific interaction; a further 9/27 did so based on both specific and general interaction and 8/27 mainly or entirely on general experience. This does imply that participants found it more difficult to separate transactions (of different metrics, or of different plots) at the same site, from each other.

Again, sample sizes were low in these two scenarios so results should be interpreted with caution.

What can we do to help participants score water companies on specific interactions?

Participants were asked if there was anything that made answering the questions difficult.

The general consensus was that there wasn't anything difficult, but some had suggestions:

- *"It would be better if there was a reference number **rather than a date** as we deal with that many things with X water."*
- *"I keep detailed records of each site. I go into that specific project, if it's just about feedback on a specific water company not a site that would be a more generalised response."*

When asked if there was anything that helped them answer about interactions, participants listed a number of things:

- *"The date, the roads and the site names. Being **consistent with site names** helps."*
- *"The **dates and the type of interaction**, help trigger the memory."*
- *"When it references how many **plots**."*
- *"To get **postcode** so I can work out the scheme because we call them something slightly different."*

Additionally, when asked if there would be anything else that would help, the general consensus was that a reference number would be very useful:

- **“Costs potentially, time it took to process, *date of submission.*”**
- **“If it had been described as a completed PoC scheme I wouldn't have gone on their portal and gone on the live but it was completed so if it was described more practically. A *reference number* would have helped me to find it easier.”**

What can we learn from how NAVs and SLPs define being “large”?

During this pilot we have asked SLPs and NAVs to self-define as “large” organisations (or not). The hugely varied responses mirrored findings from the consultation.

- 17/41 participants **have** defined themselves as large

Figure 2. Number of employees

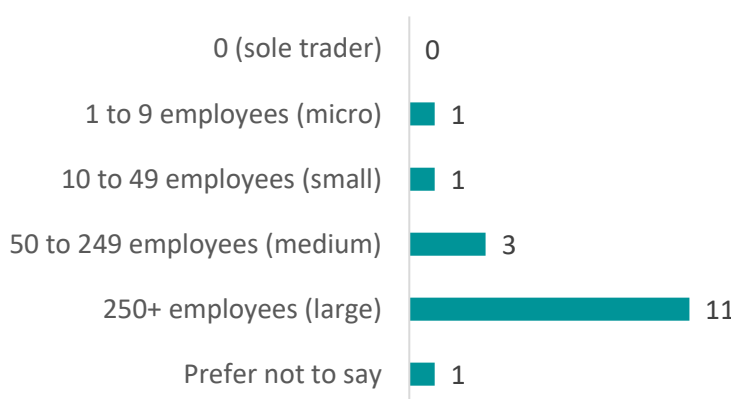
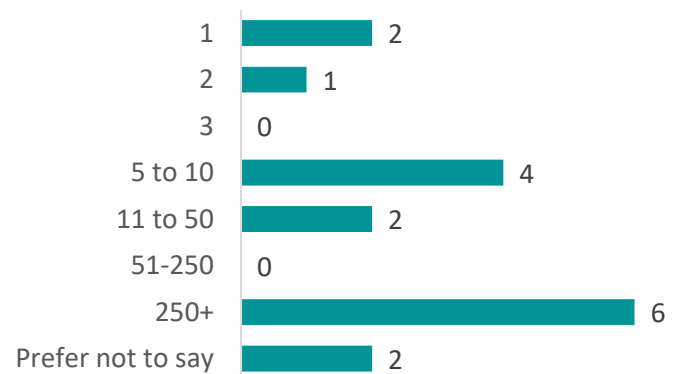


Figure 3. Number of UK sites



- 24/41 participants **have not** defined themselves as large

Figure 4. Number of employees

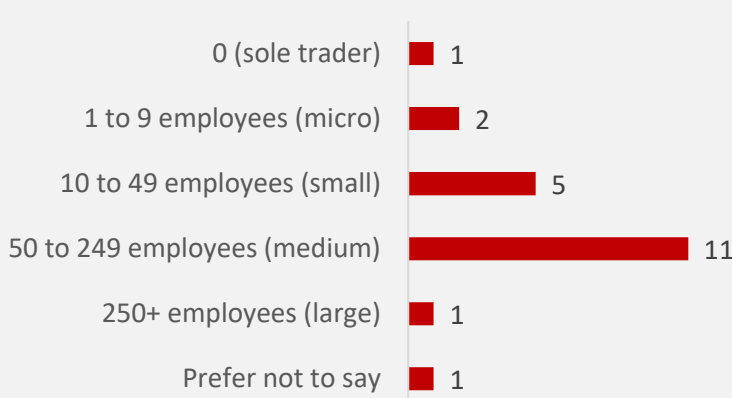
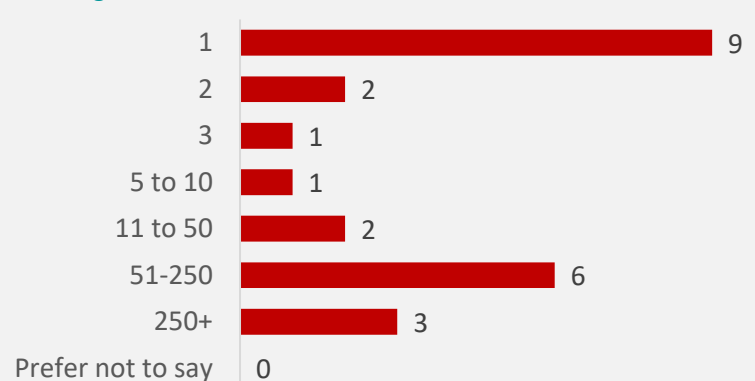


Figure 5. Number of UK sites



Here we see organisations with very few staff and with very few sites, categorising themselves as being large, alongside organisations with >250 employees and >250 UK sites *not* describing themselves as being large.

Participants were also asked why they categorised themselves as they did, and the responses are further explored below.

Organisations say they are large because:

- *“We are a NAV and we've got a lot of **employees**.”*
- *“The number of **connections** we carry out for and on behalf of X Water.”*
- *“A lot of **divisions**, owned by two big organisations.”*
- *“The **regions** that we cover, the whole of Great Britain.”*
- *“We do around 2000 **applications** a year, and we're the biggest in the UK.”*
- *“In relation to NAV we would represent the **most significant player**.”*
- *“We deal with numerous **multiple building companies** to lay their mains and utilities and designs.”*
- *“We're a NAV, developers come to us, the size of company and **amount of sites** we have. The number of sites is the total of company offices and current project sites.”*

Organisations say they are not large because:

- *“We have under 50 **staff** it can be medium to small development, it is about 100 houses max it does not do 1000's that's why I class it as small to medium.”*
- *“A medium developer, we produce 250 **plots** a year.”*
- *“I don't know, I am new to the business I am going by what everyone else around me says, it used to be a family-run business it is **growing** but is not huge.”*
- *“We're medium we're not a big or small company, we're an SLA not a developer.”*
- *“Because we do **work with large developers**, but we are a NAV incumbent.”*
- *“We're on the smaller side, based on how many **schemes** we get appointed in a year.”*
- *“We're quite specialist and we have less than 50 **employees**.”*

Summary

Returning to the primary research objectives of establishing whether it is possible to implement a mechanism for increasing the prevalence of large developers and competitors in the survey data, of whether there are benefits associated with having different surveys for different groups, and of whether it is feasible to survey customers about their experience with more than one water company in one survey the answer is certainly “yes” on all counts.

However, that is not without caveats. It was quickly established that it is not possible to materially increase the prevalence of SLP and NAV interviews by changing sampling approach and/or doing more interviews. Very few additional records that were available for selection had not been selected, and the sample that had been selected had been exhausted.

It was further established that work remains to be done on defining a large developer (if specific measures are to be taken to increase their prevalence in the data). It needs to be practicably feasible for water companies to effectively and consistently recognise developer services transactions as being related to a large developer, and this data needs to be relayed to the research agency consistently and simply within the D-MeX sample.

SLPs and NAVs were much easier to consider, due to an agreed definition existing (and transactions related to organisations of this type being clearly identified in the monthly sample sets).

The concept of processing the sample such that participants who appeared repeatedly on the sample (within and across water companies) were collated and interviewed about multiple transactions (with the suggestion being that each score is counted in its own right towards the relevant water company – meaning one interview could in theory give 5 scores towards a primary water company, and 1 each to a further 5), worked (although again not without caveats).

The sample *can* be manipulated to capture multiple interaction both within water companies and across water companies – but it is a manual and imprecise process.

SLP and NAV participants *were* able to complete the survey (and answer about multiple issues) without major issue, but in the process of understanding this a light has been shone on how participants may currently be answering to the single rating question.

Whilst it seems reasonable to assume that large developers (which in some respects are similar to these larger SLPs and NAVs), would be able to complete the survey in a similar way.

It should finally be noted, that whilst a definition of a large developer has not yet been agreed, it is possible that the question can be circumvented. By merging (and multi-counting) numerous interactions for all supplied sample records (be they SLP, NAV, Large developers, or anything else), a fieldwork design which enables organisations dealing with multiple transactions to be represented more heavily in the data would be created.

4 Appendices

4.1 Appendix A - Questionnaire

Good morning/afternoon/evening. My name is and I am calling from Accent on behalf of Ofwat, the economic regulator for the water sector in England and Wales. Could I talk to #ID#?

IF NAME ON SAMPLE SAY. I understand that you have dealt with #WATER COMPANY# in November, is that right?

IF NO NAME ON SAMPLE SAY. I understand that someone on this number has been in touch with #WATER COMPANY# in relation to a #WATER UK METRIC# – this was in November, for #SITE#. Could I talk to that person please? **IF TALKING TO CORRECT PERSON, CONTINUE. OTHERWISE ASK TO BE TRANSFERRED OR MAKE APPOINTMENT TO CALL BACK**

WHEN TALKING TO RIGHT PERSON: Accent are working with Ofwat to pilot a mechanism for better engaging with New Appointments and Variations (NAVs) and Self Lay Providers (SLPs), on the D-MeX measure (the measure Ofwat uses to measure the levels of service the various water company developer services teams provide to customers). There is a potential concern at the moment that the views of NAVs and SLPs may be being drowned out by the responses given by one-off development responses, and Ofwat would like to see if there is any way of increasing the data being collected from SLPs and NAVs, whilst making the survey more SLP and NAV friendly.

INTERVIEWER, CONTACT WAS AS FOLLOWS – pls use as necessary: #METRIC# – this was in #MONTH#, for #SITE#.

ALL: You may well have answered one of our D-MeX surveys before. Today we would like to ask you to help us develop an effective way of engaging with you, about multiple interactions at once (both within water companies, and if applicable, across companies).

Depending on how many contacts we have on record for you in the month of November, and how much you have to say about the survey design, the survey will take between 5 and 20 minutes to complete.

Accent abides by data protection laws at all times, and your responses will be treated in the strictest confidence unless you agree otherwise.

You can find out more information about Ofwat’s surveys and what is done with the information that is collected in the Privacy Policy on Ofwat’s website **Read out if necessary:** <https://www.ofwat.gov.uk/publication/privacy-policy>

Please note that this call may be monitored or recorded to verify accuracy or for training purposes.

IF NECESSARY:

- You have been selected from customers dealing with developer services at #Water Company Name#, in #MONTH#.
- Your contact details have been provided to us by #WATER COMPANY#

Can I confirm that you are happy to participate in the survey?

Yes

No **THANK & CLOSE**

INTCHECK. INTERVIEWER: PLEASE CONFIRM YOU HAVE ADVISED THE PARTICIPANT OF:

Calls being recorded

INTCHECK2. INTERVIEWER: PLEASE CONFIRM YOU HAVE ASKED AND CHECKED THAT THE PARTICIPANT IS **NOT** TAKING THE INTERVIEW ON A MOBILE DEVICE WHILE DRIVING OR OPERATING EQUIPMENT

Yes, it is safe for the participant to proceed

No, it isn't safe – we need to call back later **GO TO APPT SCREEN**

Main questionnaire

Q1 First we would like you to think about the following specific interaction:

Water company	Water UK metric	Date completed	Site address
---------------	-----------------	----------------	--------------

Taking everything into account, how satisfied or dissatisfied are you with how #WATER COMPANY# handled this particular interaction? Please use a scale of 0-10, where 0 = extremely dissatisfied, 5 = neither satisfied nor dissatisfied and 10 = extremely satisfied.

Extremely dissatisfied					Neither satisfied nor dissatisfied					Extremely satisfied
0	1	2	3	4	5	6	7	8	9	10

Q2 Why did you give that score? **PROBE FULLY**

.....

Q3 Would you say the responses you have just given were driven by the specific interaction we asked you to focus on, or by your general experience of/with the developer services team at [WATER COMPANY]?

Entirely driven by this specific interaction	Mostly driven by this specific interaction	Driven by both equally	mostly driven by general experiences	Entirely driven by general experience
1	2	3	4	5

Q4 **ASK IF Q3 = 4 or 5 ONLY** We appreciate that the interaction we asked you to focus on, was a long time ago (in November). Had we asked you about a more recent transaction (for example one in January), how much would you say your response might have been driven by the specific interaction we asked you to

focus on, as opposed to your general experience of/with the developer services team at [WATER COMPANY]?

Entirely driven by specific interaction	Mostly driven by specific interaction	Driven by both equally	mostly driven by general experiences	Entirely driven by general experience
1	2	3	4	5

Q5 **ONLY ASK IF "WATERCOMPANY 2" IS NOT NULL** Thank you. Our records indicate that you had other interactions with [WATER COMPANY] in November.

Whilst it may be difficult, we would like you to, as far as you can, tell us how satisfied or dissatisfied are you overall with how [Water Company] handled each of the following interactions using the same scale where 0 is extremely dissatisfied and 10 is extremely satisfied?

	Display fields from sample				Response Fields				
	Water company	Water UK metric	Date completed	Site address	Satisfaction Rating	Can't remember	Don't know	Why did you give that score?	Notes
					(0-10 only)	(tick box)	(tick box)	(Open)	(Open)
2	Water company 2	Water UK metric2	Date completed 2	Site2					
3	Water company 3	Water UK metric3	Date completed 3	Site3					
4	Water company 4	Water UK metric4	Date completed 4	Site4					
5	Water company 5	Water UK metric5	Date completed 5	Site5					

Q6 To what degree would you say the responses you have just given were driven by the specific interaction we asked you to focus on, compared to your general experience of/with the developer services team at [WATER COMPANY]?

Entirely driven by these specific interactions	Mostly driven by these specific interactions	Driven by both equally	mostly driven by general experiences	Entirely driven by general experience
--	--	------------------------	--------------------------------------	---------------------------------------

1	2	3	4	5
---	---	---	---	---

Q7 **ASK IF Q6 = 4 or 5 ONLY** Had we asked you about more recent transactions (for example ones in January), then how much would you say your response might have been driven by the specific interactions we asked you to focus on, as opposed to your general experience of/with the developer services team at [WATER COMPANY]?

Entirely driven by specific interactions	Mostly driven by specific interactions	Driven by both equally	mostly driven by general experiences	Entirely driven by general experience
1	2	3	4	5

Q8 Were there any particular features of the interactions we asked you to think about, or the information we used to try to prompt you to remember them, that made them **difficult** to think about individually?

.....

Q9 Were there any particular features of the interactions we asked you to think about, or the information we used to try to prompt you to remember them, that made them **easy** to think about individually?

.....

Q10 Is there any information that the developer services team at [WATER COMPANY] could have given us, that would have helped you to be able to think about the various interactions individually?

.....

Q11 **ONLY ASK IF "WATER COMPANY 6" IS NOT NULL** Thank you. Our records also indicate that you had other interactions with other water company developer services teams, in November. We will only ask you once about each one.

Whilst it may be difficult, we would like you to, as far as you can, tell us how satisfied or dissatisfied are you overall with how each water company handled the following interactions (using the same scale where 0 is extremely dissatisfied and 10 is extremely satisfied)?

<i>Display fields from sample</i>					<i>Response Fields</i>				
Water company	Water UK metric	Date completed	Site address		Satisfaction Rating	Can't remember	Don't know	Why did you give that score?	Notes
					(0-10 only)	(tick box)	(tick box)	(open)	(open)

1	Water company 6	Water UK metric6	Date complete d6	Site6					
2	Water company 7	Water UK metric7	Date complete d7	Site7					
3	Water company 8	Water UK metric8	Date complete d8	Site8					
4	Water company 9	Water UK metric9	Date complete d9	Site9					
5	Water company 10	Water UK metric10	Date complete d10	Site10					

Q12 How well were you able to consider the various developer services teams, and score them individually?

Very easy to score each team individually	easy to score each team individually	Neither easy nor difficult to score each team individually	difficult to score each team individually	Very difficult to score each team individually
1	2	3	4	5

Q13 **ASK IF Q12 = 4 or 5 ONLY** Had we asked you about more recent transactions (for example ones in January), then how well do you think you might have been able to score the various teams individually?

Very easy to score each team individually	easy to score each team individually	Neither easy nor difficult to score each team individually	difficult to score each team individually	Very difficult to score each team individually
1	2	3	4	5

Q14 And to what degree would you say the responses you have just given for each team were driven by the specific interaction we asked you to focus on, vs. by your general experience of/with each of the developer services teams?

Entirely driven by each specific interaction	Mostly driven by each specific interaction	Driven by both equally	mostly driven by general experiences	Entirely driven by general experiences
1	2	3	4	5

Q15 **ASK IF Q14 = 4 or 5 ONLY** Had we asked you about more recent transactions (for example ones in January), then how much would you say your response might have been driven by the specific interactions we asked you to focus on, as opposed to your general experience of/with the developer services team at [WATER COMPANY]?

Entirely driven by specific interaction	Mostly driven by specific interaction	Driven by both equally	mostly driven by general experiences	Entirely driven by general experience
1	2	3	4	5

Q16 Do you have any other feedback for Ofwat about the D-MeX survey, from the perspective of a participant?

.....

Business details

Q17 Thank you so much, now just a few questions about your organisation for classification purposes. Would you consider your organisation to be a “Large” developer?

- Yes
- No

Q18 What makes you say that?

Q19 How many sites in the UK does your organisation operate from?

- 1
- 2
- 3
- 4
- 5-10
- 11-50
- 51-250
- 250+
- Prefer not to say

Q20 How many employees does your organisation have in the UK?

- 0 (sole trader)
- 1 to 9 employees (micro)
- 10 to 49 employees (small)
- 50 to 249 employees (medium)
- 250+ employees (large)
- Prefer not to say

Q21 Thank you for your time and co-operation in completing this survey. On behalf of Ofwat we would like to thank you for your feedback.

Would you be willing to be contacted again if we need to clarify any of the answers you have given today?

Yes

No

Q22 **IF Q21 = YES** Please can we take a note of your name and where we can contact you?

Name: [CATI: DP, IMPORT FROM ID]

Telephone: [CATI: DP, IMPORT FROM TELNUMBER]

Interviewer Confirmation

I confirm that this interview was properly conducted and is completely confidential

Yes

No