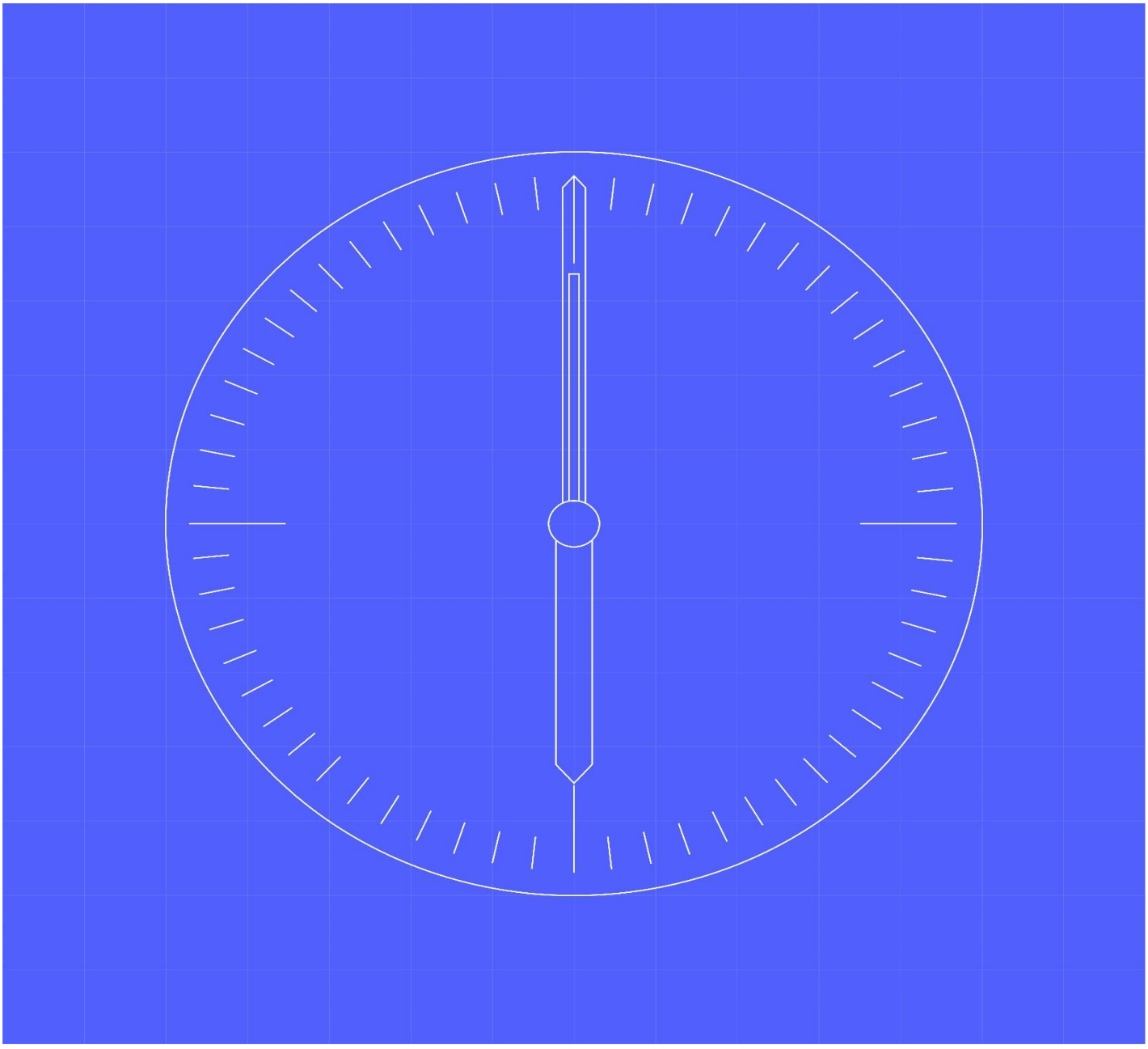
Overarching Test Data Approach & Plan



Document owner	Document number	Version
SI Test Team	**MHHS-DEL813**	**0.6**
Status:	Date	Classification
Draft	**24 May 2023**	**Public**

1. Contents

Figures

1.1 Change Record

Date	Author	Version	Change Detail
30 November 2022	Dean Bailey	0.1	Initial draft
23 December 2022	Dean Bailey	0.2	Sent for SRO review
13 April 2023	Cesar Lopes	0.3	Sent for internal LDP review
27 April 2023	Kate Goodman/John Wiggins	0.4	Draft for SRO review
05 May 2023	Cesar Lopes	0.5	Approved by SRO
24 May 2023	Cesar Lopes	0.6	Updated in response to some v0.5 industry comments.

1.2 Reviewers

Reviewer	Role
Kate Goodman	LDP Test Architect
Nigel Hunt	LDP Test Manager
John Wiggins	LDP Transition/Migration Lead
Adrian Ackroyd	SRO Client Programme Test Manager

1.3 References

Ref No.	Document/Link	Publisher	Published	Additional Information
REF-01	MHHS -DEL 315 E2E Testing & Integration Strategy	SI Testing	April 2022	
REF-02	MHHS-DEL 300 Test Data Strategy	SI Testing	May 2022	
REF-03	MHHS-DEL618 Environment Approach and Plan	SI Testing	February 2023	
REF-04	MHHS-DEL816 Population of Data Items for Testing	SI Testing	April 2023	
REF-05	MHHSP-DES138-Interface Catalogue v5.0	MHHS	February 2021	
REF-06	MHHSP-EDI021_ISD_Entities v5.0	MHHS	February 2023	
REF-07	MHHS-DEL1197 Interface Code of Connection	MHHS	May 2023	
REF-08	MHHS-DEL1210 Data Integration Platform Public Key Infrastructure (PKI) Certificate Policy	MHHS	May 2023	
NCSCDARP	NCSC Cloud Security Principles – Data at rest protection - https://www.ncsc.gov.uk/collection/cloud-security/implementing-the-cloud-security-principles/asset-protection-and-resilience#rest	NCSC		

1.4 Terminology

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

2 Executive Summary

The approach to establishing a common and fully aligned set of test data for use in the industry-wide test phases of MHHS requires activities to be undertaken by both market participants and the MHHS SI team. These activities need to be undertaken by almost all market participants (regardless of whether a participant intends to qualify via SIT or via Qualification Testing) and many activities need to be completed well in advance of the planned start of SIT.

The approach identified is:

- A data cut will be taken by each relevant Data Controller or Data Processor (participant) on the same, pre-determined date and the same point in the processing day.
- The participant must store this data cut in a form suitable for it to be loaded into its test database ahead of the start of SIT or Qualification Testing.
- The participant must provide certain data items relating to each MPAN in its data cut to the MHHS SI team via a secure file transfer system (sftp) set up by the SI team. These data items are the ones which need to be aligned across all participants' systems in order to have a coherent data set for testing.
- The SI team will conduct cleansing, augmentation and transformation as necessary and then allocate each MPAN to a supplier (for most MPANs, this will be the actual supplier at the time of the cut).
- Once manipulated, the SI will send the relevant data to each participant who must then update its test system with the augmented set in before the start of testing.
- It is the responsibility of the participant to obtain or develop appropriate means of taking and storing the data original data cut, extracting the data required by the SI, loading the data cut into its test environment and updating the data with the augmented/manipulated data provided by the SI.

Each data item related to settlement has been identified (from a combination of the MHHS Data Catalogue and the D-flows which will continue to be used in the new arrangements, such as D0149 Notification of Mapping Details giving Meter Technical Details). These are listed in an accompanying spreadsheet which identifies the method and responsibility for provision of each data item related to an MPAN. The method is one or more of: provision in a data cut, cleansing, transformation and deriving a value in a rule-based or other way.

3 Introduction

3.1 Document Purpose

This document aims to define a Test Data Approach that will deliver an aligned set of data for SIT and UIT. This document also provides a definition to MHHS Industry Test participants on the approach for test data, covering data management, allocation, cleansing, storage and archiving and maintenance.

It sets out the scope and approach for the use of data to support the successful completion of testing. This document forms the agreement between all parties associated as to how Test Data will be managed.

The following groups should read this document:

- Lead Delivery Partner (LDP)
- SRO Function (SRO)
- Data Working Group (DWG)
- Programme Participant Test Managers
- Testing and Migration Advisory Group (TMAG)

3.2 Reviews and approvals

The Overarching Test Data Approach & Plan document will go through an initial LDP review by the following team members:

- Kate Goodman, LDP Test Architect
- Nigel Hunt, LDP Test Management
- John Wiggins, LDP Transition/Migration Lead

Upon completion of the LDP review, it will then go through a formal review by the following SRO members:

- Adrian Ackroyd, SRO Client Programme Test Manager

Once comments and feedback have been incorporated, review and recommendation for approval will be requested from:

- DWG

Formal approval will then be requested from:

- Testing and Migration Advisory Group (TMAG)

3.3 Change forecast

The SI team will own this document and keep it up to date, with review and approval by MHHS programme governance as appropriate. Each new version supersedes the previous version in its entirety. Updates to this document will follow the review and approval process outlined in section 3.2.

It is acknowledged that there are assumptions and caveats associated with this version of the document, as detailed in section 3.5 Therefore, this document will be iterated once these have been addressed.

3.4 Summary of changes

Document updated following industry consultation:

- Assumptions: Updated to make clear the SI scope on data allocation for SIT, UIT and Qualification;
- Scope: Updated to make clear that data cut is required by Non-SIT participants;
- Systems Requiring Data Cut:
    - MOp: updated to make clear data cut is required for all MPANs to be able to get Meter Technical Details (MTD)
    - EES: updated to make clear the mechanism for data refreshing
    - Added paragraph to state Data cut is required by all MPIDs in an organisation
- Data Cut Process: Added paragraphs to state the participant can load data to test environment after SI runs allocation
- Sensitive Data Protection: Updated to state the elements of Data Protection and Data Sharing Agreements required to be addressed by DPIA.
- Population of Data Items for Testing: corrected all specific data items issues raised.

3.5 Assumptions

The following assumptions have been made:

- The SI will coordinate a production data cut to be applied to the SIT environment.
    - The same data is expected to be applied to the UIT environment (used for Qualification and E2E Sandbox).
- The LDP will be responsible for allocating data for E2E Sandbox testing but for Qualification the SI will assign a large volume of data to the Code Delivery Bodies for them to then manage the allocation to Qualification Test Participants.
    - The SI will segregate the MPAN data allocated for use within the Qualification Test and E2E Sandbox stages.
    - The SI will be responsible for managing the allocation of MPANs to E2E Sandbox participants only.
    - The Code Delivery Bodies will be responsible for managing the allocation of MPANs to Participants undertaking the Qualification tests.
- Real MPAN data and their respective postal codes are required to allow the execution of integrated tests without causing exceptions in validations embedded in the target systems.
    - Real electricity consumption data may be required by exception only. However, if required, the consumption will never be never linked to the original MPAN that generated it. Further details of how this will be done will be included in the Test Stage Test Data Approach and Plans for the relevant test stage, as well as the party responsible for data provision and anonymisation.
- The SI Technical team will provide allocation and maintenance of the Public Key Infrastructure (PKI) solution.
- Industry Standing Data will not all be “live” as there may be new items being introduced by MHHS, but it will be as live-like as possible.

4 Objectives

4.1 Objectives

The objectives of this Test Data Approach and Plan are to:

- Describe the approach for establishing a “Day 0” set of test data to be used for MHHS industry testing, where the various market participants test together with the Central Systems in SIT and UIT. This includes the mechanism by which the data is to be transferred between the MHHS programme and participants in the run-up to test commencement.
- Allocate responsibility to the relevant market participant for each activity, covering taking a data extract, data cleansing, data enhancement, data transformation and data load.
- Describe the measures to be taken to ensure the appropriate security of the data.
- Provide a basis on which a full Data Protection Impact Assessment (and if necessary a Legitimate Interest Assessment) can be done.

The fundamental principles have already been set out in the MHHS-DEL 300 Test Data Strategy (see REF-02). This document builds upon those principles.

5 Scope

5.1 In Scope

- The method by which test data is prepared in order to create an initial starting point for SIT and UIT.
- The production data cut required to prepare test environments for SIT and UIT.
- The production data cut required by Non-SIT participants to be used later on Non-SIT phase tests and qualification.

5.2 Out of Scope

The following activities are not in scope for this document:

- **PIT-related activities** - For PIT, which takes place on the Programme Participant’s own standalone test environment, the responsibility for planning, specifying, generating, managing, and securing the test data lies with the participant.
- **Data cleansing required for cutover to Go live** – Production data cleansing prior to migration and Go Live is covered by Migration Workstream.
    - ***Note***: It is **in the scope** of this document activities related to **Test** Data. A test data cleansing is required prior to starting the execution of SIT and UIT tests to reduce potential errors due to data inconsistencies.
- **Qualification Testing data allocation** - The mechanism and approach for allocating data to Qualification Testing participants is covered elsewhere.

6 Data Management

6.1 Systems Requiring Data Cuts

Data cuts will be required from many (but not all) systems involved in the settlement. The data cuts will be required for providing data for participants’ own systems and also for provision to the SI so that it can be suitably manipulated and allocated to participants, as described in previous sections. The SI Test Data Lead will require subsets of data files from most services to coordinate, allocate, cleanse, and provide complementary data using an SFTP environment. The participant will be responsible for loading test data items for each service on their respective test environments.

The participant is responsible for ensuring they have a mechanism in place for loading their allocated data into their own test systems.

The following two diagrams provide a high-level view of the data cut process and the responsibilities of each participant.

In Figure 1 Overview of data-cut activities (below), reference is made to production systems and legacy systems from which market participants need to take data cuts. The legacy systems are all those which will not exist in the new MHHS arrangements but which contain data relevant to new systems which will form part of the new MHHS arrangements. The production systems are those which are already in the current arrangements and will continue to exist (albeit in amended form) in the new MHHS arrangements.

It is advised that any participant standing up new systems which require legacy data for running tests should take Data Cut from their legacy systems in alignment with other participants. The data cut can be used to populate their new system on Test Environments. Note that not all legacy systems will require data cuts, as described in the list below.

The legacy systems are:

- **ECS** – data cuts are required in order to populate the new ECS systems with whatever data these systems need to operate in the new arrangements; this includes obtaining all the data required to populate MHHS Industry Standing Data (ISD) such that it aligns with the MPANs taken in the MPRS data cuts and it is understood that to be complete, the ISD may need to be obtained from several data sources

- **DC/DA and MA** – data cuts are not required because appointments to the new Data Services will be done by BAU transactions during the testing; the Data Services will start the testing with no data in their databases.

- **MOp** – data cuts are required in order to collect Meter Technical Details (MTD) to populate the Day 0 databases.
    - **Data cut is required for all MPANs.**

The production systems are:

- **MPRS** – a data cut is required of each MPRS to provide registration and some MTD data

- **Network Operator** – a data cut is required such that the MPANs line up with those in the relevant MPRS data cut

- **UMSO** - a data cut is required such that the MPANs line up with those in the relevant MPRS data cut (where UMSO services exist for a particular LDSO)

- **Supplier** – a data cut is required to allow the supplier to set up test environments.
    - The supplier may wish to manufacture consumption data and other consumer data not being used on DIP messages. Example: consumer name.

- **EES** – no data cut required, but EES is expected to conduct a full refresh before testing commences, based on the MPRS test data in those MPRS instances being used for the testing
    - As part of the initial population of test environments, the EES will be populated by the MPRS systems, which will be undertaking the SIT using the existing full MPAN refresh mechanisms.

- **CSS** – a data cut is required such that CSS can operate as part of the testing; while much of the CSS data can be obtained from the MPRS data cuts, there are some items (e.g. Domestic Premise Indicator) which are mastered in CSS and are therefore needed for the test data

- **Smart Metering** – a data cut is likely to be required for Smart Metering; details for Smart Metering tests and data will be provided in the SIT Functional Test Data Approach document (for the document release data, refer to the MHHS Programme Plan)

If an organisation has more than one MPID, a data cut is required for all their MPIDs. The usage or not of specific MPID data cut in test environments will depend on the outcomes of test reliance and required test cases.

6.2 Data Cut Process

The process by which the data cuts are taken and the responsibilities for each activity are shown in the following two diagrams.

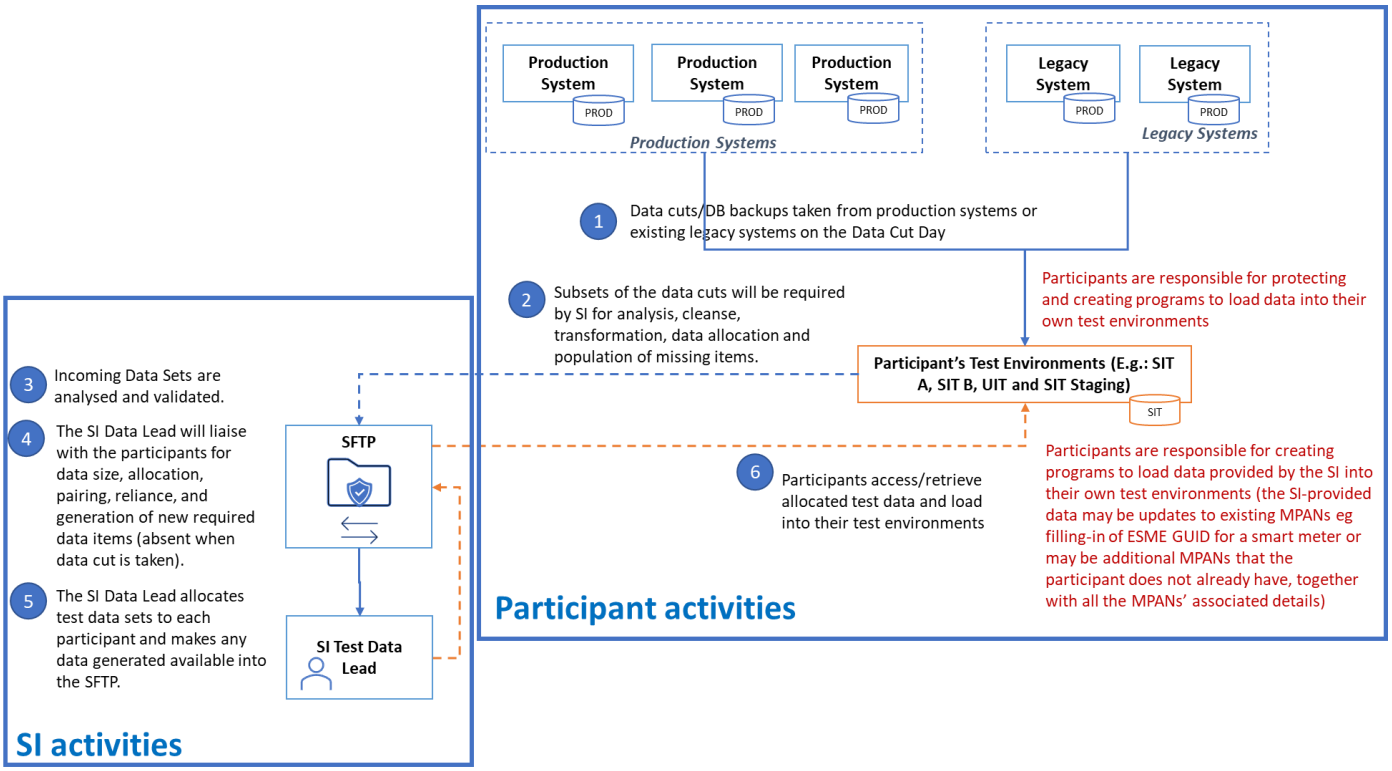


Figure 1 Overview of data-cut activities

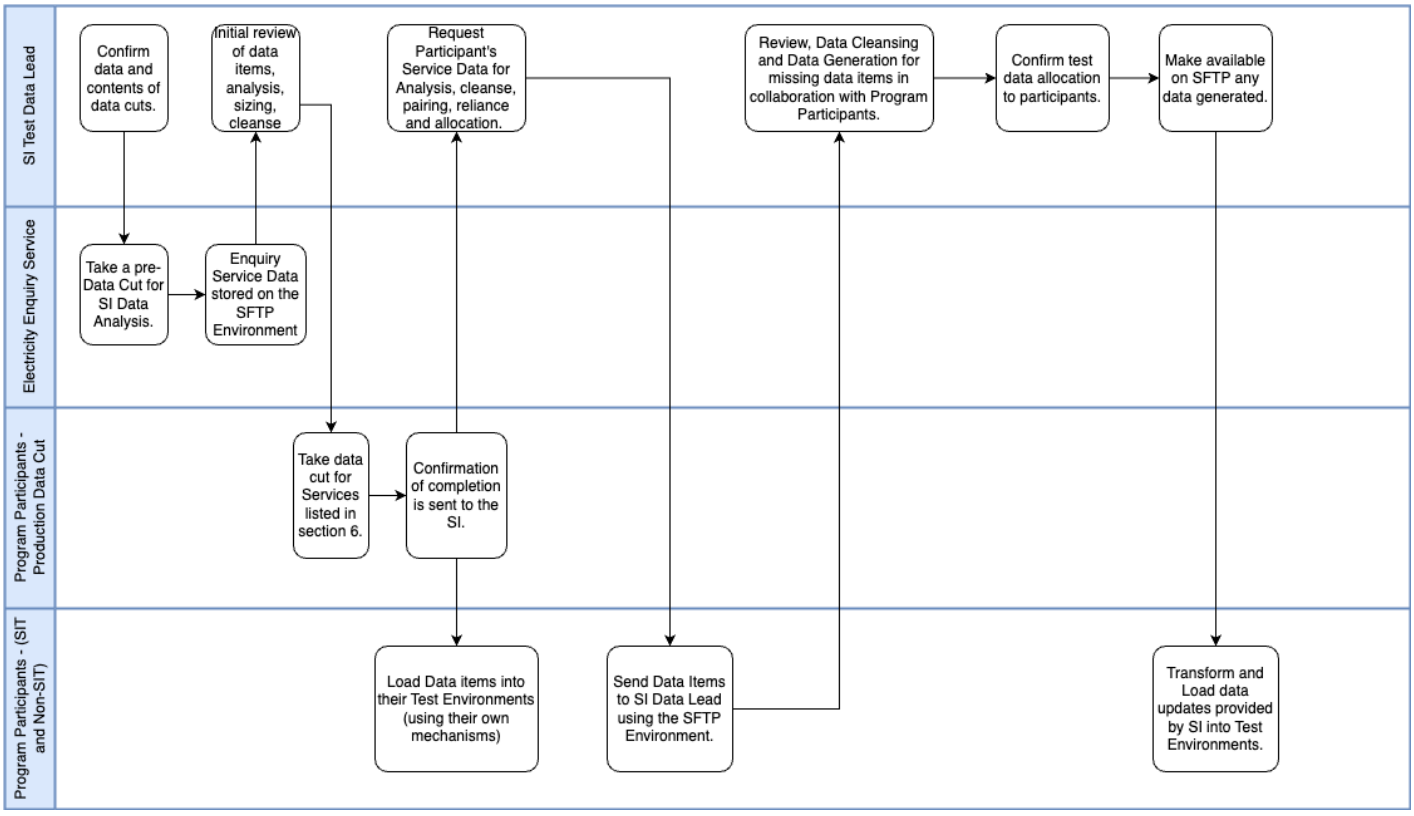


Figure 2 Overview of general activities and responsibilities

Regular data management sessions will be held between the SI Test Data Lead and participants to:

- Ensure the correct data requirements are confirmed prior to testing
- Co-ordinate the population of services
- Monitor the progress of data-related activities.

Note:

1. After the Data Cut, the SI Test Data Lead may request samples of data from the participants for analysis, checking alignment and integrity. The SI will provide feedback directly to the participant on any issues found.

2. Prior to the start of tests, the SI will allocate MPANs and related data to be used by each Programme Participant. The PP may decide to load the test data to their test environment after the SI executes data allocation for testing in order to minimise the size of the test environments.

3. The SI will avoid re-using MPANs across test phases. E.g., If an MPAN was assigned to run tests on SIT-CIT, the SI will avoid re-using the same MPAN for SIT-Functional. However, the re-use may be required if not enough data is available.

4. Complementary specific Test Data Approach & Plan per test phase will be published by SI for SIT Component Integration Test, SIT Functional Test, etc. and will cover details of data allocation.

6.3 Summary of Responsibility according to Data Domain

This can be found in section 4.9 of the MHHS-DEL 300 Test Data Strategy (see REF-02).

6.4 Test Data Creation

The SI will not need to collect all the data from each participant's data cut. The data items needed by the SI are defined in the spreadsheet which accompanies this document MHHS-DEL816 Population of Data Items for Testing (see REF-04). The data items have been identified from a combination of the MHHS Data Catalogue and the D-flows which will continue to be used in the new arrangements, such as D0149 Notification of Mapping Details giving Meter Technical Details. The service/system which must provide each data item is indicated in the column "Responsible for Population" – Column I.

An identification of settlement-related data items in DB messages passing within an LDSO's organisation is yet to be added to this list and will be included in a later version of this document.

When the SI assembles the data from all the data cuts, there will be certain items for each MPAN which may need be manipulated before the SI sends back the updated data to the participants. These data items are also described in the spreadsheet and an indication is given of how they need to be manipulated:

- Data items which **already exist in current systems** may need to be:
    - **Cleansed** – It is possible that some data items relating to a given MPAN are inconsistent (for example, MTD-related items may not always be aligned). The SI will carry out simple cleansing to obtain a sufficiently large dataset of consistent data where this is needed.
    - **Transformed** – Some existing data items (such as dates) exist in a different form from that required by MHHS, and the SI will carry out a suitable transformation.
- Data items which **do not exist in current systems** (either because they do not exist at all or they exist but are as yet unpopulated):
    - **Source or Method of Population** – How these data items will be populated is described (at a high level). Items which do not need to be populated in the Day 0 database are either "transactional" (items which are generated/used in the context of sending an IF message, such as Sender Correlation ID) or "BAU" (items of business data which are generated by one of the Services and then communicated to other Services using an IF message in Business As Usual). Those items which do need to be populated in the Day 0 database are marked "data cut" if they need to be provided by a data cut (and accompanied by an entry in Responsible for Population) or "manufactured" if the SI is going to derive them.
    - **Responsible for Population** – The organisation responsible for populating these missing items (in many cases this is the SI).

Note:

1. New data items introduced recently (e.g.: new data items in CP1558, R032) proposed to go live in June 2023 may not be fully populated during the data cut. The SI, in collaboration with participants, will manufacture values for the new data items required for testing.
2. The MHHS-DEL816 Population of Data Items for Testing (REF-04) captured the data items required by the DIP interfaces, i.e. data in motion. The Programme does not have visibility of all participant's backend systems, databases and data in rest requiring data cut to be later loaded to test environments.
3. It is the participant's responsibility to ensure the data cut contains the minimum set of data required by the DIP interfaces described in the MHHS-DEL816 (REF-04) plus the minimum set of data not visible by the DIP, but required to make their test environments functional.

6.5 Test Data Use

Each participant must set up their database for the start of testing and load it with the data from its data cut and updated by changes supplied by the SI following its manipulation and allocation of MPANs. The generated data will be available for download on the SFTP environment.

The data can only be used for the MHHS Programme and for the means of testing in line with Sensitive Data Protection processes and be compliant with UK GDPR (see section 9).

6.6 Data Quality

A Data Assessment Report will be created to measure the quality of data as it is currently held in the existing arrangements and the activities that need to take place prior to testing, to cleanse the data/add further data to achieve an adequate quality. This document is not intended as a full plan of the data cleansing/data addition activities but will be the document upon which this is based.

6.7 Test Data Cleansing

A Data Cleansing Catalogue will be used to improve the quality of the data being used for testing purposes. This catalogue will include the data items that will require cleansing and will detail the 'Cleansing Owner' and 'Date Required By'. The catalogue will also include metrics that will be monitored throughout the cleansing process and will be used to report progress of cleansing activities to both the MHHS programme and the Data Working Group.

It is recognised that cleansing activities should be started as early as possible, following confirmation that data cuts have been completed, although Supplier mastered data items, such as the Domestic Premise Indicator and Consent Granularity have been excluded from early data work.

It is expected that some data items may have updates close to the Data Cut not captured by all systems. E.g. MPRS only operates on Working Days, while Supplier's Registration Systems may operate on non-working days. A data cut executed on a Saturday may cause some data items to be out of line. To mitigate the issues, during test data allocation, the SI will exclude data items with updates two days before the data cut date and time window.

This Approach and Plan document also acknowledges R0066, R0032, BSC CP1558 and the planned deployment in June 2023 for BSC CP1568. Further detail regarding data cleansing will be provided in both the Data Cleansing catalogue and the Stage-specific Test Approach and Plan documents.

6.8 Test Data Manipulation

For the new data items, the SI Test Data Lead, in cooperation with Program participants, will generate the test data required for SIT.

- A complete reference of data required for testing is listed on [REF-05] MHHSP-DES138-Interface Catalogue v5.0, tab 01-DataCalogue.
- Data items with a current DTN ID (example: J003 for DI-063 MPAN Core) should be present on the data cut.

- Data items without DTN IDs may need to be generated for SIT or will be populated by messages created during the execution of test cases.
- The SI Test Data Lead will allocate data to ensure all participants have the specific data required by test scenarios. It may require the SI Test Data Lead to transfer MPANs data between participants, ensuring that any data transferred is anonymised. Example:
  - A participant does not have MPANs with Meter Technical Details required for a specific test case
  - The SI Test Data Lead will transfer MPANs from another participant and all required related data for the test, ensuring all data is anonymised.
- An eventual reassigning of an MPAN following established industry processes (DTN / CSS dataflows) is preferred to ensure appropriate validation is applied to any changes.

6.9 Data Refresh and Restore

The same data cut shall be used to load data into the different independent Tests Environments required at different stages of the programme. Example: the data cut should be used to load the “SIT A” Test Environment used by SIT Component Integration Testing and SIT Functional. The same data cut will then be used to load the “SIT B” Test Environment used by SIT Migration. Tests in one test environment shall not impact the others.

It is expected that the same data cut will be loaded to the UIT environment, used for Qualification and E2E Sandbox.

Although the expectation is that a single data cut will be taken for test activities, it is intended that a decision point will be provided where the MHHS Programme may decide to take an additional data cut if the original data source has become less usable over time.

Further details on data refresh can be found in section 6.6 of the [REF-03] MHHS-DEL618 Environment Approach and Plan.

6.10 Storage

Where the SI stores test data, this will be held on the Azure environment and secured as outlined in section 8.3 of this document. SFTP shall be used to protect all test data during transfers to or from the participants. For any sensitive data being stored, it will be secured with a level of encryption which is aligned to NCSC guidance [NCSCDARP].

All data will be treated under GDPR principles in storage, loading, transmission and retention. The SI’s approach for dealing with GDPR will be defined with the appropriate security guidance.

6.11 Removal of Data (Archiving)

When test data no longer needs to be moved between participants, external access to the SFTP server will be blocked. All resources and data within the SFTP resource group will be destroyed, after which the resource group will be removed.

Following the completion of SIT and UIT, each party, including the SI, shall securely delete all data from environments and storage media in accordance with relevant NCSC Secure Sanitisation of Storage Media guidelines.

If test environments and data are required to support enduring qualification, the programme will be required to update the Test Data Approach and Plan accordingly to retain any data and support the activity.

7 Domain-Specific Data

7.1 Introduction

This section describes at a high level the organisation responsible for providing data according to the different data domains (which were identified in the MHHS-DEL 300 Test Data Strategy (see REF-02). This is intended to provide context before referring to the details of responsibility on a data item basis contained in the accompanying spreadsheet MHHS-DEL816 Population of Data Items for Testing (see REF-04).

7.2 Registration Data & MTD

Registration data will be supplied by the MPRS data cuts, along with the MTD that is held in those systems. Further MTD data will be provided by MOps.

7.3 Consumption Data

The need for historical consumption data is yet to be determined. It will be clarified in the SIT Functional Test Data Approach document (for the document release data, refer to the MHHS Programme Plan).

7.4 Schedules

No schedules need to be set up before the commencement of testing.

7.5 Industry Standing Data

Industry Standing Data will be supplied by ECS.

8 Data Allocation

8.1 Directory Structures

The below figure provides a logical view of the structure that will be used to store test data.

- An individual and secured directory will be created for each Programme Participant within the SFTP environment.
- Each parent participant directory will include two subfolders:
    - From_Participant: for sharing data from the participant to the SI Test Data Lead
    - To_Participant: for sharing data from the SI Test Data Lead to the Participant
- Inside the From/To subfolders, multiple folders for data sharing events will be created following the name convention:
    - Date and Time of the data sharing event + “_” + short description
- Inside the folder created for the sharing event, multiple files can be added.

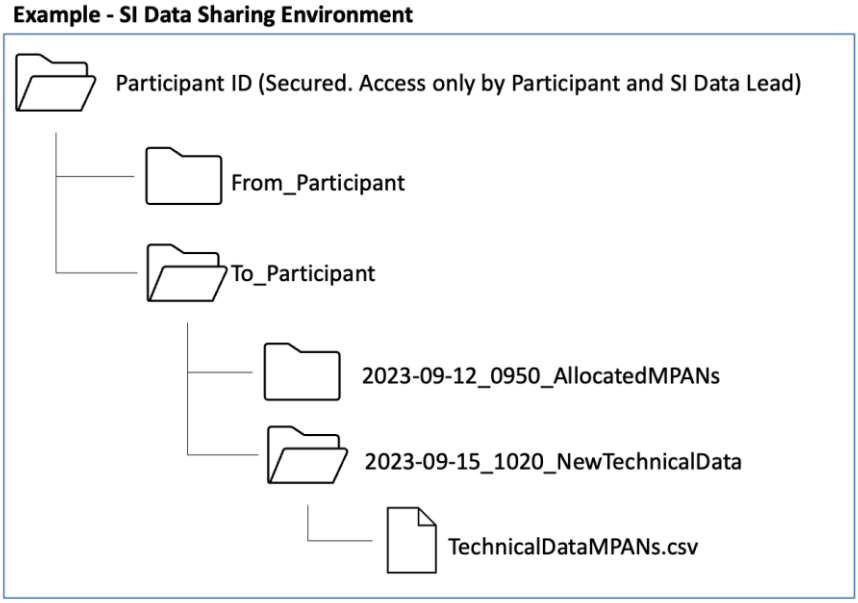


Figure 3 Logical view of allocation structure

8.2 Mechanisms for Use and Management of Live Data

Secure File Transfer Protocol (SFTP) will be used to store and transfer all test data, used by all Programme Participants. SFTP details will be provided to each participant separately due to their sensitive nature and therefore will not be detailed in this document. Each participant 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 participant uses only the data allocated by the SI and the SI may require the Programme Participants to sign an undertaking to that effect.

One or more named individuals from each participant required to access the test data transfer mechanism will be expected to request access to the SFTP service via the SI, who will validate and process the request. Participants may elect individuals from Data Processing companies to access the test data transfer mechanism on their behalf. Further information on the security protocols being implemented and the privacy of data can be found in section 9 of this document.

While the SI will manage and protect the access to data being transferred “from Participants to SI” and “SI to Participants”, the access to the Participant’s own Test Environment databases and data shall be managed and protected by each organisation. During the execution of integration tests, as part of the specific test case steps, messages and test data will be transferred between participants using their systems under test and respective PKI certificates provided by MHHS.

8.3 Format for Data Transfer

The data will be transferred to the SI and back from the SI to the participants as csv files, in a format which will be specified in a later version of this document.

8.4 Allocation Method

More information on the method of allocating test data for SIT and UIT to each participant will be detailed in a later version.

8.5 Pairing

Pairing of participants may be required for certain test scenarios, such as CoS. As discussed in the November 2022 Data Working Group, during SIT, Programme Participants may not be paired with organisations with whom they already have a commercial arrangement. Applying similar principles to those used during Faster Switching, the intention is that participants will be paired based on profile, for example, I&C only and Domestic only. Further

clarification on pairing will be provided in both the Test Stage Test Data Approach and Plans and future Data Working Groups.

8.6 Data Volumes

We understand and acknowledge the importance of determining the volume of data required for testing activities. This will be discussed further in the Data Working Group and additional information will be provided in the Test Stage Data Approach and Plan documents.

9 Sensitive Data Protection

9.1 DPIA (Data Protection Impact Assessment)

A comprehensive Data Protection Impact Assessment will be conducted and finalised before the Data cut is done and before any data is shared between participants.

No data will be collected, processed and or disseminated between organisations before the DPIA report and the implementation of required Data Protection activities and legal agreements for sharing data.

The DPIA will address:

- Account for the GDPR requirements and Legal framework agreements required for sharing data:
  - Sharing Data between Participants and the Programme
  - Sharing Data between Participants and DIP during the execution of tests
- Clearer reasoning and usage of the test data for the participants from a legal standpoint due to the use of Live like data.
- Sharing of data between parties that are not naturally resident companies within the UK, having their offices offshore.
- List all data items involved in the test data deemed as Personal Identifiable Information (PII)
- List and describe the reasoning for not anonymising any Personal Identifiable Information captured.

The following points have been noted, as referenced in MHHS-DEL 300 Test Data Strategy (see REF-02) document and will also 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, considering the associated Data Protection requirements?

9.2 Data Anonymisation

As detailed in Section 4.3 of the MHHS-DEL 300 Test Data Strategy (see REF-02), MPANs will not be anonymised to prevent issues during testing. Instead, a number of MPANs will be reserved and allocated to participants to run the tests. A participant may be assigned an MPAN that is not currently in their database to be able to run specific test scenarios.

The SI will “randomise” the consumption data (if used) with respect to MPANs meaning that the MPAN with which consumption data is associated will not be the MPAN to which it relates in the real world. The allocation of consumption data, however, will have to consider the meter COP or Connection Type to avoid validation rules being triggered unexpectedly. For example, if consumption from a COP 2 meter is assigned to a COP 10.

A Data Protection Impact Assessment will be produced to ensure the approach to test data anonymisation meets all security and GDPR requirements. Further information on the anonymisation process will be given in both the Data Working Group and later versions of this document.

9.3 Security

To ensure the security of data, public keys will be used to create a more secure method to authenticate identity and rights to access a server. There are two sets of public/private key pairs (or four keys), which are:

- Client Public Key
- Client Private Key
- Host (Server) Public Key
- Host (Server) Private Key

The clients public & private keys are a pair of keys used to authenticate a client when it connects to an SFTP server. The client's private key is kept secret and stored locally on the client's machine while the client's public key is uploaded and registered on the SFTP server the client connects to.

The servers public key (commonly known as the host key) is sent to connecting clients for validation and authentication, providing assurance that the SFTP server they are connecting to is the correct server. The servers private key is only used internally by the SFTP server/server admin and is not used by clients.

The data cut covered by this document, and the participants' Test Environments are expected to contain a subset of the Production data. The production systems and related data are protected following the [REF-07] **MHHS-DEL1197 Interface Code of Connection** and [REF-08] **MHHS-DEL1210 Data Integration Platform Public Key Infrastructure (PKI) Certificate Policy**. The test data and respective environments will follow the same security procedures unless the DPIA outcome states the possibility of applying a simpler approach.

10 Data Risk Management

Though the purpose of this document is to mitigate risks as far as possible, it is important to acknowledge them and have processes and procedures in place that are initiated if a breach of data security or unauthorised disclosure is detected or suspected.

If a breach of data security or unauthorised disclosure has taken place, or is suspected, each party shall report the breach to the SI IM. The report shall contain:

- **Analysis** - Provide as much context about the breach or disclosure as possible. This includes the initial damage, how it affected your organisation, and what caused it
- **Assessment** - Determine what data has been breached or disclosed, and the number of records affected
- **Description** - Outline the consequences of the breach or disclosure for affected parties. This will depend on the information that was compromised
- **Preventive actions** - Outline what steps have you taken, or plan to take, to mitigate any further exposure
- **Oversight** - Provide the contact details of a responsible person who can co-ordinate any follow-up activities

The SI shall provide the MHHS Programme Management team a detailed report, having made all enquiries and further analysis as is reasonably possible. The SI will follow a Security Incident Management process aligned with ISO 27035:2011 and will be available to respond to reports during normal working hours (Monday to Friday 08:00 - 18:00). Best efforts will be made to respond to reports made at other times.

The SI will set up a log to track incidents, escalations and mitigation/actions taken. All logs and incident reports will be held in a secure location in Microsoft Azure.

If, during testing activities, the MHHS Programme becomes aware of any breach or unauthorised disclosure on the Azure platform, all external access to the SFTP server shall be blocked whilst the breach or disclosure is reported to the SI, as above and an investigation takes place. The results of the investigation shall be reported back to the SI and MHHS Programme Management, along with any remedial action to be taken. External access to the SFTP server will not be re-enabled without explicit authorisation from the MHHS Programme Management.

11 Test Stage Data Approach and Plan Documents

Test Stage Test Data Approach and Plans will be submitted for each test phase. Full details on what will be included in these can be found in section 7 of the MHHS-DEL 300 Test Data Strategy (see REF-02).

12 Timescales

The participants will be required to take the data cut on a **Saturday between 7pm and midnight**.

Table 1: Data cut - Time window

Day of the week	Time Windows
Saturday	From 7pm to midnight

For the date of the data cut, and all associated activities, refer to the MHHS Programme Plan.

Appendix – Items yet to be Clarified

The following is a list of outstanding issues which will be resolved before the next version of this document:

1. **Alignment with Smart Metering** – See section 6.1 - Systems Requiring Data Cuts. The way in which the Smart Metering data is aligned with the MHHS data needs to be determined in consultation with the DCC.
2. **Rules for populating data items** – See section 6.4 - Test Data Creation. The exact rules for populating missing data items need to be determined in consultation with industry (notably St Clements Services).
3. **Historical consumption data** – See section 7.3 - Consumption Data. The need (or otherwise) for historical consumption data is yet to fully explored and will be worked out in collaboration with the Data Working Group. When this is established, the method of randomisation will be determined in consultation with the programme (and participant if necessary) InfoSec consultants
4. **Allocation of MPANs** – See section 8.4 - Allocation Method. The method of allocating MPANs to suppliers will be determined in collaboration with the Data Working Group.
5. **Format of csv files** – See section 8.3 - Format for Data Transfer. The format of the files to be transferred will be published in a later version when the responsibility and manipulation of each data item has been agreed.
6. **Data volumes** – See section 8.6 - Data Volumes. Data volumes will be added in a later version of this document.
7. **Settlement-related data in DB messages** – See section 6.4 - Test Data Creation. Settlement-related data items which appear in DB messages passed within an LDSO will be obtained and analysed for those which need to be passed to the SI. This work will be done in conjunction with St Clements Services/C&C.