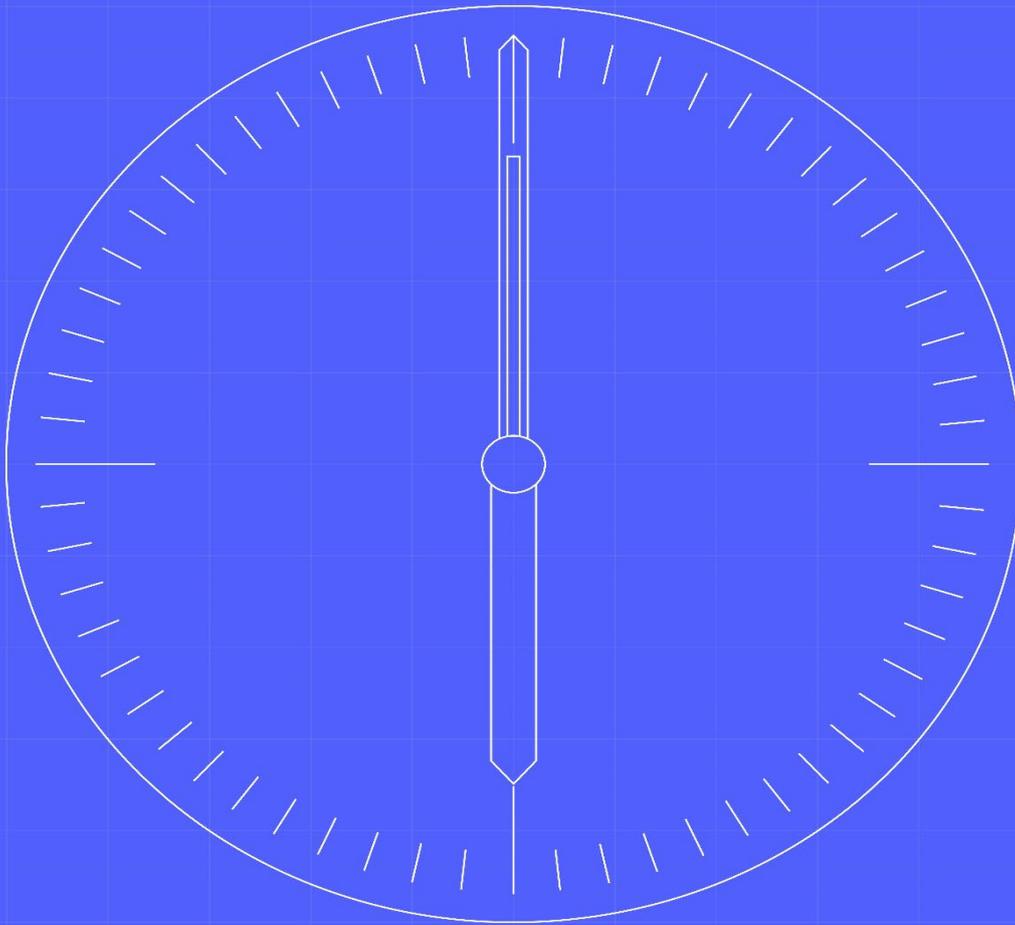




Test Data Strategy



Document owner
Kate Goodman
Status:
Approved

Document number
MHHS-DEL300
Date
18 May 2022

Version
1.0
Classification
Public



1 Contents

1 Contents	1
1.1 Change Record	2
1.2 Reviewers	2
1.3 References	2
1.4 Terminology	2
2 Introduction	3
3 Approach according to test phase	3
4 SIT & UIT approach according to data domain	4
4.1 Scope of testing and data involved	4
4.2 Data domains	4
4.3 Registration data and Meter Technical Details	5
4.4 Consumption data	5
4.5 Schedules	6
4.6 Industry standing data	6
4.7 Load shapes	6
4.8 BSC settlement data	6
4.9 Summary of responsibility according to data domain	6
5 Other general principles	7
6 Data Protection Impact Assessment	8
6.1 Introduction	8
6.2 Points for consideration in the DPIA	8
7 Test Data Plans	8

1.1 Change Record

Date	Author	Version	Change Detail
18 May 2022	Kate Goodman	1.0	Issued in first Approved version

1.2 Reviewers

Reviewer	Role
Not applicable	

1.3 References

Document/Link	Publisher	Published	Additional Information
MHHS-DEL315 - E2E Testing & Integration Strategy	MHHS	29 April 2022	Approved

1.4 Terminology

Term	Description
	For terminology, see programme glossary on the MHHS portal: Programme Glossary (sharepoint.com)

2 Introduction

Climate change presents the world with a challenge to urgently protect our planet, our economy and our populations.

The UK has set a world-leading net zero target, the first major economy to do so. Having set the target, the focus turns to delivering against that target. In the Energy White Paper 'Powering our Net Zero Future' the government has set out an ambitious programme to build a fairer, greener energy system.

Market-wide Half-Hourly Settlement (MHHS) is a key enabler of the flexibility to support the transition to Net Zero. The MHHS Programme will contribute to a more cost-effective electricity system, encouraging more flexible use of energy and helping consumers lower their bills.

Setting the strategy for end-to-end (E2E) testing and the data to be used in that testing is the responsibility of the MHHS programme and in particular of the Systems Integration (SI) team within the programme. This document has been prepared by the SI with the objective of defining fundamental principles regarding the sourcing, management and security for test data in the MHHS programme. It should be read in conjunction with the MHHS-DEL315 - E2E Testing & Integration Strategy. One of the overriding goals of the exercise of planning data management is to ensure that the data maintains its integrity across the various systems making up the MHHS arrangements.

This Strategy document does not include indication of dates or other aspects of planning. When the E2E MHHS Design has been signed off, this document will be updated if necessary. At this point, the full definition of the E2E data model will be available and an Overall Test Data Approach will be developed. This will identify all the data elements that need to be considered and define how they are sourced and made consistent across the different systems. A more detailed Test Data Plan will also be produced for each test stage at a later point in the programme. This strategy covers the data required to be co-ordinated across the systems for testing purposes. It is not intended to cover data elements which are relevant only to an individual PP, for which the PP is responsible.

3 Approach according to test phase

The MHHS testing is divided into test phases, each of which comprises one or more test stages:

Test Phase	Test Stage
Pre-Integration Testing (PIT)	Pre-Integration Testing
Systems Integration Testing (SIT)	Component Integration Testing
	Functional SIT
	Non-functional SIT
	Operational Testing
User Integration Testing (UIT)	Migration Testing
	Qualification Testing
	E2E Sandbox Testing

Responsibility for the test data differs between the test phases.

For PIT, which takes place on the Programme Participant's (PP's) own standalone test environment, responsibility for planning, specifying, generating, managing and securing the test data lies with the PP. The MHHS IM will however provide the PPs with test harnesses (test software) to generate realistic consumption data. For full details of test harnesses, see MHHS-DEL315 - E2E Testing & Integration Strategy. No harness will be provided for injecting the data into participants' systems because it is assumed that each participant has its own such test software.

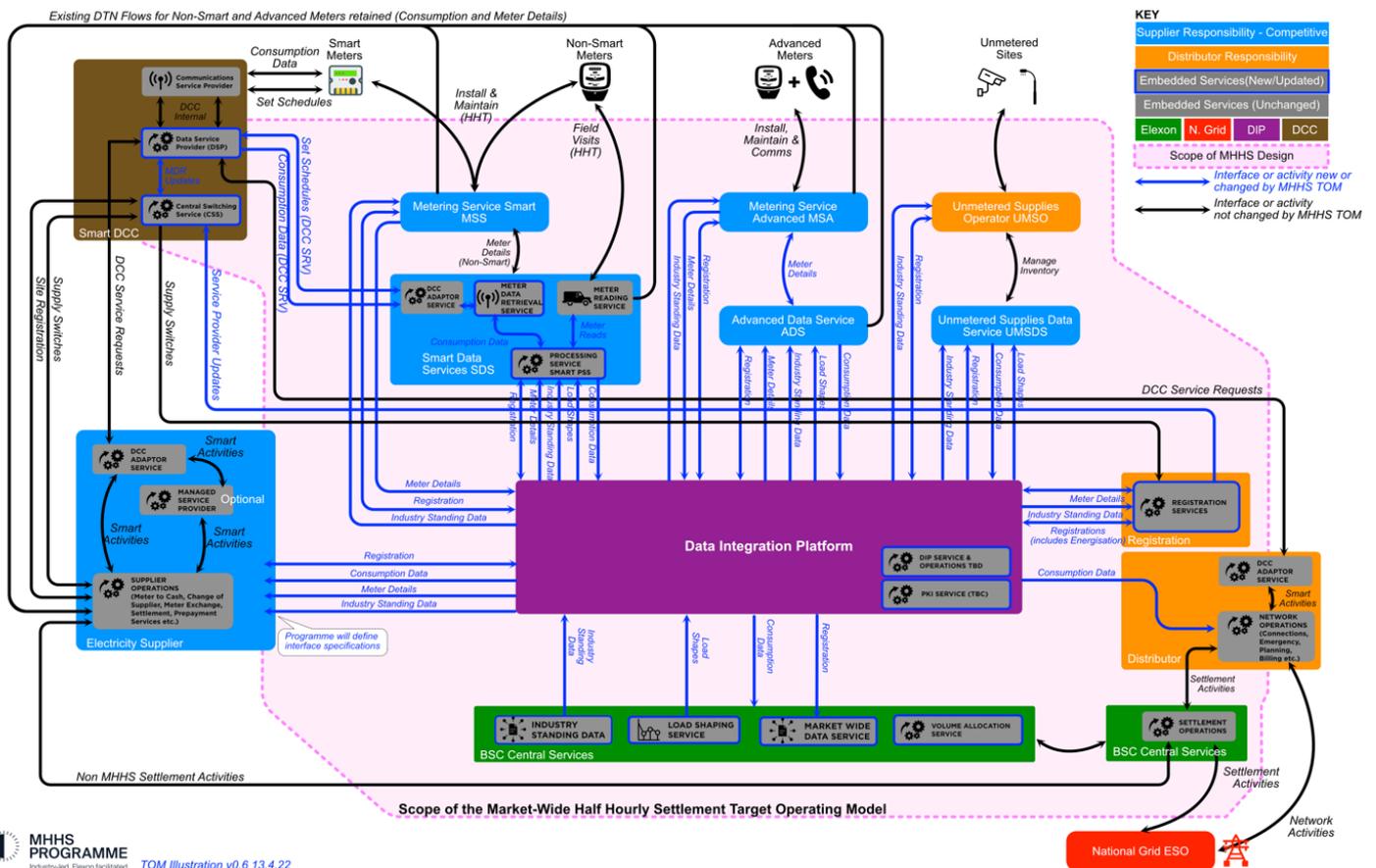
SIT and UIT both involve multiple parties testing on integrated environments. For these test phases, the MHHS IM is responsible for planning, co-ordinating and managing the test data and for allocating MPANs to each PP for their testing. The MHHS IM SI team will request the DNOs and IDNOs to take a cut of live data to be used for testing. There will be a single cut of data and in principle the same cut of data will be used for both SIT and UIT to reduce the effort associated with taking multiple data cuts. (We do however intend to provide a decision point where the programme may decide to take an additional data cut because the original data has become less usable over time.) The SI team

will allocate data to each PP on each test environment (see below for details) and will ensure that every PP has an adequate volume and diversity of data with which to test. Each PP will be responsible for the security of the data on its own systems and we assume that all will take full responsibility for this, taking account of Data Protection requirements. It is crucial that each PP uses only that data allocated by the SI and the SI may require the PPs to sign an undertaking to that effect.

4 SIT & UIT approach according to data domain

4.1 Scope of testing and data involved

The following diagram shows the scope of testing and the main data flows between the components. The “scope of MHHS design” shown by the pink-shaded shape also defines the scope of testing.



MHHS PROGRAMME
Industry-led, Elexon facilitated
TOM Illustration v0.6 13.4.22

The data described in this Strategy is that which needs to be aligned across multiple systems or is needed for settlement processing. MPANs and their associated registration and meter technical details need to be aligned and consumption data is needed for settlement so these are all covered. Where certain data is specific to a given system and does not need alignment outside that system/is not required for settlement purposes, it is assumed that the organisation responsible for that system will supply the necessary data. An example of this would be customer data in a supplier’s system, which the supplier would be expected to furnish.

It is intended that sufficient data will be pre-populated on all the systems before testing starts. It is not the intention that all MPANs will be treated as new registrations.

4.2 Data domains

As described in the diagram above, the data to be used in testing MHHS in SIT and UIT broadly falls into the following domains:

- Registration;

- Meter Technical Details (MTD);
- Consumption;
- Schedules;
- Industry standing data;
- Load shapes; and
- BSC settlement data.

This list may be revised following the baselining of the E2E MHHS Design.

4.3 Registration data and Meter Technical Details

Registration data consists of MPAN plus other items such as address. There are several challenges to be overcome when “manufacturing” MPANs, such as the presence of checksums and we understand that not all services involved in MHHS are able to use “non-live” MPANs. We therefore intend to use live MPANs in MHHS testing and will put in place appropriate safeguards. This is consistent with other industry-wide programmes. An MPAN alone is considered possibly personally-identifiable information (PII) by the Information Commissioner’s Office (ICO). We will carry out a Data Privacy Impact Assessment (DPIA), which we will do at an appropriate point, when the E2E MHHS Design has been finalised. All programme participants will be required to follow guidance or actions resulting from the DPIA.

It will be the responsibility of the DNOs and IDNOs to take a copy of live registration data on a specified date, including the associated Meter Technical Details (MTD) to ensure that registration data and MTD data are aligned and make this available to the SI. The data-cut timing will be specified as being after the overnight batch run so that all data copied will be correctly aligned. We recognise that testing will be going on over a long period and that data usefulness degrades over time. Accordingly, when planning the testing (in the Overall Test Data Approach), we will identify a point where we make a decision whether or not to take an additional data cut. The requirements with respect to PSR customers will be respected in taking this data cut and those customers who have expressed a wish for their data not to be included will be excluded from the cut. Note that if a second data cut is taken, then previous data will have to be removed and new allocations of data made based on the new data cut.

Since it is helpful to suppliers to have their “own” MPANs to use in testing, we will not anonymise the MPANs. (In other words we will not change the registered supplier.) However, we will reserve a number of MPANs, taken from the data sets of the large suppliers, for use by software and service providers. We will make it clear to the affected suppliers (as well as the software and service providers) which MPANs they can and cannot use. In addition, a number of MPANs will be reserved so that they can be used for scenarios involving new registrations. The SI will be responsible for managing and securing this data set and for communicating it to PPs in a suitable manner. Initial advice from InfoSec experts indicates that this is a valid approach but it will be confirmed when we conduct the Data Protection Impact Assessment (see section 6 - Data Protection Impact Assessment).

It is likely that suppliers will wish to take a cut of their data at the same time to aid their testing. We appreciate the preparation involved in doing this, so we will give plenty of notice of the date of the data cut.

4.4 Consumption data

For consumption data, in line with good practice, we intend to conduct some testing using old live data captured onto our test environments. The method of capturing this and the associated responsibilities will be defined in the Test Data Plan for the relevant test stage. The ICO has ruled that the combination of MPAN and consumption data is PII, the use of which must be adequately justified and the data itself correctly risk-assessed and secured. We believe that the use of live data in at least a portion of the testing is important to flush out defects which may not be found using manufactured data (because incorrect assumptions may be made in the data manufacture). We recognise the importance of treating this data with the correct care and therefore will “randomise” the consumption data with respect to the MPANs. This means that we will use real consumption data but the MPAN with which it will be associated on our test systems will not be the MPAN to which it relates in the real world. Further details of the way this will be done will be included in the Test Data Plan for the relevant test stage, as well as the party responsible for data provision and data anonymisation. There are clearly important considerations for the anonymisation, such as complying with the ICO’s Code of Conduct for such processes, ensuring security of transit by use of encryption, ensuring adequate procedural

and physical controls for access to the data before it is anonymised, as well as taking account of meter technical details. These will be described in the Test Data Plan. We are assuming that the SI will carry out this anonymisation but this too (as well as the source of the data) will be confirmed in the Test Data Plan.

However, doing all of the SIT and UIT testing with live consumption data is not necessary and for the bulk of the testing we will use manufactured data (as the equivalent of smart meter data, advanced meter data or dumb meter readings for processing and sending on to Elexon Central Services). The SI team will produce test software which can produce consumption data for intake by the Data Services. We believe this will provide maximum flexibility because managing old live consumption data and lining up dates and ensuring its integrity will be a significant challenge. Instead, the manufactured data will be easily parametrizable for dates and lists of MPANs and we see this as a significant aid to the testing. We will set rules for the generation of test consumption data that will ensure that consumption data is reasonable for the particular registration and MTD combinations. We will also rely on test software to produce settlement period data suitable for intake by the Load Shaping Service and the Market Data Service (produced either by the SI team or by Project Helix – to be confirmed). The consumption data generated will need to be parameterised and be realistic in order to pass the validation checks carried out by various services participating in MHHS. In order to produce sufficiently realistic data, we are likely to need some real data on which to either base the patterns of generated data or for it to be directly re-used. This will be detailed in the relevant Test Data Plan.

4.5 Schedules

The suppliers and supplier agents participating in testing will be expected to generate schedules appropriate to the tests being executed and aligned with the planned schedule being co-ordinated by the SI.

4.6 Industry standing data

Industry standing data is the equivalent of the existing Market Domain Data. Live industry standing data will be obtained to align with the data cut of MPANs.

4.7 Load shapes

Load shapes will be generated by the Load Shaping Service (LSS) itself when that forms part of the testing. When the LSS is not being used, then the SI team will request Project Helix to provide sufficient examples of load shapes for the required testing.

4.8 BSC settlement data

GSP Group metered data is required from CVA systems in order to apply a GSP Group Correction Factor to ensure that consumption data matches in the Elexon Settlement Systems. This will need to be generated to match a realistic level of consumption data for tests where VAS systems are included. We expect that any other BSC Central Systems test data required to retain the integrity of testing will be identified and generated by the Elexon Helix project in collaboration with the MHHS Programme.

4.9 Summary of responsibility according to data domain

Where the SI stores test data, this will be held on the Azure environment and secured in the standard way for Microsoft Azure.

Data domain		Live/ Manufactured Data required?	Source	Managing data	Notes
Registration	Cut of live	<input checked="" type="checkbox"/>	(I)DNOs	MHHS IM	MHHS IM anonymises supplier
	Manufactured	<input checked="" type="checkbox"/>	n/a	n/a	

Data domain		Live/ Manufactured Data required?	Source	Managing data	Notes
MTD	Cut of live	<input checked="" type="checkbox"/>	(I)DNOs	MHHS IM	
	Manufactured	<input checked="" type="checkbox"/>	n/a	n/a	
Consumption	Cut of live	<input checked="" type="checkbox"/>	Suppliers /agents/ (I)DNOs	MHHS IM	The method of anonymisation and responsibility for carrying it out will be determined when the full data model is available following MHHS E2E Design being baselined.
	Manufactured	<input checked="" type="checkbox"/>	Supplier/ agent using MHHS IM test harness	Supplier/ agent	The participating supplier or agent will be responsible for managing this data because they will generate it and use it
Schedules	Cut of live	<input checked="" type="checkbox"/>	n/a	n/a	
	Manufactured	<input checked="" type="checkbox"/>	Suppliers /agents	Suppliers /agents	
Industry standing data	Cut of live	<input checked="" type="checkbox"/>	Elexon Central Systems	MHHS IM	This data will not be all "live" because there may be new items being introduced by MHHS but it should be as live-like as possible
	Manufactured	<input checked="" type="checkbox"/>	n/a	n/a	
Load shapes	Cut of live	<input checked="" type="checkbox"/>	n/a	n/a	
	Manufactured	<input checked="" type="checkbox"/>	Elexon Central Systems	MHHS IM	This is considered manufactured because it will come from the tests systems of Project Helix and not the existing live systems.
BSC settlement data	Cut of live	<input checked="" type="checkbox"/>	n/a	n/a	
	Manufactured	<input checked="" type="checkbox"/>	Elexon	MHHS IM	This could be based on live data but as long as it is representative, the source is not important.

5 Other general principles

Other principles underpinning the test data approach include the following:

- Data provided will be adequate to support the planned number of participating PPs;
- A particular set of data will be allocated to each PP, as well as a number of MPANs reserved for initial registration by each PP;
- Customer names will not be included in the data provided;

- Each PP will be allocated data with which to test;
- The data one PP is using for testing will be exclusively for its use and will not be revealed to another PP (other than for scenarios which require sharing data, for example CoS);
- Data will be reviewed before testing starts;
- The test data will be designed with the test scenarios in mind, including adequate volumes of data for non-functional volume/performance testing;
- Test data needs to be able to be manipulated where necessary to exercise negative testing;
- Storage and use of data will comply with the UK Data Protection Act and GDPR; and
- A policy regarding data restoration will be defined in the Overall Test Data Approach document. This will either be that data will be able to be restored (if needed because of a serious failure) at some point during the testing or that all data will be expected to be able to move forward without restoration.

6 Data Protection Impact Assessment

6.1 Introduction

A Data Protection Impact Assessment (DPIA) is a tool to help organisations find the most effective ways of complying with data protection obligations and meet individuals' expectations of privacy. In the UK, the Data Protection Act 2018 and General Data Protection Regulations (GDPR) are the cornerstones of data privacy legislation. DPIAs are a key element of a 'privacy by design' approach - one that builds in privacy and data protection compliance from the outset.

The MHHS IM will commission a DPIA once the E2E MHHS Design has been baselined.

The MHHS IM will also conduct a Legitimate Interest Assessment.

It has been noted that the most appropriate way of ensuring all parties' compliance with the policies defined is via a framework agreement to which all parties are signatories (rather than by many bilateral agreements). This is the standard method for this type of multi-party requirement and therefore the method we plan to use.

6.2 Points for consideration in the DPIA

The following points have been noted and will be considered as part of the DPIA:

- **MPAN reference IDs** – Is it appropriate to use a reference table where MPANs are assigned identifiers for the purposes of all parties communicating about which MPAN they are using in testing? This gives a level of anonymisation and avoids having real MPANs being passed around in testing communications. But it does introduce a level of complexity, where one of the guiding principles of our testing approach is to ensure simplicity wherever possible.
- **Allocating MPANs to software and service providers** – The intention is to allocate certain MPANs from the large suppliers for the sole use of software and service providers. Is there any DP issue with this?.
- **Method of anonymising the consumption data** – How will the consumption data be anonymised, taking into account the associated Data Protection requirements?

7 Test Data Plans

Full details of the test data to be used will be specified in the Test Data Plan for each test stage. These will cover the process and tool used to manage the data and for each item of data needed for system set-up identify:

- Party responsible for providing/maintaining/anonymising the data;
- How the data relates to other data required for testing;
- Source and content of the data; and
- Quantity of data required.

The Test Data Plan will also cover (as relevant):

- The method for capturing consumption data;
- The method for anonymising the consumption data;

- How data is allocated to each PP, including that set aside for software and service providers;
- How data is set aside for CoS and new registrations;
- How data allocation is communicated; and
- How co-ordination will ensure no overlap/conflict between PPs when testing.

The UIT environment will use the same set of test data as the SIT environment.