

D-MeX Pilot For PR19

REPORT PREPARED FOR



31 January 2019

Redacted version



PROC.01.0599

CONTENTS

<u>SECTION</u>	<u>PAGE</u>
1 BACKGROUND & OBJECTIVES	1
1.1 Background	2
1.2 Objectives	4
1.3 D-MeX pilot development	5
1.4 Shadow year	7
2 SUMMARY OF RECOMMENDATIONS AND OUTSTANDING ISSUES ...	8
2.1 Recommendations for the shadow year	8
2.2 Outstanding issues to be tested in the shadow year	10
3 APPROPRIATENESS OF A CONTACT FOLLOW-UP SURVEY	11
4 CHOICE OF RATING SCALE	12
5 THE TRANSACTIONS FOLLOW-UP SURVEY	14
5.1 Transactions to follow up	14
5.2 Inclusion of Water UK metrics in the quantitative element	18
5.3 Data collection method	20
5.4 Self-lay providers	21
5.5 New Appointments and Variations	22
5.6 Consultants	22
5.7 Questions to be asked	22
5.8 Question(s) on which to base the incentive	24
5.9 Sample provision and preparation	27
5.10 Sample size feasibility	28
5.11 Survey frequency	30
5.12 Confidence intervals	31
5.13 Sample size options	32
5.14 Weighting the transactions survey results	36
5.15 League table vs performance improvement	38
5.16 Budget implications	39
6 THE RELATIONSHIPS SURVEY	40
6.1 Sampling	42
6.2 An annual industry survey	44
6.3 Questions	44
6.4 Sample size	47
6.5 Combining the results from the two surveys	47
6.6 Budget implications	47

APPENDIX: RESPONSES RECEIVED TO OFWAT'S REQUEST FOR FURTHER FEEDBACK, FOLLOWING THE PILOT WAVE 2 WORKSHOP

Please note that this document presents the results of a year of testing approaches to the D-MeX survey methodology. Any results in this document are to be read in the context that this is a pilot and not are not necessarily a reflection of company performance.

Where information is sensitive it has been redacted for confidentiality reasons. Any redactions are marked with [X]

1 BACKGROUND & OBJECTIVES

1.1 Background

With a growing population, rising customer expectations and climate change affecting the water industry, Ofwat (the economic regulator of the water and wastewater sectors in England and Wales) is looking to develop newer and more flexible ways of regulating.

One element of Ofwat's role is to set the price, investment and service package customers receive through its price review every 5 years. To encourage water companies to provide an excellent level of customer service, the Service Incentive Mechanism (SIM) was introduced in 2010¹. This was (partially) a customer satisfaction measure and significant improvements in company performance have been seen over time as a result of the associated reputational and financial incentives. However, it is felt that there are a number of limitations with SIM, one being that it does not cover the service experience of Developer Services customers.

In July 2017, Ofwat consulted the industry on an alternative set of mechanisms aimed at incentivising water companies to provide excellent customer service for residential customers (C-MeX) and Developer Services customers (D-MeX). Ofwat outlined that D-MeX would apply to 'developer services' (or new connections) customers, which comprise builders, developers, SLPs (self-lay providers) and NAVs (new appointments and variations).

Following further development through separate working groups including relevant stakeholders in the industry, the decision to replace SIM with C-MeX and D-MeX was made in December 2017.

¹ From March 2010 to April 2016, this survey was run and managed by McCallum Layton who changed their name to Allto Consulting in January 2017

Ofwat's December 2017 Final Methodology document² outlined that 'Since July 2015, Water UK has published quarterly performance results for water companies, against a set of standards that developer services customers can expect for the provision of a range of new connection services which, in turn, enable housing growth. Despite this, developer services (new connections) customers continue to raise concerns with us about poor quality service from water companies. While the Water UK reporting has encouraged improvements in the timescales for service delivery, we consider that more needs to be done to improve the quality of customer service received.'

The key features of D-MeX as set out in Ofwat's Final Methodology are outlined below.

D-MeX – based on a single score:-

- Developer services customer satisfaction survey
- Quantitative measure of water company performance against a set of key metrics based on Water UK's existing metrics of service levels for Developer Services customers, and any future iterations of the metrics

Company performance will be ranked and published annually and financial performance payments of up to 2.5% of annual developer services revenue or penalties of up to 5% of annual developer services revenue will be applied to the best and worst performers.

Navigator Research, and its sub-contractor Allto Consulting, were commissioned to design and run pilots of the qualitative element of the mechanism (the customer satisfaction survey) to test its efficacy, so that Ofwat can refine the methodology in advance of its becoming fully operational in April 2020. The scope of this work did not include supporting Ofwat in finalising the selection of the quantitative measures to be included in D-MeX.

This document outlines the results of the qualitative pilots in assessing the options for specific aspects of the methodology, and provides a number of recommendations for Ofwat to consider.

² <https://www.ofwat.gov.uk/publication/delivering-water-2020-final-methodology-2019-price-review-appendix-3-customer-measure-experience-c-mex-developer-services-measure-experience-d-mex>

1.2 **Objectives**

The overall objective of the project, as stated by Ofwat, was to:

Test all aspects of D-MeX so the incentive meets the following success criteria: -

- encourages companies to improve customer experiences and innovate;
- simple and meaningful for companies and customers;
- proportionate;
- practical to implement;
- measures performance across companies consistently, reliably and fairly; and
- reflects customer behaviour changes and market changes.

And develop the mechanisms to the point where they can be run in shadow form (which is intended to be a dry run) in 2019-20, and have financial incentives applied on an annual basis from April 2020 onwards. This involved designing and running at least two waves of pilot customer surveys for D-MeX that were consistent with the design decisions of the incentive outlined in the methodology statement 'Delivering Water 2020: our methodology for the 2019 price review'.

A subsidiary objective was to provide advice and make evidence-based recommendations on how to adapt D-MeX, if necessary, based on the results of the D-MeX pilot.

1.3 D-MeX pilot development

For the purposes of the pilot, various approaches were trialled in Wave 1 and Wave 2 to determine the most appropriate ones to be recommended for the shadow year. In summary, these are outlined below.

Wave 1 (June - July 2018)

Our initial approach for the D-MeX customer satisfaction survey could realistically only cover the first pilot phase. Discussions in the 22 March 2018 workshop and feedback from the responses submitted to the pre-workshop request for information indicated there were still a number of significant issues to resolve before we could recommend how the measure should be designed for the future. The Wave 1 pilot, however, afforded a good opportunity to start to develop a detailed approach.

Industry input pointed to a need to separate out the survey programme into two elements - a transactional follow-up survey to assess the quality of service delivered in relation to specific pieces of work, and a relationship survey to examine issues around how the Developer Services departments manage their dealings with the larger customers they are in regular contact with. The companies considered that both elements were required in order to give a balanced view.

Looking only at operational service delivery would weight the outcomes towards smaller customers, who account for the bulk of the individual transactions a company will deal with day to day. Larger customers like the major housing developers will have fewer transactions in any given month, as a group, than the smaller builder population, but it is how these large developers are serviced that could have greater impact on improving industry efficiency and helping stimulate growth in residential completions.

We therefore recommended that the initial pilot comprise these two elements.

For the transactions survey, companies were asked to provide details of all Developer Services transactions under a selection of key Water UK metrics, that had been completed in May 2018, by 22 June (shortly after they would have provided their May return to Water UK). They were asked to provide all relevant transactions - such that the numbers completed, regardless of whether in time target or not, would tally with their Water UK returns. These were to include transactions carried out for large developers, Self-Lay Providers (SLPs) etc. as well as smaller customers. From this list, we completed a total of 930 structured phone interviews with the relevant customer contact for the transaction in question.

For the relationships survey, companies were asked to provide lists of key individuals within their largest customer firms, with whom the company was in regular contact. From this list, we completed a further 452 structured telephone interviews.

Wave 2 (October - November 2018)

On the basis of our experience at Wave 1, some adjustments were made to the method employed for pilot Wave 2; these were:

- Tailored approaches to the SLP and New Appointment and Variations (NAV) markets
- Exploration of ways to sample the various survey elements as objectively as possible. (Any records provided by the water companies need to be fully auditable, and we need to avoid any potential for companies to be selective in the contacts they provide)
- A site-based rather than individual transaction-based approach to the SLP and NAV operational survey elements. In order to gather more data from these much smaller market segments, we sought ratings and feedback on stages across the entire experience to date in relation to a particular site, rather than in relation to a single work stage only (as has been the approach for builders and developers) - as this might have involved more than one customer contact at the SLP/NAV, these surveys were carried out online rather than by phone

The Wave 2 pilot wave, then, comprised:

- 1,030 phone interviews following up August transactions, with builders and developers
- 109 relationship interviews with large developer contacts, sampled independently
- 10 relationship interviews and 23 site-based operational interviews with SLPs, sampled independently (but with reference to SLP transactions carried out in August)
- 6 relationship interviews with NAVs (sampled independently)

Some adjustments to the survey questionnaires were also made:

- Revision of the individual service factors being rated. Some that were included at Wave 1 were dropped, if their correlation with overall satisfaction was low (indicating little impact on overall perception and therefore lower importance to customers); and some new factors were added (where responses to the open questions showed other issues are important that had not been included at Wave 1)
- Some separate questions for SLPs and NAVs to those used with builders and developers. Industry consultation pointed to a need for the inclusion of specific measures that would be particularly important to SLPs (e.g. around adoption agreements) and NAVs (e.g. around Point of Connection/Point of Discharge applications and NAV framework agreements)
- The use of a numeric 10-0 scale throughout, for all rating questions. At Wave 1, we also piloted a verbal 5 point scale in half of the interviews (randomly), but determined that the 10-0 scale gives greater differentiation in the results and should therefore be consistently employed

1.4 Shadow year

It is intended to run the surveys in shadow form from April 2019 – March 2020, to give companies time to make any systems changes needed before the financial incentives are applied from April 2020 onwards. Another purpose of the shadow year is to allow Ofwat to monitor the operation of D-MeX incentives and make adjustments which they consider appropriate for the period 2020-2025.

Our recommendations outlined in this report relate to the qualitative element of D-MeX for this shadow year, during which any outstanding issues can be tested, and discussed with stakeholders. Final D-MeX guidance will be published by March 2020.

2 **SUMMARY OF RECOMMENDATIONS AND OUTSTANDING ISSUES**

2.1 **Recommendations for the shadow year**

The following table lists our recommendations for running D-MeX through the shadow year from April 2019 – March 2020. Our discussion of and rationale for making these recommendations, and further information from the pilots, can be found in the relevant section of this document as shown in the table.

Table 1 – Recommendations for the Shadow Year

Recommendation	See section:
General	
That D-MeX does <u>not</u> take a contact follow-up approach, but focuses instead on operational customer service delivery	3
To adopt the 10-0 rating scale for D-MeX	4
Transactions	
To employ a selection of the existing Water UK Levels of Service (LoS) metrics as the framework for D-MeX transactions survey sampling for the shadow year	5.1
To include in the sample selection list for the shadow year the further SLP metrics W25.1, W26.1, W28.1 and W29.1 (that were not included in the pilots), to ensure completeness in the coverage of SLP customer journey stages	5.1
Regardless of whether/how the Water UK metrics may be changed in future, that the D-MeX framework should continue to reflect the key stages in the Developer Services customer journey, and companies should continue to collect data on these, including for D-MeX purposes	5.1
To use the shadow year to develop an appropriate way to incorporate NAV activities within the data extraction process for D-MeX transactions survey sampling	5.1
That the transactional follow-up survey be carried out by phone, as the method that will provide the most robust results	5.3
To survey SLPs in the same way as the builders and developers and follow up individual transactions with them by phone, rather than running a separate online site-based operational survey for SLPs	5.4
That shadow year guidance should provide clear instructions on flagging SLP records within the data sets, and the correct identification of these customers be further checked by Ofwat's appointed agent, by reference to the WIRS database held by Lloyd's Register	5.4
Not to carry out a separate operational survey for NAVs, as this process was not effective at pilot	5.5
That consultants and agents be included in the D-MeX survey in the same way as end-customers	5.6

To use the pilot wave 2 questionnaire content, moving the overall satisfaction question from the beginning to later in the interview when the respondent has answered open and specific ratings questions about the transaction and therefore thought through how the matter progressed	5.7
That in terms of which question(s) the incentive will be based on, Ofwat considers keeping this simple and easy for all to understand and work with, by using the single overall satisfaction score as the measure on which companies will be ranked, in preference to an amalgamation of scores against specific service factors	5.8
An ongoing sampling and fieldwork process throughout the year for the transactions survey, with files provided by the water companies each month	5.11
To aim for a sampling proportion of around 20% of customers for each water company, with a boost to 100 for those companies where 20% would fall below this minimum number	5.13
If the transactions survey results are to be weighted by customer type, that Ofwat considers a banding system to give greater prominence in the overall scores to larger customers (ie the largest spenders, most frequent service users and/or most high-volume builders), in preference weighting by to actual revenue or unit/connection numbers; this could be simpler, and would be more comparable from company to company and hence fairer in terms of the ultimate scoring outcome	5.14
That Ofwat use the shadow year to develop an appropriate means by which the transactions data can be weighted by customer size/type, to arrive at a solution that will be simple, clear and workable for all companies	5.14
That consideration be given to incorporating in the qualitative D-MeX outcome some form of individual improvement target as well as league table position. Potential ways in which this could be done can be tested in the shadow year	5.15
Relationships	
In our view, the transactions survey is the core requirement of the D-MeX customer satisfaction tracking programme. A relationships survey would give further weight within the programme to larger developers, SLPs and NAVs, and would capture a wider range of customer perspectives, but as these features do not significantly impact the shape of the outcome, then the additional cost associated with the relationships survey may not be justifiable. On balance, we believe that a transactions survey alone will be sufficient to achieve Ofwat's objective of a simple, proportionate and practical mechanism that measures performance fairly and encourages companies to improve	6
If it is decided to carry out a separate relationships survey, to de-couple this from the transactions data and run an annual phone survey of key contacts at all large developers, SLPs and NAVs, using lists sourced by Ofwat's appointed agent	6.2

To test out weighting options for the relationships survey scores within the overall qualitative measure, in the shadow year, if this element of the survey programme is carried out	6.5
--	-----

2.2 Outstanding issues to be tested in the shadow year

The key issues that are still to be resolved before D-MeX goes live in 2020 are:

- The annual customer population size for each company (calculated from a count of unique customer contacts throughout the year), to confirm sample size and confidence interval estimates
- How companies could best report delivery of work for NAVs, to include in the transactions survey sampling
- A consistent and workable means of classifying customer types across the whole industry, if it is decided to give greater weight in the results to larger customers
- Water company systems readiness to provide the data from which the survey would be sampled

3 APPROPRIATENESS OF A CONTACT FOLLOW-UP SURVEY

Both waves of the pilot programme have included a transactional follow-up element (to measure the quality of day to day work and service delivery) and a wider customer relationship element (to balance the programme coverage that could otherwise over-emphasise the voices of smaller customers, undertaking the bulk of transactions).

It has been suggested that an alternative approach could be a simple contact follow-up survey (along the lines of SIM in the household market and the Customer Service element of the new C-MeX survey programme). On the basis both of company input and of customer feedback through the pilot surveys, we advise against this approach, on two key grounds:

- Many of the water companies do not systematically log all Developer Services customer contact, within a central call centre system. If they were required to do so in order to provide a sample frame for D-MeX, this could have unintended consequences in terms of how companies would manage customer contact, to the detriment of customers (who often talk about wanting to be able to get in touch with individual staff members with queries, disliking call centres that can slow up their ability to get an answer quickly)
- A contact follow-up survey would by its nature be led by how the contact was handled rather than how the work itself is being delivered, and would therefore not give the correct focus on what needs to be improved within the industry (contact handling is only one part of this, along with the speed and quality of the work itself)

We therefore recommend that D-MeX does not take a contact follow-up approach, but focuses instead on operational service delivery for customers.

4 CHOICE OF RATING SCALE

One of the issues with the SIM survey league tables is that scores for water companies have tended to converge over time leaving little to discriminate between companies in statistical terms. It was hypothesised that extending the range of scores that customers can award for their satisfaction with their water company should provide greater discrimination between water companies. To test this, in the Pilot Wave 1 of C-MeX, half of respondents were asked about satisfaction with their water company using a 5 point scale from very dissatisfied to very satisfied and half were asked on an 11 point scale from 0 (extremely dissatisfied) to 10 (extremely satisfied).

On the basis that it will make sense for the satisfaction scoring systems on C-MeX and D-MeX to match, we applied the same test to the rating questions on the D-MeX surveys for Wave 1.

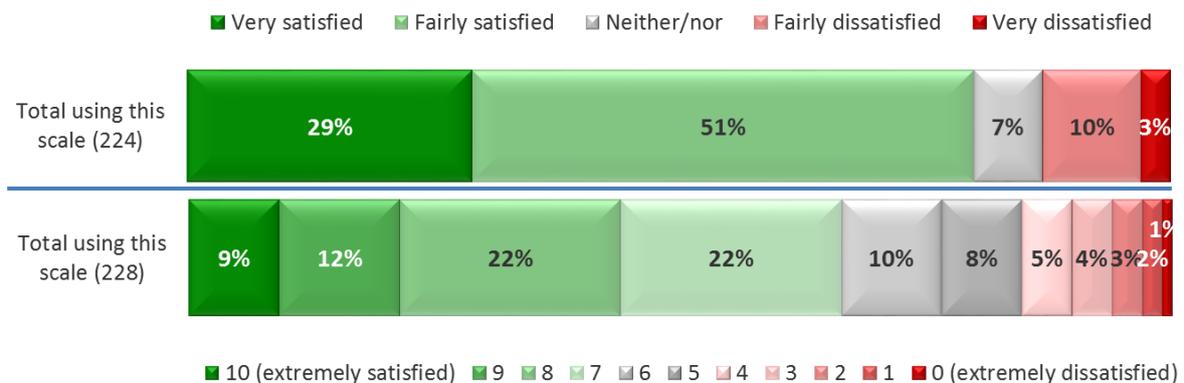
The spread of responses to the overall satisfaction question on each of the transactions and relationships survey at Wave 1 was as follows:

Chart 1 – Comparison of rating scale outcomes, Pilot Wave 1

Transactions



Relationships



In the case of both surveys, the use of the 10-0 scale gives us:

- fairly similar proportions of actively dissatisfied respondents (if we think of 0-4 out of 10 as obviously dissatisfied)
- but rather more in the middle range of the scale (5-6) when compared to the proportion saying 'neither satisfied nor dissatisfied' on the verbal 5 point scale, where respondents may feel they ought to swing one way or the other even if their view is fairly neutral
- and greater differentiation at the positive end, with the blunt 'very satisfied' potentially encompassing scores of 8-10 out of 10 which then enables us to distinguish far more between the best performers

Our recommendation is therefore to adopt the 10-0 scale for D-MeX.

5 THE TRANSACTIONS FOLLOW-UP SURVEY

5.1 Transactions to follow up

D-MeX is intended to be an incentive mechanism to drive improvement in Developer Services customer experience. A framework is required, defining the point(s) of reference against which customers are asked to rate their satisfaction with the service they receive, such that results can be compared consistently between companies and over time. This framework should relate to operational service delivery, and companies should be able to provide customer details for elements within this framework, for follow-up in a customer survey. The details provided by companies should be representative of the work they carry out and of their customer base, and auditable.

The logical initial choice for such a framework has been Water UK's Levels of Service (LoS) metrics system, as this already exists, and all companies already report performance data to this system on a monthly basis.

Water UK is a membership organisation representing all the major statutory water and wastewater service providers in the UK. The organisation has produced a set of standards that developers and others can expect from incumbent water companies in England and Wales, in relation to the provision of infrastructure for housing development. These standards relate to the timescales within which companies should respond or deliver, at key points within a customer's journey through the new connection process; water companies submit performance data monthly, and Water UK publishes the results each quarter.

The existing performance reporting, then, focuses on timescales for the completion of key stages of the Developer Services journey. The D-MeX customer satisfaction survey should build on this, by examining the quality of the service experience. Employing the Water UK framework for D-MeX at this point avoids any need to establish a separate set of points of reference, with attendant potential issues over definitions, consistency of interpretation, ability to obtain data from existing IT/records systems, compliance checking and auditing etc.

The key customer journey points are defined as a set of metrics, listing the range of activities essential for housing development. For the D-MeX pilot, we took a selection of these to question customers about.

Not all of the metrics have been used – activities felt to be purely administrative (such as application acknowledgements) have been excluded, as have lines in the metrics that are for information only. The full list of current metrics, and those selected for inclusion in the D-MeX pilots, are as shown in Table 2 below.

Table 2 – Water UK LoS metrics used for sampling the D-MeX pilots

WATER			Included
W1.1	Pre-development enquiry – reports issued within target	PERF	✓
W2.1	s45 applications – written acknowledgements within target	PERF	
W2.1a	s45 applications - refused/returned/questioned	INFO	
W3.1	s45 quotations - within target	PERF	✓
W4.1	s45 service pipe connections - within target	PERF	✓
W4.1a	s45 service pipe connections - within extended target	INFO	
W5.1	Mains design - written acknowledgement within target	PERF	
W5.1a	Mains design - forms refused/returned/questioned	INFO	
W6.1	Mains design <500 plots - quotations within target	PERF	✓
W7.1	Mains design >500 plots - quotations within target	PERF	✓
W7.1a	Mains designs >500 plots - as % of total mainlaying jobs	INFO	
W7.1b	Mains designs >500 plots - % where extension agreed	INFO	
W8.1	Mains construction within target	PERF	✓
W8.1a	Mains construction within extended target - as % of all mainlaying jobs	INFO	
W16.1	Mains diversions - written acknowledgements within target	PERF	
W16.1a	Mains diversions - applications returned/refused/questioned	INFO	
W17.1	Mains diversions (without constraints) - quotations within target	PERF	✓
W17.2	Mains diversions (with constraints) - quotations within target	PERF	✓
W17.2a	Mains diversion quotations - % where agreed extension	INFO	
W18.1	Mains diversions - construction/commissioning within target	PERF	✓
W19.1	Self-lay point of connection applications - written acknowledgements within target	PERF	
W20.1	Self-lay Point of Connection report < 500 plots etc - reports issued within target	PERF	✓
W21.1	Self-lay Point of Connection reports >500 plots etc - reports issued within target	PERF	✓
W22.1	Self-lay design approval and terms request application - written acknowledgements within target	PERF	
W23.1	Self-lay design and terms request <500 plots etc - quotations within target	PERF	✓
W24.1	Self-lay design and terms request >500 plots etc - quotations within target	PERF	✓
W25.1	Self-lay signed agreement - acknowledgements within target	PERF	
W26.1	Self-lay water for pressure/bacteriological testing - provided within target	PERF	
W27.1	Self-lay permanent water supply - provided within target	PERF	✓
W28.1	Self-lay vesting certificates - issued within target	PERF	
W29.1	Self-lay Asset Payments - issued within target	PERF	
W30.1	Self-lay plot references and costing details - issued within target	PERF	✓

SEWERAGE			Included
S1.1	Pre-development enquiry – reports issued within target	PERF	✓
S2.1	Sewer requisition - written acknowledgement of applications within target	PERF	
S2.1a	Sewer requisition - applications refused/returned/questioned	INFO	
S3.1	Sewer requisition design – offers issued within target	PERF	✓
S4.1	Sewer requisition – constructed and commissioned within agreed extension	PERF	✓
S4.1a	Sewer requisition – constructed and commissioned - extensions agreed	INFO	
S5.1	Technical vetting of adoptions & diversions – acknowledgements within target	PERF	
S5.1a	Technical vetting of adoptions & diversions – applications refused/returned/questioned	INFO	
S6.1	Technical vetting of adoptions & diversions – approval or rejection letters within target	PERF	✓
S6.1a	Technical vetting of adoptions & diversions – extensions agreed	INFO	
S7.1	Adoption legal agreement – draft agreements issued within target	PERF	✓
S8.1	s106 sewer connection - approval letters issued within target	PERF	✓
S9.1	s106 sewer connection - rejection letters issued within target	PERF	✓

The same selection of metrics was used at Wave 1 and Wave 2 of the pilot. It encompasses the requirements on the industry to provide quotations, water mains and sewer requisition design and construction, mains and sewer diversions, sewer adoption, connections, and the elements of service provided specifically to SLPs; in addition to these statutory elements, we included pre-development reports (and self-lay point of connection enquiries) as these often mark the initial point of contact for a customer in relation to a site, and can set the 'tone' of future dealings.

We recommend continuing to use this list as the framework for the D-MeX transactional survey sampling for the shadow year. We would, however, recommend including the further SLP metrics W25.1 (signed agreements – even though this is an acknowledgement metric, it is the only one that refers to self-lay adoption agreements), W26.1 (source of water supply for SLP testing, an earlier stage to the permanent supply) and W28.1 and W29.1 (vesting certificates and asset payments, as these can be required in stages over the development of an SLP site). Including these additional metrics will ensure completeness in the coverage of SLP customer journey stages.

We are aware that discussions are ongoing in the industry about these metrics, and that the list may be amended, reduced or added to in future. We note that it is the water companies who would make these changes (being the members of Water UK), which means there is a potential risk that the metrics framework could be amended to exclude work stages that are typically more difficult and/or less likely to be carried out to customers' satisfaction, in order to avoid customers being surveyed on these and thereby improve scores in D-MeX. (Any changes that are made purely to clarify terms of reference, better reflect what matters to customers, simplify the cases where a stage is split into two and/or to speed up the delivery targets, though, could be beneficial and would not have any implications for D-MeX transactions survey sampling.)

Regardless of whether/how the Water UK metrics may be changed in future, we recommend that the D-MeX framework should continue to reflect the key stages in the Developer Services customer journey, and companies should continue to collect data on these, including for D-MeX purposes.

The pilot survey work indicates that these key stages are, for builders/developers:

- pre-development enquiry
- quotation
- mains design (water) / design approval (waste)
- mains construction
- connection
- adoption (waste)

The existing metrics cover all of these stages (although mains design is split into two according to the size of the development, and mains diversion is split into two depending on whether the scheme is with or without constraints). An additional stage covered in the metrics that has not come through clearly in the pilot work as a key stage in itself is the technical vetting of sewer adoptions and diversions - but we believe that this has value in its own right as a stage in the journey, and should therefore be included in D-MeX, when approval or rejection could significantly impact on the customer's ability to proceed with the project.

As far as D-MeX is concerned, the two water mains design metrics W6.1 and W7.1 could be combined without detriment, as could the two mains diversion quotation metrics W17.1 and W17.2 (there would be no sample duplication as a single customer would currently appear under one metric or the other, and if they happened to appear under both, de-duplication would take out one of these cases anyway), although companies may still wish to retain the distinctions being made here.

For SLPs, the pilot survey work indicates that the key stages are:

- Point of Connection report
- Design approval
- Adoption agreement
- Source of supply for testing
- Permanent water supply
- Vesting certificate(s)
- Asset payment(s)
- Service connection approvals

Again, the existing metrics cover these, and our recommended additions to those used in the pilots for sampling transactions in D-MeX will reflect these stages.

The SLP point of connection report and design and terms request metrics are split into two, for sites that are <500 and >500 plots (W20.1/21.1 and W23.1/24.1). Again as far as D-MeX is concerned, W20.1 and W21.1 could be combined without detriment, as could W23.1 and W24.1, although SLPs may still wish to retain the site size distinctions being made here (if they feel there is a continued need to have service levels on larger sites identified separately).

The one customer segment not accounted for in the metrics is NAVs. For these, key stages identified in the pilot work are:

- Point of Connection/Point of Discharge application process
- Quotation
- Off-site design approval
- Contractual framework agreement
- On-site build (if applicable)
- On-site connection

Until such time as relevant NAV metrics are added to the existing Water UK framework (if indeed they are), we recommend using the shadow year to develop an appropriate way to incorporate NAV activities within the data extraction process for D-MeX transactions survey sampling.

5.2 Inclusion of Water UK metrics in the quantitative element

The discussion in section 5.1 has related to the metrics which are used to sample the D-MeX transactions survey from. A separate issue is which of these metrics are used by Ofwat to set the quantitative element of the mechanism. While the purpose of this document is primarily to report on developing the survey element of D-MeX, we have been asked to consider which are the most important metrics as far as the customer is concerned, to inform decisions around setting the quantitative element.

On the basis of the pilot outcomes, we consider that they are all important, as they each represent a stage in a process that the customer has to go through (subject to the nature of the project), and the quality of the service received at each point directs how smoothly the work can

proceed to the next stage. We cannot therefore recommend a sub-set of the metrics on the basis of which are the most important, other than to suggest excluding acknowledgements, as we have done in defining the survey sampling approach.

An alternative method could be to use customer satisfaction on each of the metrics as a means of determining which to include in the quantitative measure (focusing on those where performance is weaker, to encourage improvement in those areas). There are a number of problems with this idea, though. One is base sizes (see the table below, showing satisfaction by metric from pilot wave 2) – some stages have very few or even no data points, so it is not possible safely to judge where they sit. Another is the nature of the work at each point - one of the lower scored stages is sewer connection rejections, and satisfaction there may be impacted as much by outcome as by service, and other lower scoring stages are mains design and construction, which only apply to large developers and not across the board, plus the sewerage metrics will only apply to WASCs. A further issue is that satisfaction levels will change over time, as company performance improves with increased focus on individual areas of work, such that metrics with lower scores now rise up the table and cease to be priority targets.

Table 3 – Sample bases and satisfaction scores by metric, Pilot Wave 2

METRIC (excluding SLP metrics, collected separately)	BASE (interviews carried out on this transaction type)	PERFORMANCE (mean average overall satisfaction)
S3.1 Sewer requisition design – offers issued	1	8.00
W27.1 Self-lay permanent water supply - provided	1	8.00
W30.1 Self-lay plot references and costing details - issued	1	8.00
S1.1 Pre-development enquiry – reports issued	49	7.90
S7.1 Adoption legal agreement – draft agreements issued	15	7.73
W1.1 Pre-development enquiry – reports issued	26	7.35
S8.1 s106 sewer connection - approval letters issued	104	7.29
W4.1 s45 service pipe connections	307	7.15
AVERAGE	1027	7.03
W3.1 s45 quotations	367	6.97
S6.1 Technical vetting of adoptions & diversions – approval or rejection letters	29	6.83
W6.1 Mains design <500 plots - quotations	39	6.56
W18.1 Mains diversions - construction/commissioning	4	6.25
S9.1 s106 sewer connection - rejection letters issued	43	6.12
W8.1 Mains construction	34	6.03
W17.1 Mains diversions (without constraints) - quotations	7	5.71
S4.1 Sewer requisition – constructed and commissioned	0	
W17.2 Mains diversions (with constraints) - quotations	0	
W7.1 Mains design >500 plots - quotations	0	

Our view as to which metrics are used to comprise the quantitative element of D-MeX, then, is that they are all relevant (excluding administrative acknowledgements and information only items) and that including them all is simpler and safer than trying to reduce these to a sub-set.

5.3 Data collection method

The main method used for this survey at both Wave 1 and Wave 2 was telephone interviews.

We recognise that dealings in relation to any particular site may be handled by more than one individual at a customer firm, and where this is the case, an invitation to take part in an online survey, that can be shared around the team for each to contribute their own perspective, might be more appropriate than approaching a single individual by phone. We tested this method at Wave 2 with SLPs, and attempted also to test it with NAVs; the completion rate on the online SLP operational follow-up was 20% of all contacts attempted and 38% of those who actively agreed to take part when approached, while for NAVs we did not receive any responses to the online survey (note, though, that this latter point is likely to be as much about the nature of NAV interactions with Developer Services teams – see section 5.5 - as about the survey method). Among the builders and developers surveyed by phone, the response rate (interviews as a proportion of interviews plus refusals) was 55%. The conversion rate on the phone survey (records used per interview achieved, including other outcomes and possible callbacks still in the system at the end of the fieldwork period) was 5.7; for the SLP operational survey this was rather lower, at 5.0, but this was because particular effort was put into pursuing the much smaller number of SLP contacts with a view to securing an outcome from all of them.

Achieving as representative a sample as possible is crucial to successfully delivering a fair financial incentive mechanism, and this means minimising non-response. The potential gain in terms of capturing the views of multiple team members has to be balanced against the likelihood of their actually taking part, and working on the basis of hearing about around half of targeted transactions from a single perspective has to be safer than limiting ourselves to perhaps a fifth from multiple perspectives.

It could be argued that we have not piloted running the survey online among builders and developers, and the response could be better within those market segments. From experience, our view is that the response online from the 'mass market' would not be any better, and could indeed be lower.

As well as generating a higher response rate and hence better sample representativeness, a conversation over the phone led by a professional interviewer will often provide a higher quality of response, as the interviewer can probe incomplete or inconsistent answers and encourage due consideration of rating scores given.

There is always the option of allowing customers to complete the survey online if they specifically request this, but the option should not be actively offered (as respondents may agree to take part online simply to close the call).

We recommend that the transactional follow-up survey be carried out by phone, as the method that will provide the most robust results.

5.4 Self-lay providers

At Wave 1, SLPs were interviewed along with builders and developers in the transactions survey. The sample achieved was relatively small, and it was not always possible to identify SLPs within the records provided by the water companies.

It was decided at Wave 2 to survey this market segment separately in order to gain a more targeted view of their needs and views, and an online site-based operational survey was trialled. As noted earlier, the response rate to this was noticeably lower than that seen on the phone survey. In addition, relating the questionnaire to the firm's whole experience over the life of the site in question (rather than just a single work stage) meant that the results could not be rolled together with the rest so a separate analysis phase was required for what was ultimately a relatively small set of results.

On the basis of the pilot, we recommend using the Wave 1 approach, treating SLPs in the same way as the builders and developers and following up individual transactions with them by phone. This approach will simplify the programme design for the shadow year, and because the level of SLP activity is so variable between water company territories, will avoid the possibility of SLP feedback being over- or under-represented in the results as a whole (as they will be surveyed in proportion to their presence in each company's records, in the same way as the other customer groups).

Extending the Water UK metrics list used for D-MeX sampling to include acknowledgement of adoption agreements, source of supply for testing, vesting certificates and asset payments will ensure SLP inclusion in the survey can be maximised.

Shadow year guidance should provide clear instructions on flagging SLP records within the data sets, and the correct identification of these customers can be further checked by Ofwat's appointed agent, by reference to the WIRS database held by Lloyd's Register.

5.5 New Appointments and Variations

There are currently 6 NAV companies. At Wave 2, we carried out a relationships interview with all of these and found quite limited engagement with water company Developer Services teams. We attempted to carry out a site-based operational survey (similar to the approach taken with SLPs) but received no responses other than a comment from one that they did not feel they could fairly rate any incumbents as their dealings with them are too limited.

Again with a view to achieving a design approach for the shadow year that is not overly complex, we recommend not repeating this survey element in 2019.

There is a NAV desire to include some metrics in the Water UK Levels of Service set referring specifically to NAV activity, relating to Time to Connect and Time to Quote. If these are adopted then they can be added to the set of metrics used in D-MeX and the NAV customers surveyed within the overall transactions programme alongside the other market segments.

5.6 Consultants

During the pilot, there was some debate as to whether consultants and agents engaged by builders and developers should be included in the survey, on the grounds that it is the needs of builders and developers that companies are required to meet.

Our view is that these consultants and agents should be included in the D-MeX survey in the same way as the end-customers, as the developers will in these cases have chosen to engage professionals to act on their behalf and the needs and views of the consultant are then commensurate with those of the end-customer.

Indeed, it could be argued that SLPs are themselves acting as consultants or agents on behalf of end-customer developers and it is clearly important that the experiences of this specific market group are captured in D-MeX.

5.7 Questions to be asked

The questionnaires used at Wave 1 and Wave 2 were broadly the same, although the set of satisfaction rating questions relating to specific service factors was updated for Wave 2 in light of open-ended comments made at Wave 1 about customer priorities and where they would most like to see Developer Services improve.

In both cases, the average interview length came out at around 13 minutes, which is reasonable for this type of feedback survey. There may be some concern that this is on the long side and a target of nearer 6-8 minutes would be desirable. We consider, though, that respondents must be allowed the opportunity to think about and comment on their experience in their own words rather than simply responding to a battery of rating questions, and taking a few more minutes to enable this makes for a more satisfying interview encounter for respondents (which will help future response rates) as well as gaining a more considered set of results.

We therefore recommend using the Wave 2 questionnaire content in the shadow year, moving the overall satisfaction question from the beginning to later in the interview when the respondent has answered open and specific ratings questions about the transaction and therefore thought through how the matter progressed.

Ofwat may wish to consider whether or not to continue asking the question used in the pilots about how transparent the company is about developers being able to use alternative infrastructure providers for certain services – while this issue is a priority for Ofwat, it is less relevant for Developer Services customers (if they do not know about competition, they cannot answer, and if they do, and are already using other providers – or are themselves an SLP – the company’s transparency is not really significant).

The question set may therefore be as follows:

1. What do you feel that [WATER COMPANY] did well, if anything, in relation to this particular transaction? (open)
2. And what do you feel that they could have done better, if anything - again in relation to this particular transaction? (open)
3. I’m now going to read out a number of aspects of service, and I’d like you to tell me how satisfied you are, on each of these, still thinking about your dealings with Developer Services at [WATER COMPANY], in relation to this specific transaction (rating scales):
 - a) Ease of contacting them
 - b) The quality of the information available on their website
 - c) Understanding your needs
 - d) Timeliness of response to queries and requests
 - e) Keeping you informed on progress, where required
 - f) Offering value for money
 - g) Completing the work within a timescale that is reasonable
 - h) Meeting agreed deadlines
 - i) Their efficiency in handling this stage of the work
 - j) Accuracy and completeness of any documentation provided (eg quotations, plans, reports etc)
 - k) Any advice and guidance they gave you, to help progress the work
4. How satisfied are you overall with how [WATER COMPANY] handled this particular transaction? (rating scale)

5. Thinking more generally about dealing with Developer Services – what would you personally say are the top 3 most important things you are looking for, in terms of how they deliver their service? (open)
6. If there was one thing that you'd suggest that the Developer Services team at [WATER COMPANY] could do, that would have most impact on how easy you find them to deal with, what would it be? (open)
7. Respondent name and job title
8. Whilst this survey is being carried out on behalf of Ofwat, [WATER COMPANY] is keen to learn from your feedback and improve its customer service. With this in mind are you willing for us to pass your answers back to them, identified by name?
9. That is the end of the interview. Are there any other comments you would like to make, before we close?

5.8 Question(s) on which to base the incentive

The options here include using:

- The overall satisfaction score alone, or
- A combination of rating questions that reflect customer priorities

The first of these options is the more simple and straightforward. The second could be more effective in focusing minds on what matters most.

Key themes emerge from the open comments and from quadrant chart analysis of derived importance by satisfaction, pointing to a particular need at this point for improvements to be made in:

- Speed of turnaround on delivery of work
- Meeting deadlines
- Keeping the customer informed of progress/issues
- Speed of response to queries and requests
- Proactive collaboration, responsiveness to commercial business needs and to work programmes
- Ease of doing business (communication, advice & guidance, access to the right people)

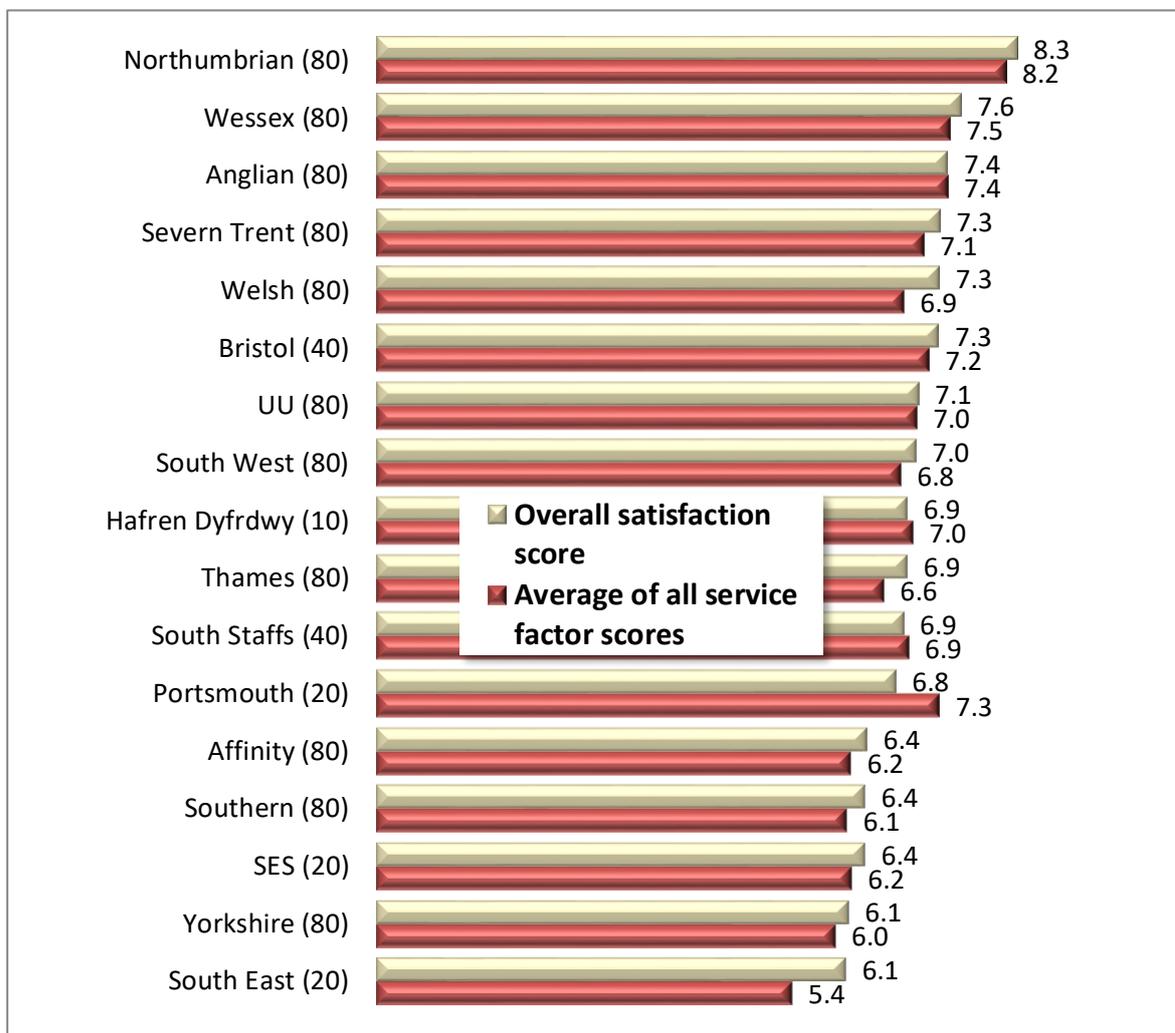
These are all covered in the service factor elements to be rated. It would therefore be feasible to define an amalgamated score using some or all of the rated service specifics to generate the measure on which the incentive will be based. This would be worth doing if an average of the amalgamated scores came out noticeably different to the overall satisfaction for each company (meaning that doing this added value to the performance ranking process) and/or if company performance on each service factor varied such that they were good in one area and poor in another (meaning that this process would ensure they addressed those poorer performing areas).

As mentioned previously, please note that as the results that follow are based on a pilot, they may not necessarily reflect company performance.

Looking at the first of these scenarios, the industry average overall satisfaction score seen at Wave 2 was 7.03; an average of all the service factor satisfaction scores was similar, at 6.86. Considering this by company (see the chart below), in most cases the difference between the two measures is no more than 0.1 – with some exceptions:

- Portsmouth Water does better on the average of the service factors than on overall satisfaction (overall perceptions of them may be being negatively affected by something else – although their base of 20 interviews is small)
- South East Water, Southern Water, Thames Water and Welsh Water do better on overall satisfaction than on the average of the service factors (negative impressions of aspects of service delivery are perhaps being mitigated by other aspects that are stronger)

Chart 2 – Overall and aggregated satisfaction scores, Pilot Wave 2



In general, the differences between the overall satisfaction score and the average service factor score are quite small, though.

The second scenario referred to above which might lead us to consider using the service factor scores in defining the incentive question(s) is if individual companies do particularly well on some and poorly on others, as this could encourage them to focus on improving the under-performing aspects. This is true in some cases and not others:

- Two companies do well across the board, emerging in the top 5 for each of the service factors rated, and five companies do poorly across the board
- One company, on the other hand, does well on most factors but under-performs on website information and value for money; another is in the top 3 for several factors but falls down on understanding needs, keeping the customer informed, meeting deadlines and advice & guidance

There could be a case, then, for taking the service factors into account in setting the incentive. However, it would be necessary to summarise them in some way to arrive at a single figure and as we have seen above, taking a straight average does not significantly alter many of the positions vs simply using the overall satisfaction question. Any other means of summarising the factor scores (such as taking an average of each company's ranking within each factor individually) starts to become quite complex for companies to be able to unpick, communicate internally and act upon.

A further issue we have in considering using the service factor scores in this way is that it may potentially be subject to changing market needs. If we selected factors to include within the measure that are particularly important to customers now, companies would rightly focus on improving their service on that specific factor (albeit possibly to the detriment of other service areas), which would in time lead to that factor becoming less significant to customers because it is no longer a problem for them. A way round this is to weight the factor satisfaction by the derived importance score (generated by correlating individual factors against overall satisfaction, which would go down if the factor ceases to be as important in future); the derived importance would have to be calculated for each factor for each company individually, as these will vary, so again the process becomes quite complex for companies to manage.

Overall, and especially given that this is a completely new measure, we recommend that Ofwat considers keeping this aspect of the design simple and easy for all to understand and work with, by using the single overall satisfaction score as the measure on which companies will be ranked.

The issue of timescales (a key priority for customers) will be factored into the measure through the application of Water UK Levels of Service performance data as a quantitative element applied to the qualitative overall satisfaction data.

The interview should continue to seek feedback on the service specifics, even if it is decided not to employ these directly in the incentive calculations, and raw data should be passed back to the companies (attributed or anonymised, according to the customer's preference) so that they can carry out their own diagnostic analysis on the drivers of overall satisfaction.

We do not propose that call recordings from these business interviews be shared with companies. The closer nature of the relationship that customers will have with the Developer Services teams they deal with (compared to a household customer with the company as a whole) means that it could potentially cause embarrassment for the customer next time they need work doing, if the company has heard exactly how they have expressed their opinions – which would detrimentally impact on future response rates. In addition, while they would of course be asked at the end of the interview if they are prepared for the recording to be shared, in this type of interview it is good practice to inform the respondent at the start that they will be asked this later, so that they can temper their comments/language accordingly if they wish, and this can reduce the detail and the value of the feedback they give.

5.9 Sample provision and preparation

Details of transactions completed within the month in question should be provided by the water companies as soon as possible after the end of the month. In reality, this is likely to be shortly after the companies submit their return for the month to Water UK, although some may be in a position to generate the data sooner (and it is to be hoped that those whose systems are not currently geared up to combine information on the transaction with information on the relevant customer contact will be able to address this over the course of the shadow year). The sooner a customer contact can be followed up about their experience of a particular matter, the fresher their memory of the events will be.

Clear guidance will need to be provided to the water companies on the data content and format required, in good time, to support those who may need to allocate resource and/or secure budget to adapt their systems.

The guidance should also cover the requirement that details of all transactions be provided. There is no need to seek customer opt-in to being approached to take part in this survey (as one of the water companies has assumed), as the survey is part of Ofwat's price control work in relation to regulated companies.

5.10 Sample size feasibility

Our recommendations for setting appropriate sample sizes for D-MeX are largely based on practicalities. While taking confidence intervals into account would generally be important for any financial incentive mechanism to be fair and robust, the potential survey size will also be determined by the volume of customers that can be consulted.

For Wave 2, we received a total of 24,020 records from the water companies representing the transactions reported to Water UK for August 2018. The process of cleaning the data (removing records with missing contact names and/or phone numbers) and de-duplicating (removing individual contacts occurring more than one) brought this down to 6,662 usable builder and developer records (SLP contacts were removed for a separate survey approach in Wave 2). From this, we achieved 1,030 interviews, aiming for higher conversion proportions where the sample numbers were low and setting a maximum target of 80 where the numbers were high.

By company and metric, the usable sample volumes and numbers of interviews achieved were:

Table 4 – Sampling and completion statistics, Pilot Wave 2

	Records loaded	Interviews completed	%		Records loaded	Interviews completed	%
TOTAL	6662	1030	15%	TOTAL	6662	1030	15%
Affinity	339	80	24%	W1.1 Pre-development enquiry reports	124	26	21%
Anglian	1032	80	8%	W3.1 s45 quotations	2239	369	16%
Bristol	180	40	22%	W4.1 s45 service pipe connections	2086	308	15%
Hafren Dyfrdwy	28	10	36%	W6.1 Mains design quotations <500 plots	192	39	20%
Northumbrian	326	80	25%	W7.1 Mains design quotations >500 plots	3	0	0%
Portsmouth	76	20	26%	W8.1 Mains construction	200	34	17%
Severn Trent	409	80	20%	W17.1 Mains diversion quotations, without constraints	54	7	13%
South East	83	20	24%	W17.2 Mains diversion quotations, with constraints	1	0	0%
South Staffs	205	40	20%	W18.1 Mains diversion construction/commissioning	24	4	17%
South West	571	80	14%	W20.1 Self-lay Point of Connection reports <500 plots	0		
Southern	449	80	18%	W21.1 Self-lay Point of Connection reports >500 plots	0		
SES	52	20	38%	W23.1 Self-lay design and terms quotations <500 plots	5	0	0%
Thames	362	80	22%	W24.1 Self-lay design and terms quotations >500 plots	0		
UU	690	80	12%	W27.1 Self-lay permanent water supply provided	4	1	25%
Welsh	766	80	10%	W30.1 Self-lay plot references and costing details	8	1	13%
Wessex	388	80	21%	S1.1 Pre-development enquiry reports	340	49	14%
Yorkshire	706	80	11%	S3.1 Sewer requisition design offers	10	1	10%
				S4.1 Sewer requisition construction/commissioning	2	0	0%
				S6.1 Technical vetting of adoptions & diversions	225	29	13%
				S7.1 Adoption legal agreements	148	15	10%
				S8.1 s106 sewer connection approvals	675	104	15%
				S9.1 s106 sewer connection rejections	322	43	13%

Note that two companies were only able to provide a sample of their records with customer contact details appended, rather than the full set. Another had sought an opt-in from their customers to take part in the survey, so we were unable to use those of their records where the customer had declined to be contacted.

The number of usable records in this one month ranged from around 30 to over 1,000 per company. When further files are provided for another month, some of the same contacts will appear again while others will be fresh. As time goes on, it is likely that many customer contacts will appear time and again within the records. It will not be reasonable to keep contacting these individuals every time they crop up in the records, as they will soon suffer from survey fatigue.

A minimum cycle will need to be applied in these cases, suppressing contacts who have previously been approached to take part within a set time period. What this time period should be is not set in stone; at this stage, we would suggest that 6 months could be appropriate, but this can be tested in the shadow year.

Note that applying this principle runs the risk that any individual contact who appears multiple times from month to month might only ever be interviewed in relation to an early stage of their dealings with the company in question, and suppressed thereafter. This risk will need to be managed in the sample preparation process to ensure they are consulted across a range of work stages.

The level of de-duplication needed within and between sample files in any given month already reduces the number of records provided by the water companies that will be usable for the survey, as seen above, leaving perhaps around 1 in 4. Suppression of contacts who have taken part in the last 6 months will further reduce the usable number although we do not at this stage have the data to say to what extent.

In order to suggest the kind of sample sizes it could be practical to achieve from the records available and hence provide estimated options for sample sizes for the shadow year, we will assume that over the course of a year the annual usable records could amount to 9 times the total for August as shown in the table above (rather than 12 times, which it could be if there were no repeat customer contacts from month to month). That would mean annual sample totals of between around 250 and 9-10,000 per company.

A conversion rate of up to around 35% could be feasible (as achieved for Hafren Dyfrdwy and Sutton & East Surrey Water at Wave 2, which had the fewest records), but targeting this level of sampling for the largest companies would be excessive in comparison.

In section 5.13, we show the options available for setting appropriate sample sizes for D-MeX based on the discussion above of feasibility, plus (in the next two sections) consideration of survey frequency and statistical robustness.

5.11 Survey frequency

The transactions survey can either be run on an ongoing basis, with sample provision and fieldwork taking place each month throughout the year, or periodically, focusing on one month per period (e.g. quarterly).

An ongoing process would have the following key benefits:

- The ability to access and make use of as many records as possible for the smaller companies, where the level of their monthly submissions severely restricts the volume of interviews that can be achieved (and if the survey is to be ongoing for the smaller companies, it will make sense to do the same for the larger ones as well, for process consistency)
- Avoidance of the possibility of seasonal peaks and troughs impacting on the results (and also the possibility of more effort being put into customer relations in some periods of the year than others)

We therefore recommend an ongoing sampling and fieldwork process throughout the year for the transactions survey, with files provided by the water companies each month.

We are aware that some of the water companies may want to do their own internal customer surveying alongside the D-MeX survey. It may be possible for this to be facilitated for the larger companies, if Ofwat's appointed agent takes a sample of the full set of records provided in a month for inclusion in the D-MeX survey and returns the rest to the water company for them to use internally. This could, however, have implications in terms of sample management, as an individual customer could be approached separately for D-MeX and for a water company (even, potentially, for more than one water company) within a short period, which over time could depress D-MeX response rates through survey fatigue. Given the relatively small pool of Developer Services customers and the already stringent need for careful sample management, it will be safer to assume that all regular, structured customer satisfaction surveying among transactional customer contacts be carried out within D-MeX and not also through individual water company programmes. (This does not preclude the possibility of the water companies periodically carrying out other ad hoc pieces of customer research, for example focused specifically on testing or developing new services.)

We are also aware that some companies may find it administratively stretching to provide monthly files at this stage. They will, however, have to gear up their internal systems to be able to export these data easily in future, so the additional focus on doing this now may be beneficial in the long run. If we decided only to take periodic samples across the board on the basis of those who currently have to undertake a fair amount of manual work to pull the required data together, the robustness and hence value of the shadow year operation would suffer.

5.12 Confidence intervals

Confidence intervals demonstrate the range within which a survey result would fall if everyone in the relevant population were asked the question rather than just a sample. When seeking to compare and rank company performance, it is clearly important that confidence intervals are as low as possible, as the degree of confidence around their scores could mean that one company looks on the face of it to have done better than another (and could then be rewarded) but actually might not have done because their confidence intervals overlap which means the ranking is unreliable.

Around a mean score such as the overall satisfaction rating referred to here, confidence intervals are calculated by reference to the mean score itself and also the standard deviation (a measure of how spread out the individual scores are, so how consistently or inconsistently the company is felt to perform).

In the case of large markets such as the household sector for water companies, the size of the 'population' is irrelevant in calculating confidence intervals around a survey result. For Developer Services, though, where the total number of people in any one water company's customer population is much smaller, an adjustment to the standard formula needs to be made to account for this.

At Wave 2, the mean average overall satisfaction score across the industry was 7.03 out of 10, and the standard deviation 2.69. Using these figures, and the potential total population sizes for companies across the size range, the confidence intervals (at the 95% level) around this result at various sample sizes would be:

Table 5 – Confidence intervals at various population and sample sizes

Total number of customers dealt with over the course of a year (population)	Confidence intervals at the following sample sizes - +/-:			
	100	400	800	1,600
250	0.41	-	-	-
500	0.47	0.12	-	-
1,000	0.50	0.20	0.08	-
5,000	0.52	0.25	0.17	0.11
10,000	0.52	0.26	0.18	0.12

(Note that these confidence intervals will change slightly, with different mean score and standard deviation figures.)

This statistical consideration has also to be balanced against what is practical, of course. Securing 400 interviews from 500 records, for example, may provide a particularly reliable result in theory, but 80% conversion on a survey could not be achieved in reality.

5.13 Sample size options

There are broadly 3 approaches that could be taken to determining how many interviews to carry out per annum on the transactions survey:

- A fixed number per water company
- The number of interviews that would be required to achieve a fixed confidence range around the satisfaction score
- A fixed proportion of each company's customer population

Fixed number of interviews

A fixed number per company per year is simple, and can easily be determined, and costed, in advance. However, given the huge variability in the size of each company's customer base, a single fixed number does not make sense, so a banding system would have to be used. The simplest system might be 3 bands for small, medium and large companies; aiming for, say, 100 interviews for the smallest companies, 800/1,000 for the largest and 400/500 for those in the middle. Even within these bands, however, there is a great deal of variation

The latest full year Water UK data (2017) shows the following total numbers recorded by company (these are of course Levels of Service metric line items and not customers, but the principle is the same):

Table 6 – Water UK LoS reporting totals, 2017

Thames	147,050
Northumbrian	50,633
Affinity	50,621
Anglian	49,125
Severn Trent	44,826
United Utilities	44,177
Yorkshire	44,057
South East	35,747
Southern	23,876
Dŵr Cymru	23,306
South West	22,951
South Staffs	18,569
Wessex	14,731
Bristol Water	8,373
Sutton & East Surrey	4,917
Bournemouth	3,889
Portsmouth	3,489
Dee Valley	782

Thames Water and Hafren Dyfrdwy are outliers at either end, so would need to be special cases. Otherwise, the 'large' set could include Northumbrian Water down to Yorkshire Water, the 'middle' set South East Water down to Wessex Water, and the 'small' set Bristol Water down to

Portsmouth Water. In that case, though, Portsmouth Water would get the same number of interviews as Bristol Water, which has twice the volume, and Wessex Water would get the same as South East Water which has nearly 3 times the volume. The number of bands could be extended, but how many this should be and how the companies should be grouped would be fairly arbitrary. This solution is therefore not statistically robust.

Fixed confidence interval

The second option is to aim for a fixed confidence interval around the mean average overall satisfaction score.

If the incentive is to be set on the basis of a company's position in a league table, then it is important to ensure that results are as accurate as possible – one company that misses out on a performance payment or is subject to an under-performance payment behind another company whose score is only slightly better will want to know whether the difference is statistically significant or might just be down to chance.

A target fixed confidence interval can be set at any level of accuracy, depending on the practicalities of available sample numbers, and budget. The average overall satisfaction scores at the pilot wave 2 transactions survey came out as follows:

Chart 3 – Overall satisfaction by company, Pilot Wave 2



Again as mentioned previously, please note that as these results are based on a pilot, they may not necessarily reflect company performance.

The difference between the lowest of the top 6 companies and the next one down was 0.26, and between the highest of the bottom 5 and the next one up, 0.37. These differences would need to be halved in order to determine the confidence intervals beyond which the companies would be differentiated statistically; Bristol's 7.31 +/- 0.12 (worst case 7.19) and UU's 7.05 +/- 0.12 (best case 7.17) would be significantly different but at +/- 0.14 or more, they would not, and at the other end, Affinity's 6.38 +/- 0.18 (best case 6.56) and Portsmouth's 6.75 +/- 0.18 (worst case 6.57) would be significantly different.

As target confidence intervals, these would be a stretch. As we have seen in section 5.12, they are simply unachievable for the smaller companies, and would call for over 1,500 interviews in a year for the largest. If this approach were taken, a more realistic target might be +/- 0.2, accepting the consequence of less accuracy in differentiating between companies in a league table. What the sample sizes would need to be to achieve this will have to be based on certain estimations at this point, as confidence intervals around a mean score vary according to the score and the standard deviation, as well as the population size, as discussed earlier. It is possible to table a reasonable estimation for this purpose, though, using the pilot wave 2 industry average performance and a view of each company's likely population size as a multiplier of their monthly customer numbers at pilot wave 2. The number of unique customers over the course of a full year will be less than 12 times a single month, allowing for duplication, as discussed in section 5.10; what the multiplier will actually be can be tested in the shadow year, but we are suggesting that this might be 9 times a single month, for the sake of this exercise.

Using these estimations, the sample sizes that would be needed to achieve +/- 0.2 would be:

Table 7 – Required sample sizes for +/- 0.2 confidence

	Sample size required for +/- 0.2 confidence	Proportion of customers this may equate to
Affinity	550	18%
Anglian	620	7%
Bristol	500	31%
Hafren Dyfrdwy	185	73%
Northumbrian	550	19%
Portsmouth	350	51%
Severn Trent	650	9%
South East	360	48%
South Staffs	500	27%
South West	600	12%
Southern	600	15%
SES	280	60%
Thames	650	7%
UU	650	10%
Welsh	650	9%
Wessex	600	17%
Yorkshire	650	10%
	8945	

This approach would not be possible for Hafren Dyfrdwy, Portsmouth Water and Sutton & East Surrey Water, where obtaining the sample sizes to achieve a confidence interval of +/- 0.2 would mean converting half or more of their records, which will not be practical – they would need to be capped at 100 interviews each, with attendant confidence intervals of around 0.41 to 0.49.

For the larger companies, this would set a cap at around 650 interviews in a year (that may only represent less than 10% of their contacts). We are concerned that this does not give enough scope to ensure good representation of their customer base, especially in the shadow year when issues around weighting the results by customer group still have to be resolved.

Target confidence intervals could also be set in bands, such that small, medium and large companies each have their own target (with Thames Water and Hafren Dyfrdwy again potentially as outliers), but this would negate the whole purpose of using a set level of accuracy to differentiate between companies.

Fixed percentage of customers

The third approach would be to aim for a sample size representing a fixed proportion of the number of customers that each company deals with in the course of a year. This is the option that would relate most directly to the volume of work that each company handles, and thereby ensure maximum representation of the spread of their own customer base.

What this proportion should be set at has fewer practical issues, as it can always be achievable in terms of customer numbers regardless of company size (as long as it is kept at or below around 30%, shown to be fully feasible in pilot wave 2).

At different potential proportions, the approximate total number of interviews per year may be:

- At 10% sampling – 7,100
- At 15% sampling – 10,500
- At 20% sampling – 14,000

Excluding the smallest companies that would need their annual totals boosting to 100 to achieve a reasonably robust set of data (where the confidence intervals would be around +/- 0.41 to 0.49 simply because their populations are so small), the range of approximate confidence intervals around each company's mean average overall satisfaction score at different sample size proportions could be as follows:

- At 10% sampling, +/- 0.16 to 0.39
- At 15% sampling, +/- 0.13 to 0.31
- At 20% sampling, +/- 0.11 to 0.26

Our recommendation for the shadow year would be to aim for a sampling proportion of around 20% of customers for each water company, with a boost to 100 for those companies where 20% would fall below this minimum number. This approach has a systematic basis, maximises the reliability of the data within the bounds of what it will be practical to achieve, and ensures effective representation of the range of customer types that each company services.

5.14 Weighting the transactions survey results

It was suggested by the industry that the transactions survey covering day to day operational service delivery should be run alongside a relationships survey with key customer contacts, as the results would otherwise be skewed towards the outcomes experienced by small customers who account for a large proportion of individual transactions. We have therefore run relationships surveys at both Wave 1 and Wave 2, employing slightly different approaches in each case (detailed in section 6) but with a view to ensuring feedback is obtained from key individuals at the larger customers.

Whether or not a separate relationships survey forms part of D-MeX (but especially so, if not), then it may be appropriate to weight the transactions survey results, applying a value to each overall satisfaction score based on the relative 'importance' of the customer giving it. These values could be determined on the basis of:

- A specific profile figure, such as spend with the company over the previous year, the number of housing units they complete or connections they request in a year etc.
- Or, a banding system to give weights of, for example, 1, 2 or 3 according to the customer type, where 1 applies to one-off connections, 2 to small builders with occasional transactions and 3 to large developers (and SLPs, potentially)

Applying specific values relating to spend, connections etc. would be a more sophisticated approach, but with the downsides that it would not be feasible to audit any figures provided by the water companies (and some may have difficulty appending these values to their transactions records anyway), scores given by large customers of the likes of Thames Water could far outweigh those of companies in regions where there is less development (but where their largest customers are relatively just as important to them), and this would potentially mean that the views of customers requesting one-off connections (that need more hand-holding through the process) more or less disappear in the weighted results.

If the transactions survey results are to be weighted by customer type, we recommend that Ofwat considers a banding system to give greater prominence in the overall scores to larger customers (ie the largest spenders, most frequent service users and/or most high-volume builders), in preference to weighting by actual revenue or unit/connection numbers; this could be simpler, and would be more comparable from company to company and hence fairer in terms of the ultimate scoring outcome.

Which band a customer is assigned to can either be determined by the water companies, flagging their transactions records accordingly before delivery (subject to being provided with clear guidance as to how to do this, to ensure consistency of approach), or by Ofwat's appointed agent on review of how often firms appear in the sample records from month to month (this has been put forward by the industry as a means of ensuring objectivity on the classification).

The potential drawback of the second approach is that any individual builder or developer may use different consultants for specific parts of the process – without knowing on whose behalf the consultant is acting, Ofwat's appointed agent would not be able to band these customer contacts appropriately. The way round this is to request both that water companies band their records initially and also append the name of the end-customer in case of consultants. In this way, Ofwat's appointed agent has a starting point for classifying customer type and also a means of verifying the banding by reference to the frequency with which end-customers appear in the monthly samples.

We recommend that Ofwat uses the shadow year to develop an appropriate means by which the transactions data can be weighted by customer size/type, to arrive at a solution that will be simple, clear and workable for all companies.

5.15 League table vs performance improvement

We have seen in section 5.13 that it may be problematic to differentiate between companies in a league table due to variable and/or relatively large confidence intervals in their mean scores, such that the score achieved by a company towards the top of the table may not be statistically significantly different to that of another one further down, even if the sample sizes aimed for over the course of a year represent the largest bases that are practical from the records available. This arises because of the highly variable customer population sizes between the companies, from large to very small.

It may then be worth considering other options for determining whether or not performance or under-performance payments may be applied to a particular company, either instead of or alongside league table positioning. One such possibility would be to set individual company improvement targets from year to year.

Over the course of the shadow year, a robust data set will be gathered for each company. From this, it will be possible to calculate what a statistically significant improvement from the shadow year performance would be, for each company individually. Whether or not this improvement is achieved over the course of the following year(s) could form part of the basis for deciding whether performance or under-performance payments are applied.

As the data builds from month to month, Ofwat's appointed agent would need compile a full analysis file for each company with a view to calculating what the final mean scores, standard deviations and population sizes are likely to be, so that they can assess what a significantly better score for that company would be (i.e. their target). This could either be on the basis of their overall satisfaction score, or potentially one or more of the service specifics where a company achieves under-average satisfaction scores.

The potential drawback to using improvement targets as part of the incentive mechanism, though, is that those companies that are already performing strongly now may find it harder to improve significantly than those towards the bottom, and a fixed improvement target may therefore be de-motivating.

Nonetheless, we recommend that consideration be given to incorporating in the qualitative D-MeX outcome some form of individual improvement target as well as league table position. Potential ways in which this could be done can be tested in the shadow year.

5.16 Budget implications

Assuming that the survey will run on an ongoing monthly basis, our recommended approach for the transactions shadow year is to aim for sample sizes of around 20% of customers for all companies, but boosting the smallest up to 100 in the year (a simple and consistent approach). At this stage, we cannot be sure exactly how many interviews this will generate, as our estimate of total customer contacts over a year, after de-duplicating, of 9 x a single month has still to be tested, but if that estimate works out to be approximately correct, then this could produce a total sample over a year of around **14,000 interviews**. On that basis, the total cost for the shadow year on the transactions survey would be in the region of **£[X] plus VAT**.

This budget indication allows for the fact that there will still be a fair amount of work to do in cleaning/preparing the data submissions from companies each wave. Despite the fact that some have indicated that the review and de-duping process could be relatively straightforward, we know that it will still not be - particularly given that several companies are still only in the early stages of aligning their internal systems to produce this information cleanly. Hopefully over the course of the shadow year, these issues can be ironed out ready for going live the following year. We are therefore also allowing for ongoing conversations with each of the companies, as required, to arrive at a process that is as straightforward as possible at both ends.

The other options discussed in section 5.13, then, would carry the following approximate budgets:

- 15% sampling (ca. 10,500 interviews) - **£[X] plus VAT**
- 10% sampling (ca. 7,100 interviews) - **£[X] plus VAT**
- However many interviews were required to achieve a fixed confidence range around each company's result (if +/-0.2, then ca. 9,000 interviews) - **£[X] plus VAT**
- Fixed numbers of interviews per company, in 3 bands according to size (eg 100, 400 and 800, total ca. 7,300 interviews) - **£[X] plus VAT**

For reference, the fixed costs for the full shadow year, covering project management, monthly sample file cleaning and preparation, monthly results tracking for Ofwat and raw data provision to the water companies, and a final report at the end of the year, are **£[X] plus VAT**. The cost per interview for fieldwork and data processing, depending on volume, is in the region of **£[X] - £[X] plus VAT**.

6 THE RELATIONSHIPS SURVEY

As noted earlier, both waves of the pilot phase included a relationships survey, as companies suggested that a separate survey would be required to examine issues around how the Developer Services departments manage their dealings with key stakeholders in the larger customers they are in regular contact with.

There are a number of reasons why a separate relationships survey may not be pursued:

- It is reasonable to assume that water company effort put into building and maintaining good relationships with key individuals at large customers will show through in how day to day operational service delivery is scored through the transactions survey, such that it is not necessary to seek separate scores from individuals of different types within large customer firms
- There are challenges in identifying and sampling key customer contacts objectively (not insurmountable, but affecting the cost)
- Available budgets may be such that spreading the investment over the two approaches means smaller sample sizes on each
- Where key contacts deal with more than one water company, there are challenges involved in determining the most appropriate focus of the interview (not straightforward, and could be open to challenge if the allocation goes against one company over another)
- And, it would be possible to weight the transactions survey results (as discussed in section 5.14) to give greater emphasis to the feedback obtained from large customers, if that is felt to be appropriate, which would negate the need to carry out additional interviews with this group to ensure their views are adequately represented

In terms of meeting the D-MeX objectives, while a relationships survey would reflect market structures and potentially differing customer needs, it is less practical to implement than the transactions survey, and measuring performance by this means in a way that is consistent, reliable and fair would be complex, such that the associated cost may not be proportionate. The transactions survey alone could meet the objective of encouraging companies to improve customer experiences and innovate.

At the workshop held in December 2018 to report back on the pilot wave 2 findings and present initial shadow year recommendations for discussion, we suggested that D-MeX is not intended as a diagnostic tool but a means of monitoring company performance, and that a transactions survey alone would provide sufficient evidence for this purpose; on the day, this appeared to be generally accepted, but participants were invited to consider all the recommendations again afterwards and submit any further thoughts by email.

Of the 5 water companies who took the opportunity to make any comments on the qualitative measure after the event, one indicated they would have no concerns adopting a purely transactional approach but the other four felt that it would be important to incorporate a separate relationships survey because this boosts the coverage of their larger customers and ensures a full range of customer perspectives (strategic as well as operational) are counted within the measure.

Provided the transactions survey sample sizes are large enough, and a suitable means is found to identify and upweight large customers in the transactions data (as outlined in the previous chapter), then boosting the coverage of these large customers through a separate survey should not be necessary.

The issue around different customer perspectives would be valid, if it were the case that high level strategic contacts with whom a company has an ongoing relationship, and day to day operational contacts for whom a company delivers specific pieces of work, have fundamentally different wants and needs and felt fundamentally differently about how the company works with them. The pilot survey evidence indicates that this is not in fact the case. Major issues across the board relate to turnaround times, meeting deadlines, effective communication, commercial awareness of the customer's needs and proactivity in meeting these. The mean average overall satisfaction scores accorded each company in the two separate surveys have been broadly similar.

In our view, the transactions survey is the core requirement of the D-MeX customer satisfaction tracking programme. A relationships survey would give further weight within the programme to larger developers, SLPs and NAVs, and would capture a wider range of customer perspectives, but as these features do not significantly impact the shape of the outcome, then the additional cost associated with the relationships survey may not be justifiable. On balance, we believe that a transactions survey alone will be sufficient to achieve Ofwat's objective of a simple, proportionate and practical mechanism that measures performance fairly and encourages companies to improve.

The companies may continue to carry out their own diagnostic customer satisfaction programmes with key stakeholders, and use the results to inform improvement in their operational service delivery, which should then feed through to movement in the D-MeX transactions survey scores.

The following sections deal with how the relationships survey could be developed, if it should be decided for further testing purposes to include this in the D-MeX shadow year.

6.1 **Sampling**

At Wave 1 of the pilot, the water companies were asked to provide lists of the key customers they are in regular contact with. These lists comprised developers, consultants, SLPs and a small number of NAVs. After cleaning and de-duplication between company lists, a total of 2,345 usable records remained, from which we achieved 452 interviews, with a response rate (interviews as a proportion of interviews plus refusals) of 59% and conversion rate (number of records used to complete one interview) of 4.1.

The survey results reflected a spread of opinion on the companies, with respondents giving constructive criticism as well as positive feedback. Nonetheless, we recognised that if these interviews are sampled from lists provided by the companies in future, there could be an incentive to filter which customers are passed through and which not, and we would have no way of auditing the representativeness of the sample.

For Wave 2, then, we took a different approach and sampled all of the relationships interviews independently, as follows.

For large developers

- Contact details in the transactions data are often a site or office based individual. The relationships survey is intended to target Technical Directors or equivalent, who have responsibility for the development lead-in process and can therefore comment on ongoing working relationships with utility suppliers. We set out to identify relevant large developer organisations in the transactions data, then independently seek out and approach their Technical Directors (as well as including their site/office contacts in the transactions survey)
- At a D-MeX sub-group meeting, it was suggested that companies flag their transactions records as either 'on site' or 'off site' to distinguish between customer organisations involved in projects that require new infrastructure ('on site' - likely to be large developers, hence warranting a relationship interview) and those for whom projects simply involve connections to the existing network ('off site'). The Wave 2 guidance therefore requested that transactions data be flagged in this way
- However, records flagged as 'on site' were clearly not all large developers, and some well-known large developers were flagged as 'off site', so this method of identifying appropriate customer organisations to target for a relationships survey appeared flawed. In addition, not all companies acted upon the classification request in the guidance. This approach would therefore not be reliable; the Wave 2 schedule did not allow time to go back to the companies to clarify or re-run their data, so instead we purchased PropertyData's House Builder Directory, identified the top 100 developers in their database (by turnover and/or unit completions), and flagged these ourselves in the transactions data, to approach for a relationships interview

- As a starting point for a relevant individual to speak to, we took the Technical Director or nearest equivalent from PropertyData's contacts database – then approached these by phone and took referrals if appropriate (within the same region, where applicable)
- After de-duplicating contacts (at random), we were able to load a total of 185 records - head or regional offices known to have worked with a company in August. From this list we achieved a total of 97 interviews with a response rate of 84% and a conversion rate of 1.9

For SLPs

- 6 of the target Water UK metrics relate specifically to SLP work. Some companies flagged customer contacts in their transactions data as SLP. These two indications could have enabled us to pull out SLP customers from the rest, to approach them separately
- However, some records flagged as SLP contacts were not, and vice versa, so this approach would not be reliable. Instead, we identified registered SLPs using Lloyd's Register's WIRS database, available online, and flagged these ourselves in the transactions data to approach for the SLP surveys, removing them from the transactions sample this Wave
- We then took the key contact listed in the WIRS database, approached these by phone and took referrals if appropriate
- After de-duping (at random), we were able to load a total of 31 records for relationships - SLPs known to have worked with a company in August. From this list we achieved a total of 10 interviews with a response rate of 59% and a conversion rate of 3.1

For NAVs

- From Ofwat's NAV register, we identified 6 NAV organisations that are not themselves linked to an incumbent water company, found a key contact at each (from Ofwat's records and/or the Water UK council membership list available on their website) and contacted these to carry out a relationships survey
- NAV circumstances and needs are quite different to those of builders, developers and SLPs. In order properly to understand their perspectives, the relationships interviews in this segment were carried out by means of exploratory depth interviews by phone rather than a strictly structured interview
- Interviews were achieved with all 6

The process adopted at Wave 2 was relatively time-consuming. We would not recommend repeating this on a regular basis as many of the same developer and SLP companies would appear in the transactions data each month and the costs of managing and de-duplicating the sample each time would be disproportionate to the outcomes. Instead, we would suggest that the approach we outline in the next section would be a more effective solution.

6.2 **An annual industry survey**

If Ofwat decides to run a relationships survey in the shadow year, then in order to meet the objective of independent sampling and to ensure the process is practical and cost-effective, we recommend de-coupling the relationships survey from the transactions data and carrying out an annual phone survey of key contacts at all large developers, SLPs and NAVs, using lists sourced by Ofwat's appointed agent.

Despite the fact that the Wave 2 response rates were very high, it would be important to avoid asking these individuals for help with the survey too often. In addition, as the focus of the interview is on the relationship and not on a specific piece of work, it is unlikely that perceptions would change a great deal over short periods of time. An annual approach would therefore be appropriate.

Sample sources to be used could be:

- For large developers - PropertyData's House Builder Directory (which lists developers by size and contains contact details for key individuals - regionally where applicable), or similar. This is the most comprehensive source we have identified so far, although there may be others we have not yet come across (To purchase)
- For SLPs - Lloyd's Register's WIRS database (Available online)
- For NAVs - Ofwat's NAV register (Available online)

This just leaves a potential gap in terms of large consultants and agents employed by developers, that water companies will have ongoing relationships with. If it is felt necessary, it would be possible to explore whether there are any suitable sample sources that could be used for this segment.

The nature of the interview in this case, discussing the quality of the relationship that the individual customer has with the water company in question, points to a conversation over the phone being the most appropriate method.

6.3 **Questions**

In the same way as we have discussed in relation to the transactions survey, we suggest that the overall satisfaction rating be used as the question to base the D-MeX incentive on, if it is decided to run a relationships survey in the shadow year, rather than any amalgamation of scores on specific service factors.

At Wave 2, large developers were simply asked about the one water company from whose transactions data the customer firm was picked out. SLPs and NAVs were asked which companies they deal with at all and of these, which they deal with most - the interview focused primarily on the main company dealt with, but respondents were also asked for an overall satisfaction score for each of the other companies they have experience of (in order to increase the number of data points obtained).

This same approach could be used with the large developers, in order to avoid any potential issues over our selecting which company they primarily talk about. While this does limit the extent to which sample sizes by water company can be anticipated in advance and controlled, the relevance of the data provided by each respondent would be stronger.

The interview could therefore run along the following lines:

1. What would you say are the top 3 most important things you personally are looking for, from a water and sewerage company's Developer Services team? (open)
2. Which of the water companies do you personally have any dealings with? [If more than one:] I'd like to focus the rest of the interview on a single water company, so that we can collect specific feedback on industry performance. Which one of these would you say you deal with most?
3. How satisfied are you with the following aspects of the service you receive from Developer Services at [WATER COMPANY]? (rating scales):

Business dealings	a) Their responsiveness to your business needs and priorities b) Ease of doing business with them
Contact	c) Frequency of contact d) Contact at an appropriate level of seniority or skill e) Ease of contacting them when you want to
Customer Service	f) The extent to which they take a proactive approach to collaborating with you g) Timeliness of response to queries and requests h) The rapport you have with the staff you deal with i) Getting things right first time j) The quality of the information available on their website k) How joined-up the service is, between different parts of the organisation l) Consistency of approach, between different individual members of staff
Pricing	m) How well they explain their charging arrangements n) How reasonable their charges are
Work stages [DEVELOPERS AND SLPs]	o) Their technical expertise p) Their responsiveness to fitting in with your work programmes q) The time taken to complete processes r) Their commitment to meeting deadlines s) [SLPs ONLY] Ease of obtaining adoption agreements

<p>Work stages [NAVs]</p>	<p>t) The clarity and ease of use of their published NAV connection/discharge application framework</p> <p>u) Adherence to their published service levels in the framework</p> <p>v) Their model for serving you (e.g. whether you have a single point of contact who is empowered to resolve your queries; their flexibility)</p> <p>w) The flexibility of their approach to pricing, where services change</p> <p>x) Their complaints and dispute resolution process</p> <p>y) Providing you with sufficient information to enable you to self-serve, where appropriate</p>
--------------------------------------	---

4. How often are you personally in contact with them? (pre-coded list)
5. Can you suggest any ways in which their customer service could be improved?
6. **[DEVELOPERS AND SLPs ONLY]** Which specific work stages or processes, if any, would you like to see improved, and in what way(s)?
7. **[NAVs ONLY]** Which elements of their approach to dealing specifically with NAV customers, if any, would you like to see improved, and in what way(s)?
8. How satisfied are you with your experience of dealing with Developer Services at [WATER COMPANY], overall? (rating scale)
9. What do you feel they do particularly well, if anything? (open)
10. And if there was one thing that you'd suggest that the Developer Services team at [WATER COMPANY] could do, that would have most impact on the relationship you have with them, what would it be? (open)
11. And for comparison, could you give me an overall satisfaction score for [each of] the other water [company/companies] you deal with? So, using our scale of 0-10, how satisfied are you overall with your experience of dealing with Developer Services at [WATER COMPANY 2]? And [WATER COMPANY 3]? Etc.
12. Respondent name and job title
13. Whilst this survey is being carried out on behalf of Ofwat, [WATER COMPANY] is keen to learn from your feedback and improve its customer service. With this in mind are you willing for us to pass your answers back to them, identified by name?
14. That is the end of the interview. Are there any other comments you would like to make, before we close?

6.4 Sample size

From a review of the Wave 2 transactions records that had been flagged as on site or off site, around a third of customers had been classified by companies completing this exercise as on site (hence large by implication). This figure is not verifiable by reference to any actual size information, and in any case it will include multiple contacts at any given firm.

At this point, then, we cannot firm up how many interviews an annual relationships survey could be likely to comprise, but for the sake of an estimate for the shadow year we could say the number of relationships interviews per company could come in at 10% of their annual transactions sample, so 1,200 interviews in total on our initial transactions option of 20% described in section 5.13 above (this level looks to be feasible, and again would be a consistent approach for all companies).

6.5 Combining the results from the two surveys

In order to arrive at a single qualitative score for D-MeX, the results of the transactions and relationships surveys would need to be combined in some way. How this is done may depend on whether or not the transactions survey data are weighted according to customer value as described earlier.

Adding the two data sets together would be the simplest approach, perhaps weighting the relationships scores with the same value applied to the large customers in the transactions data (for example multiplying by 3 if the 1,2,3 banding scheme suggested in section 5.15 were employed). The potential issue here is that the largest customers of each water company could end up over-represented in the overall score.

We would suggest that it would be appropriate to test out weighting options for the relationships survey scores within the overall qualitative measure, in the shadow year, if this element of the survey programme is carried out.

6.6 Budget implications

On the basis of the possible sample size of 1,200 interviews outlined in section 6.4, the total cost for the shadow year relationships element would be in the region of **£[X] plus VAT**.

APPENDIX:

Responses received to Ofwat's request for any further feedback, and information on quality assurance processes, following the pilot wave 2 workshop

POST-WORKSHOP RESPONSES, DECEMBER 2018

A D-MeX workshop was held on 4 December 2018 to report back on Pilot Wave 2 and discuss possible outline recommendations for the shadow year. Following this, attendees were asked to provide any further feedback on the options and recommendations presented; and, in the case of the water companies, to set out the steps they take to quality assure the accuracy of the underlying data that they submit to Water UK for the Levels of Service reporting (which was raised as a potential concern at the meeting during discussions about the quantitative element of D-MeX).

RESPONSES WERE RECEIVED FROM:

8 water and sewerage companies (WaSCs), 1 water only company (WoC) and three customer representatives.

SUMMARY OF POINTS MADE, AND OUR RESPONSE (IN BOXES): (THE QUALITY ASSURANCE INFORMATION PROVIDED IS SHOWN IN FULL, LATER)

[X]

Described how their online portal has been developed to allow reporting on Water UK and D-MeX; confident that data are accurate.

Of the options presented at the workshop, preference is for option 1 (replicating pilot wave 2), saying this is auditable and should flag missing information. Preference for monthly data submissions to mirror Water UK but quarterly fieldwork to avoid survey fatigue. Regarding metrics to be used, preference is for a sub-set, identified by reference to customer priorities, as not all carry equal weight.

Ofwat will determine if a separate relationships survey (included in Option 1) is appropriate. Monthly fieldwork will maximise potential sample sizes for the smaller companies; the risk of survey fatigue can be managed through effective sample management. For D-MeX customer survey sampling, using all of the relevant metrics ensures a comprehensive data.

[X]

Submitted quality assurance documents detailing approach to Water UK process and reporting. No feedback on workshop content or recommendations.

[X]

Forwarded a copy of a letter written to Ofwat outlining differences between the regulatory regime in England and Wales that could impact on a straight comparison in D-MeX between [X] and the English companies. A key point raised is that all new sewers and lateral drains that will connect to the existing public system in Wales must now be adopted, and that customers do not welcome having to comply with these new requirements; this leads to a greater proportion of the DS work in Wales relating to waste, and potentially lower satisfaction scores in Wales for the sewer adoption

services provided by [X]. The requirement in Wales for new houses to have sprinklers installed (if mains-fed, as developers prefer, the water pressure needs to be higher than the existing minimum requirement), and [X]'s role as a Statutory Consultee in the planning process may also negatively impact customer satisfaction scores compared to companies in England.

Further concerns expressed about D-MeX as currently conceived:

- Customers wish to see greater consistency between companies in the service provided but D-MeX may discourage the collaboration that could achieve this aim
- Variations between WASCs and WOCs in the services provided, together with variation in the local markets of each company (in terms of the extent of SLP and NAV activity, as well as the urban/rural mix that will affect the nature of the developments in their area) mean that direct comparison of results may not be fair
- And therefore a solution could be to set an improvement target for each company, based on plans tailored to their own specific market, rather than reward and penalise on the basis of position within a league table

These are fair points to raise, and the pilot data does not allow us to judge either way whether any differences in customer satisfaction scores between Wales and England, between WASCs and WOCs and between those that do and those that do not have, for example, significant SLP activity in their area, could be down to these issues. Ofwat has, however, indicated that 'While we recognise that national and regional differences exist we are not minded to apply company-specific or other adjustments to D-MeX, to avoid undue complexity'.

We have suggested that further thought be given to including some performance improvement measure within D-MeX, along with direct comparison between companies, and this will be for Ofwat to decide.

[X]

Described their data collation, review and analysis process for monthly submissions, and internal audit and improvement cycle.

No comments on workshop content or recommendations – felt to be comprehensive and valuable.

[X]

Briefly described a documented methodology, training and compliance check process.

Feedback on the recommendations:

- No concerns adopting a purely transactional approach
- Balance required in surveying all stages of the customer journey, and fair distribution/weighting in terms of service and customer type
- Predefined template for data submissions needed – happy to help develop this, based on the pilot wave 2 version
- 10-0 scale preferable to 5-point verbal, for differentiation
- Phone surveys preferable to email – more personable and representative

- Overall satisfaction as the key question for measurement is preferable, asked towards the end of the interview – simple for companies and customers
- Critical that survey is representative for the smaller companies. As [X] is so very small, suggests taking them out of the results tables
- Suggests a rationalised list of metrics to account for key aspects of the journey:
 - Water: Pre-Development Enquiry – Design – Quote – Construction (source of water, mains requisitions, connect) – Pay (Asset Value Payments)
 - Waste: Pre-Development Enquiry – Design Approval – Quote – Construction – Adoption
- Chosen metrics should be closely aligned to existing due to implications on systems of making changes
- To avoid survey fatigue, any one customer should not be approached more than twice a year
- Preference against monthly sampling and data provision, due to resource required and possible survey fatigue
- Equal weighting between qual and quant measures is appropriate

Would welcome clarification on the methodology as early as possible to enable effective system configuration. Would also welcome clarity on how the financial aspects of the mechanism will work

Most points here indicate agreement with the recommendations made at the workshop.

It is unlikely that Ofwat would exclude [X] entirely from D-MeX, although how they should be incentivised (along with or separately to [X]), given their extremely small footprint, will be for Ofwat to determine.

The metrics to be used in transactions survey sample selection should be comprehensive; those used in defining the quant measure will be for Ofwat to decide.

Monthly sampling is preferable if coverage of the smaller companies is to be representative (which is accepted in the feedback as critical) and survey fatigue will be avoided through sample management – although we appreciate the manual effort that will be required for some companies while new data extraction systems are put in place.

[X]

Gave a brief statement on quality assurance processes and offered to provide copies of procedures if we want them.

For D-MeX, keen that both relationships and transactions are included, while accepting this is extra work; an annual relationships survey would demonstrate that companies are building relationships with regular developers for the longer term, which is 'a very important part of delivering DS'.

Keen to understand urgently the data required from April to enable them to amend systems to automate transactions data collection as opposed to doing this manually.

We have outlined the ways in which a separate relationships survey may be seen as valuable to the industry, and the reasons why it may not be necessary to include one in D-MeX. Whether or not to have a separate relationships survey within D-MeX will be for Ofwat to determine.

[X]

Submitted documentation that had been provided to [X], outlining their work management system, and commented that the company is currently developing a replacement system.

No feedback on workshop content or recommendations.

[X]

Described levels of responsibility for data review and approval. Currently undertaking work to ensure staff understand the metrics definitions and the importance of accurate data.

Feedback on the recommendations:

- Sampling transactions as they fall out naturally seems the fairest way to apply the survey across all companies
- Provided the relationships survey remains, to ensure larger segments have sufficient representation (and therefore drive behaviour to focus on improving service across these segments)
- As the number of NAVs is very small and some companies have no NAV customers, suggest not adopting a specific approach or weighting for these but roll them together with SLPs (as the service provided to each is similar)
- Suggest 50/50 weighting between qual and quant, and within qual, 50/50 between transactions (falling out naturally) and relationships (50/50 large developers and SLP/NAV)
- To identify sample for relationships, either use PropertyData as at pilot wave 2 or seek lists from companies of those spending above a specified amount or building above a set number of units (which would vary from company to company) – no preference expressed. Recognises this means consultants would not be captured
- Overall satisfaction preferred over indices, for simplicity (but interested in knowing more about the correlation method)
- Appropriate contact to survey is likely always to be the applicant for small customers but may be a technical person or site contact (when it comes to construction) at larger ones. [X] will proceed to collect applicant details for the quotation stage and site contact details for construction stage, in anticipation
- Suggests monthly transactions data and fieldwork to ensure sufficient sample for small companies, provide parity of approach and avoid any seasonal effects. De-duplication should avoid skewing of the data by service
- Suggest relationships carried out no more than 6 monthly
- Sufficient updates and data through the shadow year should be provided that the system is tested and reviewed throughout
- SLPs and NAVs should score all companies dealt with, as at pilot wave 2
- For the quant measure, a subset of the metrics should be used; the volume weighted average (option 2) would not be appropriate as this would over-represent small customers

Most points here indicate agreement with the recommendations made at the workshop.

We have outlined the ways in which a separate relationships survey may be seen as valuable to the industry, and the reasons why it may not be necessary to include one in D-MeX. Whether or not to have a separate relationships survey within D-MeX will be for Ofwat to determine.

[X]

Described the extraction and review of data from their case management system; once D-MeX requirements are finalised the process will be subject to board assurance external audits.

Suggested an annual relationships survey; feedback from their key stakeholders says they would like to be included, 'with the aim for a focus on an overall improvement across all water companies'.

We have outlined the ways in which a separate relationships survey may be seen as valuable to the industry, and the reasons why it may not be necessary to include one in D-MeX. Whether or not to have a separate relationships survey within D-MeX will be for Ofwat to determine.

[X]

Detailed feedback provided on the recommendations:

- It cannot be for companies to decide what works for them in structuring D-MeX
- Proposals are too heavily skewed towards customers who just want water connections. A means needs to be found to ensure customer input is proportionate to the sector they represent
- Customers may fall into a wide range of segments. (Over 25 units a year may be a suitable threshold to differentiate against 'one-off' applicants)
- Self-lay jobs are a very small number so may not trigger many transactional interviews
- Sewerage accounts for less than 10% of total connection jobs – with de-duping, sewerage bases could be small
- Need to continue with relationships survey to mitigate against small samples of important segments in transactions
- Regularly surveying customers that account for a significant amount of work may be unsustainable
- Results would be skewed for the likes of [X] if SLPs are surveyed on a project basis as the bulk of their scores would relate to easy one-offs or tricky highways
- Customers have misgivings about the metrics; an independent audit is required
- A sub-set of metrics needs to be used to exclude those of a purely administrative nature
- The metrics were introduced by the companies with no customer input. They are not challenging so do not differentiate
- Results should be weighted to reflect the value of the work (not the cost) – eg a main for 50 properties should be weighted at 50 times one on-site connection, and services in the highway (more costly to construct) should be upweighted
- The scoring mechanism should categorise companies in bands rather than a league table

The potential skew towards high volume low value customers will be addressed through the proposed customer weighting system to be developed through the shadow year.

Self-lay and sewerage jobs will indeed account for small volumes within the total, but taking an approach that is representative of each company's operational workload will most closely reflect their performance overall. Our recommendations to include all SLP metrics within the sample selection process, and monthly data extraction which will maximise the bases for the smaller metrics, will ensure as wide coverage as possible across all customer groups and activities.

The transaction survey questionnaire asks for feedback on what customers see as improvement priorities for the company under discussion, which gives the opportunity to comment on service more widely than simply the transaction in focus, for example for an SLP being surveyed about an easy one-off job who wishes to raise a wider problem they have

Whether or not to include a separate relationships survey within D-MeX will be for Ofwat to decide. Concern over the possibility that such a survey may not be sustainable could form part of their reasoning in coming to their decision.

We have already excluded metrics of an administrative nature, in the sample selection approach.

It does appear that the Water UK metrics are not challenging, and this has implications for Ofwat's thinking around the nature of the quant measure. The satisfaction survey is an entirely separate exercise, generating outcomes that are based entirely on customer input.

[X]

Expressed concern that Ofwat is keener on 'appeasing' the water companies than designing a measure 'driven by the needs of the customer'. Described working on OFGEM's GSOP (Guaranteed Standards of Performance) to improve energy companies' customer service in this sector, that ended up as a set of KPIs (with fines for failure) and no survey element. Believes that time and money put into customer surveying in D-MeX could be better spent.

Ofwat have already determined that D-MeX will be based in large part on a customer satisfaction survey programme. Our work on the pilot waves has demonstrated that such a survey programme is feasible, and capable of generating robust scoring data and customer comment. Our recommendations for building on the pilot work through the shadow year will enable Ofwat to test outstanding issues and come to a position on the ultimate survey design.

In our view, any systematic and objective programme of asking customers what they think of the service they're getting, publishing the results, comparing companies against each other and distributing rewards and penalties as a direct consequence, cannot fail to focus companies' minds on better aligning how they deliver their service with what customers need.

[X]

Expressed concern that D-MeX is constrained by companies' ability to collect and record data and does not truly represent the quality of service provided or provide assurance to customers that service will improve. Recognises however that the measure has to be implementable so confines comments to the recommendations as they stand:

- Current LoS metrics are not comprehensive (eg missing NAV bulk supply agreements); use existing metrics but expand to cover all aspects of connections process
- Concerned about quality and consistency of company data and welcomes Ofwat's request for systems information. Water UK should clarify areas where there is inconsistency

- For the quant measure, suggests weighting each sector (option 3); a consistent scaling factor across all types of customer would be number of premises to connect
- Suggests simple 50/50 weighting between qual and quant measures
- For the qual measure approach – suggests either option 1 (replicate pilot wave 2) or option 3 ([X])’s proposal for a survey of small customers and panel of large ones); preference for option 3 with a firm commitment to ensure representative sampling of all customer types

We have commented in our final report on the request for relevant NAV metrics to be include in the Water UK set, and recommended that a way be found in the shadow year to obtain details of NAV activity from which transactions interviews can be sampled.

We note the concern over the quality and consistency of company data currently reported. Using the Water UK metrics as the point of reference for sampling the transactions satisfaction survey remains the most practical way to structure a programme to obtain customer feedback that represents the spread of work being undertaken by the water company Developer Services teams.

RESPONSES REGARDING QUALITY ASSURANCE PROCESSES, IN FULL:

[X]

As [X] has an online portal system, we have the ability to create reports that pull data from information submitted/entered on our system. When an application has been submitted or uploaded a ‘task’ to review the application is auto generated. This is repeated at each stage of the application and allows us to track and manage each application.

We use our system to report on ‘tasks’ that are built into the system, which are linked to each Water UK metric. This allows us to break down the report into each metric, see the total number of ‘tasks’ completed within a calendar month and the time taken to complete from the start date. This will then display as a pass or fail for each ‘task’ depending on the time taken to complete.

Once the reports are generated, we review the information presented. This allows us to quality assure the accuracy of the data we are submitting to Water UK.

To ensure our D-MeX data is in line with our Water UK reporting, we have mirrored the report we use for Water UK with the additional fields built in to gather requirements for D-MeX reporting e.g. application contact details, customer group etc. We also carry out data cleansing on a monthly basis, so are we able to have confidence on the information we are reporting on.

Once the report has been produced we will check through the lines of data with the aim of removing any fields containing errors, such as contact numbers displayed as ‘0777777777’. We cross check our communication log and replace information, such as the example above, with information we are sure is valued.

Based on the way our system is built, as well as the reporting linked in, I have complete confidence that both [X] Water UK and D-MeX reporting is accurate.

[X]

Please see the attached quality assurance documents for [X] detailing our approach to Water UK process and reporting. [6 detailed process documents attached, for different work types]

[X]

You also asked us to confirm the governance we have around the collation and submission of our Water UK Developers' Charter data and I hope the following is sufficient.

- Our starting position is that since we started reporting this data 3 years ago we have maintained consistency by using a well-trained data analyst who is not only an expert in data analysis and our customer contact and work management systems but also has previous experience of working within the Developer Services team.
- Once the raw data is collated there is peer review between the relevant team leaders across all of the departments responsible for service delivery which allows internal challenge of the results. Where there is any doubt whatsoever on whether something was within level of service or not, we count it as a fail. An example would be where there is evidence that a conversation has taken place with a developer customer to agree an alternative date but there is no written evidence we count that as a fail. We also use this to identify root cause.
- We also carry out an analytical review of the reporting from the previous quarter to ensure that results are as expected given any operational issues that may have happened in the quarter.
- The data is then signed off by a senior manager before submission

Our Internal Audit team have carried out reviews of the processes in this area to ensure that we are compliant with the relevant guidelines and that our processes are robust and capable of producing reliable data that can be agreed to underlying records. This audits involve re-performing our reporting using base data. Any recommendations made to strengthen processes are implemented.

Internal Audit plan to carry out a further review in early 2019.

In addition to this we were part of the Water UK Horizontal Audit and our processes were found to be robust with zero major non-conformities. There were some recommendations on what was seen as best practice in other companies and we have, where appropriate, sought to introduce these improvements.

[X]

In terms of quality assurance our regulatory submissions include:

- A documented and transparent methodology
- Documented internal training/guidance
- 1st and 2nd line assurance to include: compliance with requirements, outcome validation, sampling and volume trending prior to Head of Department sign off

[X]

To quality assure our data we have detailed procedures for each line element of the Levels of Service. These identify a data producer, data reviewer and data sign off. All of these are different people and set at levels of Management within the business. On top of this we have internal audit to review our data on their audit programme. Levels of service is also added to our external audit review undertaken once a year. If you require copies of our procedures, please let us know.

[X]

Not sure of the detail you wish for me to go into regarding our data surety, so if the below is not sufficient please just let me know.

Attached is a document, which we provided (along with the process maps) to the independent auditors ([X]) which conducted the Water UK audit into LoS data provision. This provides an outline of the work management system we currently utilise, we are in the process of developing a replacement system.

With our online application portal and work management system our transactional items are time stamped by the system throughout the process (no manual input to this aspect) and the reports are auto generated to email out on a weekly and monthly basis. We also have a dashboard to track applications through the process stages which updates ever 30 mins, so we are able to see approaching deadlines.

Some LoS information is from our construction PMs and contractors are issued from P6 and their performance trackers.

The information is first checked in its entirety by our support services partner ([X]) through a 2 stage check and review, a quality and performance manager and then a senior manager, and then spot checked by a senior manager in [X] Developer Services prior to submission to Water UK.

[X]

We have a routine monthly governance cycle for our Water UK metrics.

Each measure has a responsible manager who has to check the raw data for each measure to ensure themselves that they are happy that the rules for that measure have been appropriately applied. The collective measures for Water and Waste are then reviewed and approved by an accountable manager within Developer Services before the entire submission is signed off by either the head of Developer Services or delegate prior to submission to Water UK.

We are also in the process of doing a deep dive on all of our measures to ensure that all individuals involved in the preparation and review of data understand the definitions of the metrics and the importance of accuracy as we move towards their inclusion in some form within DMEx.

[X]

In regards to our steps for to quality assure the accuracy of the data that we submit to Water UK, I have detailed these below:

A performance analyst extracts the SLA data from the case management system.

Data is sent to relevant process owners to validate their data by checking a sample of work. If any errors are found, larger samples are checked to determine if the completed items need to be reviewed and ensure that we are reporting correctly.

Results of the data checks are sent to the Head of Developer Services, and if any issues are highlighted we have additional governance to validate the data by undertaking further sample checks.

We also have internal audits for additional governance and compliance to ensure the data provided is reliable accurate and complete.

Once this becomes part of our ODI, this will be subject to regulatory reporting procedures which include board assurance external audits.