MHHS PROGRAMME

Industry-led, Elexon facilitated

MHHS Programme
SIT Industry Retrospective Summary

MHHS-DEL4213

Version 1.0

SIT Lessons Learned - Industry Participation

The programme recognises the value of capturing lessons learned from the delivery of MHHS Systems Integration Testing (SIT) and has a responsibility to document these insights to support the planning and delivery of future industry programmes of a similar nature.

The programme will be conducting its own internal retrospective but also proposed an industry testing retrospective, the outcomes from which would be documented and form a section of the content within a MHHS planned 'Blueprint for subsequent Industry Programmes' document.

It's understood that SIT participant teams are likely to be focused on completing critical internal activities to achieve M10, it is however also acknowledged that key Industry test team members may also be rolling off the programme shortly, and so with this in mind, the programme facilitated four 1-hour retrospective sessions over the 26 & 27 August, where industry participants attended to reflect on the MHHS testing and defect management journey, thus providing the opportunity for views to be logged on what worked well, and to identify lessons that could help streamline future industry programme delivery.

The sessions were intended for Participant test team members only, and to be punchy and constructive, and were split out as follows:

Session	When	Duration	Format	Facilitated By	Session Focus
1	26 Aug (11am)	1 Hr	Teams Call	SI Test / PPC	Component Integration Testing (CIT) / SIT Functional / SIT Migration
2	26 Aug (4pm)	1 Hr	Teams Call	SI Test / PPC	Settlements Testing
3	27 Aug (9am)	1 Hr	Teams Call	SI Test / PPC	SIT Non-Functional Testing (NFT) & Pre-Integration Testing (PIT)
4	27 Aug (4pm)	1 Hr	Teams Call	SI Test / PPC	SIT Operational (Ops) / Environments, Defects, Release Management

Recordings / transcripts made available to those who attended.

SIT Industry Retrospective Sessions

4 x 1hr sessions are an opportunity for SIT Participants to provide their views on the MHHS SIT lifecycle and consider what lessons can be learnt for future programmes

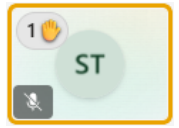
PPs were asked to consider:

- Positive and negative experiences:
    a) What went well?
    b) What didn't go so well?
- How items raised impacted:
    a) Participant(s) individually?
    b) The programme overall?
- Constructive suggestions on how things could be done differently in future programmes?

Approach to discussion:

- Being constructive and considering what has been learned that could be passed on to future teams
- The focus areas for each session were a loose structure, so PPs could raise items from other focus areas if they wished

- Attendees were asked to raise their hand in Teams when they wanted to contribute to the discussion:

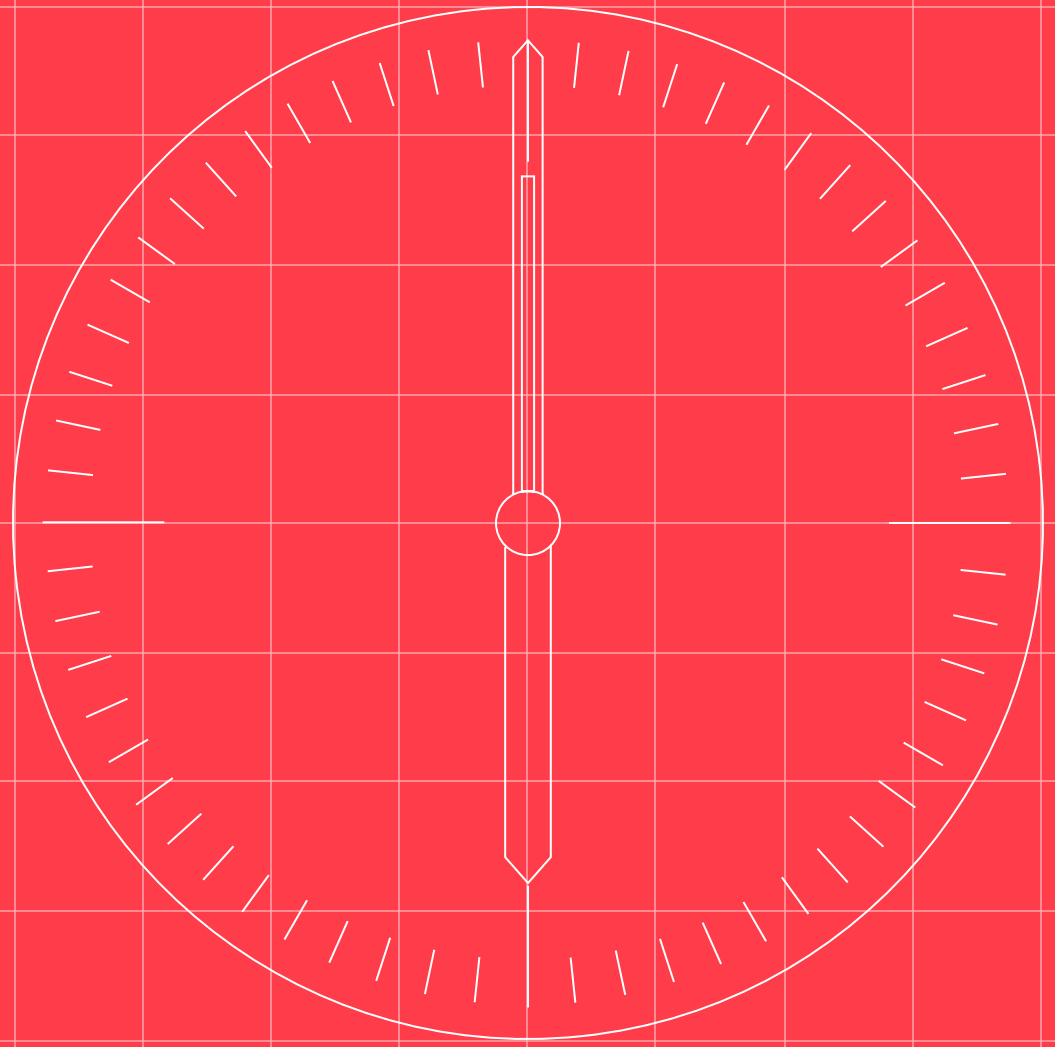


- PPs were asked to consider the whole SIT lifecycle, the SI facilitator prompted discussion on new topics to keep the session flowing e.g.
    - Test Cases & Data
    - Test Process
    - Engagement & Communications
    - Test Tooling & Reporting
    - General

Post session commitments:

- Recordings and transcripts to be published to attendees
- A summary of themes will be played back at an appropriate industry forum e.g. eSITWG.
- Lessons logged will feed into the 'Blueprint for subsequent Industry Programmes' document

Summary from Retrospective Sessions



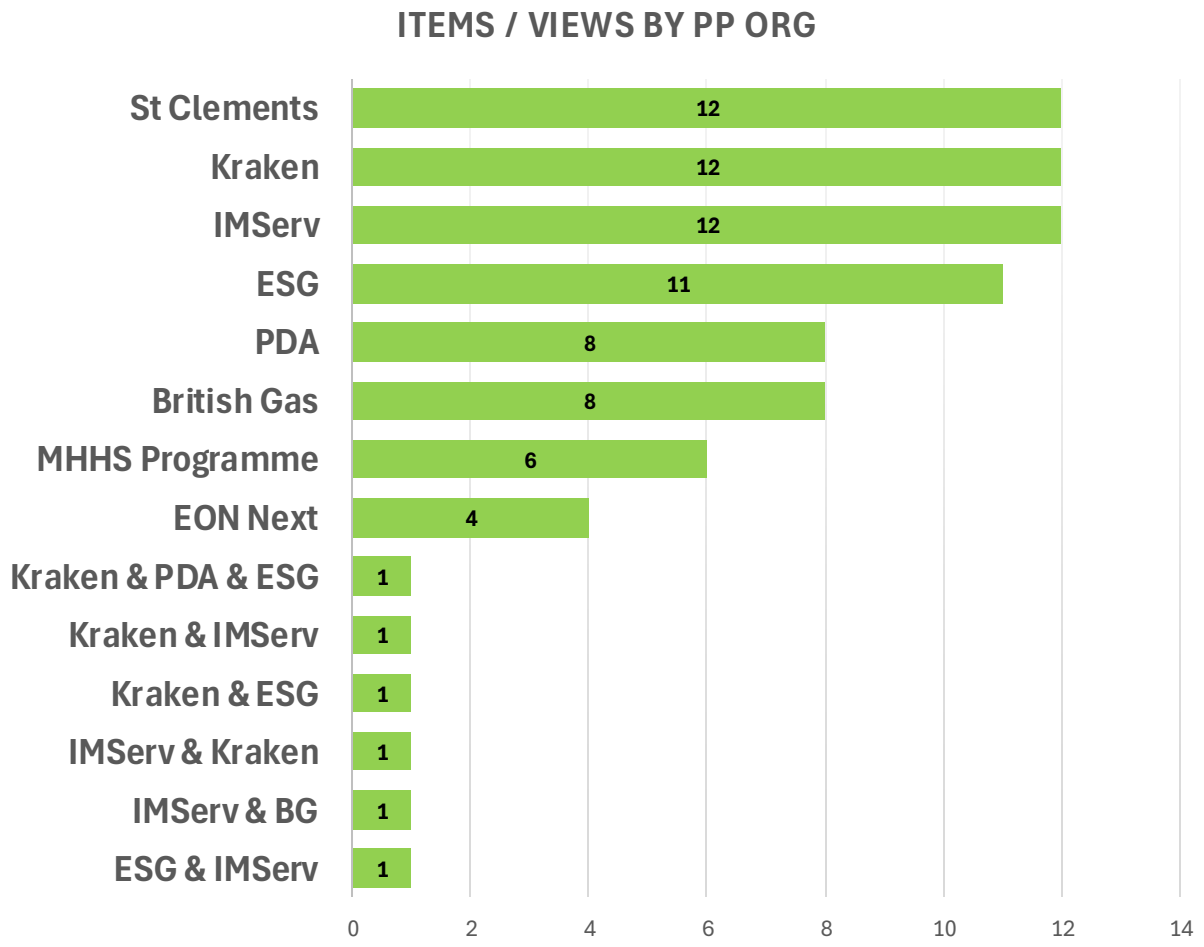
SIT Industry Retrospective Sessions – Summary

- All 4 Retrospective sessions were recorded in Teams, and transcripts have been compiled in MS Excel to record the comments raised by each PP / Organisation in attendance in the sessions

✓ Engagement in the sessions was both active and constructive with 79 items / viewpoints raised in total:

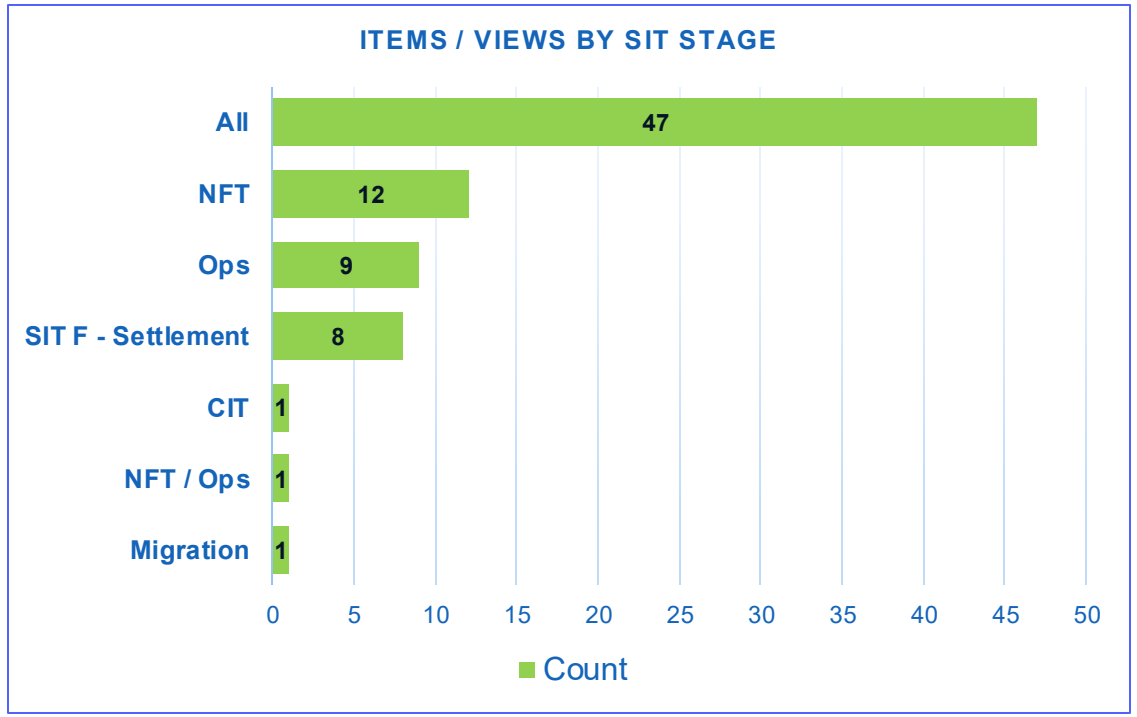
Session	Session Focus	Items / Views Raised
1	CIT / SIT F / SIT M	27
2	Settlements Testing	12
3	SIT NFT & PIT	12
4	SIT Ops, Environments, Defects and Release Management	28
	Grand Total	**79**

✓ Eight individual PP organisations raised items in the discussions (note for 6 items raised, during the flow of the discussion, a viewpoint had been echoed by more than one PP / Org and so these have been logged as 1 item raised, but logged against multiple PPs / Orgs)



SIT Industry Retrospective Sessions – Item / View Analysis

SI Test has compiled all Items / Views raised and categorised by the applicable 'SIT Stage' and into 16 distinct 'Themes':



Themes	Count
Test Approach & Structure	16
Test Cases (Design, Review & Maintenance)	12
Data	8
Defect Management	8
Collaboration	6
Test Scope	5
Test Preparation & Execution Timelines	5
Communication	3
Programme Knowledge	3
Test Tooling	3
Design	2
Meetings	2
Test Assurance	2
Reporting / MI	2
Release Management	1
Environment Management	1
Grand Total	79

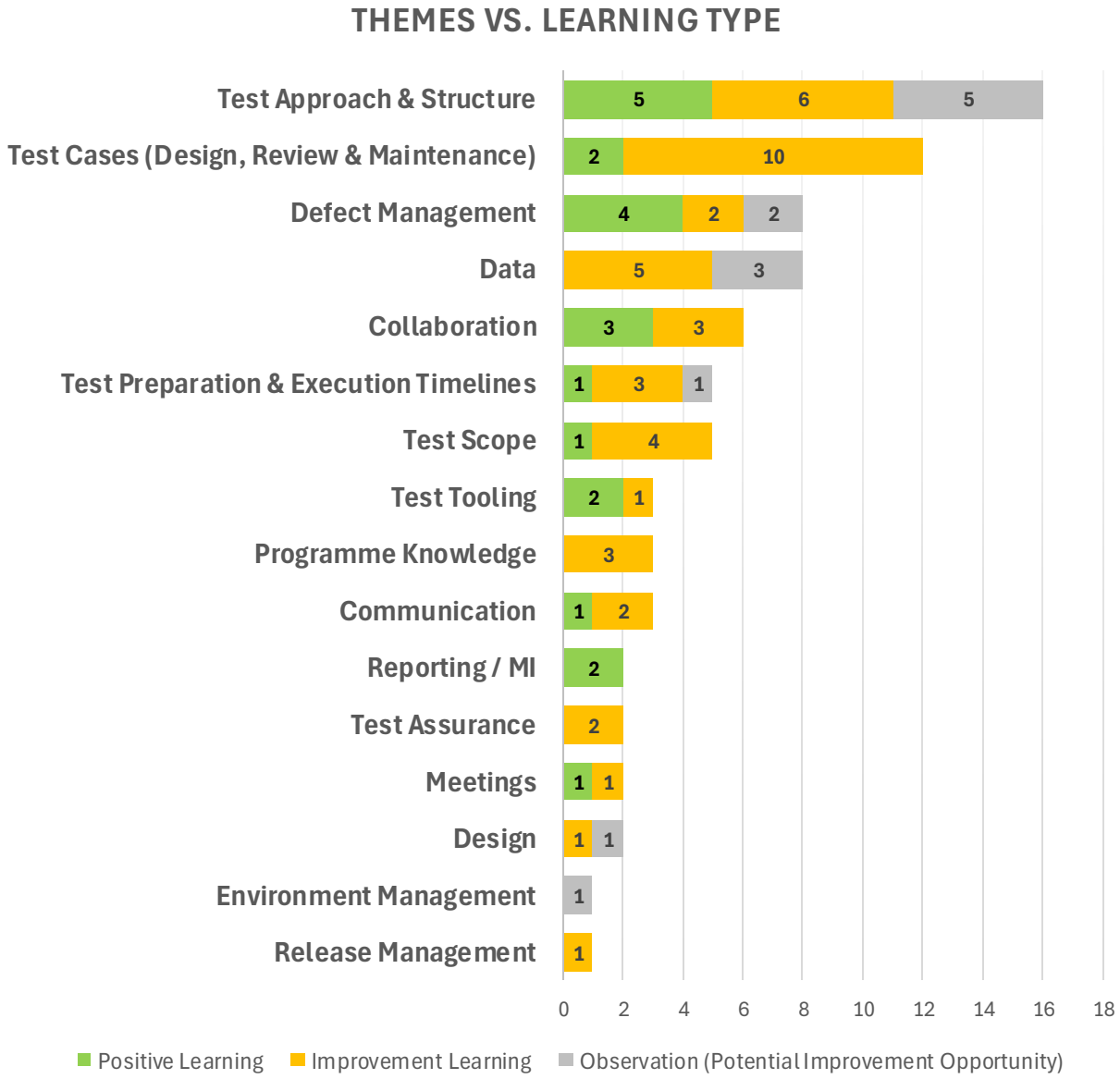
All items were then assessed and assigned into 1 of 3 'Learning Types':

#	Learning Type	Summary	Count
1	**Positive Learning**	These were items or viewpoints expressed where the contributor felt this resulted in more efficient or positive outcomes, and therefore would recommend future industry programmes to consider implementing	22
2	**Improvement Learning**	These were items, where in retrospect, things could have be done differently to improve efficiency or outcomes, and should be recommended by future programmes to mitigate or avoid	44
3	**Observation (Potential Improvement Opportunity)**	These items may have been specific to MHHS SIT and therefore it is debatable if they would be relevant to subsequent industry programmes	13
			79

SIT Industry Retrospective Sessions – Themes Analysis

	SIT Stage							
Themes	All Stages	NFT	Ops	NFT / Ops	SIT F - Settlement	Migration	CIT	Grand Total
Test Approach & Structure	8	6				1	1	16
Test Cases (Design, Review & Maintenance)	6		2		4			12
Data	5	1			2			8
Defect Management	8							8
Collaboration	1	1	2	1	1			6
Test Scope	1	1	3					5
Test Preparation & Execution Timelines	4	1						5
Communication	1	1	1					3
Programme Knowledge	3							3
Test Tooling	2		1					3
Design	2							2
Meetings	1				1			2
Test Assurance	2							2
Reporting / MI	2							2
Release Management	1							1
Environment Management		1						1
Grand Total	47	12	9	1	8	1	1	79

- Note that in the majority of cases the comments raised can be seen as applicable to all stages of SIT



SIT Industry Retrospective Sessions – Theme Breakdown & Synopsis

The retrospective session transcripts are stored on the MHHS Collaboration Base - here

What follows is a synopsis of the points raised by session attendees within each of the 16 ‘Themes’, which has been grouped by ‘Learning Type’ with an SI Test team context summary provided, interpreting the key takeaways and recommendations for future industry programmes.

Contents by Theme (Click on Link)

SIT Industry Retrospective Session Theme Synopsis: Test Approach & Structure (1 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** - The Cohort structure was positive and noted the accountability this enabled between parties, and the paired Cohorts and how this enabled CoS and CoA testing to be well organised (Kraken). • **All Stages** - Cohort and paired cohort stand ups were very useful and facilitated well by SI test coordinators, these enabled all members of the cohort to keep on top of testing status (British Gas). • **All Stages** - The management and coordination of test case execution handover between parties and how this was set up in ADO was smooth, and the role the coordinators played in this was critical to success. The early engagement with PPs and the DITL sessions prior to test execution were very helpful and provided clarity for SIT PPs on how they should engage with the programme (IMServ). • **CIT** - The approach to staggering CIT into intervals instead of 'big bang' with everyone trying to do the same things on day one was seen as an efficient approach (IMServ). • **All Stages** - Moving from self-selection of tests in Cycle 1 to smaller chunks assigned by the programme within a sprint approach from Cycle 2 onwards was a great improvement, the agile principles brought in helped ensure greater progress achievement (E.on).	Future industry programmes that need to incorporate multiple market roles in an integrated testing stage, i.e. one that mimics market conditions, should pay careful attention and planning as to how participants are organised into units that enable the E2E exercising of the design and business processes under test, and should also not underestimate the central overhead and resources required to operate and maintain such a structure in terms of test tooling, coordination, meetings, test support functions and MI. In such a complex setting and structure, being clear in the definition and documentation of what is expected of testing participants on a day-to-day basis, and the DITL processes that underpin this is highly recommended (e.g. Ways of working, test process, data, meetings, test tooling, defect management, environment and release management, escalation routes etc…), it is also crucial to not assume that participant test teams on the ground will be able to understand this without walking this through in sessions with participants, and that these be recorded and available to access for new starters on the programme and PP teams during the whole testing lifecycle. With large numbers of participants engaged in the testing process it was a benefit to stagger the introduction of PPs into testing in a structured manner (in the case of MHHS this was by market role types), this enabled core systems and processes to stabilise and bed down and learnings to be established, both of which benefited subsequent testing entrants. On the directing of test execution and progress, whilst a balance should be struck between test participant responsibility to progress and centralised control of execution priorities, it should be noted that the experience of the MHHS programme showed that participants required more central direction on how and where to progress than originally anticipated, and until this was acknowledged and action taken to change the approach, testing progress was less than optimal. PPs responded to being assigned small chunks of work within short 2-week sprints and this maintained a greater focus and pace for all PP teams, in addition, testing MI became easier for stakeholders to digest and understand the path to completion.

SIT Industry Retrospective Session Theme Synopsis: Test Approach & Structure (2 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - At the outset of SIT, by maintaining a policy of participant anonymity, this drove how to structure things like meetings and ADO, which then meant that there was quite a complex set up in terms of the overhead, the number of meetings and separate ADO projects to maintain, this also reduced overall transparency and made the visibility of the testing status position across all cohorts and participants more difficult (this in turn impacted programme and PPs ability to clearly see the full picture and may have limited opportunities to understand impacts and steer progress). Once it was agreed to remove anonymity post cycle 2, it became much easier to manage testing processes and reporting, as well as reduce meeting and ADO maintenance overheads and gave the programme and PPs better information to make day-to-day decisions on testing, however there were still some residual inefficiencies due to the original structuring that could not be changed at that point mid-flow (MHHS Programme). • **All Stages** - Repeated full DITL pack walkthroughs at later intervals in the programme would have benefited new joiners and existing team members, rather than just an overview of what had changed (IMServ). • **All Stages** - greater central monitoring of Participant test team resource availability and leave could have improved planning and day-to-day progress (PDA).	In the early stages of SIT the MHHS programme adopted a policy which maintained organisational anonymity for participants which had volunteered to participate in SIT i.e. non-Central Parties. These test participants were not publicly referred to by their organisation name, and instead only by the market roles they were fulfilling in SIT. To uphold this policy required that PPs be segregated into multiple ADO projects and comms channels, plus they were anonymised within MI and status reporting, and defects needed to be managed in a way that continued to protect the identity of these participants from each other, which required a parent & child solution to be implemented that created much duplication of effort and challenges to keep all defects up to date. This policy created an overhead in the complexity of these structural and technical arrangements which was inefficient but also obscured the true status of testing and made it more difficult for all participants to benefit from transparency and collaboration across parties and cohorts. Eventually, it was agreed with all test participants that this was not benefitting any participant, and was potentially constraining testing progress, and so the policy was removed, at which point all test participants moved into a single ADO project, and all participants could then see all defects and the status of all testing, and that resulted in greater efficiencies and collaboration and testing throughput was seen to positively increase. It is recommended that future industry programmes implement a policy of full transparency from the outset, and that any SIT participant accepts this as a condition of entry into testing. Over a long testing schedule, it is realistic to assume that central and test participant resources will churn, and it is recommended that all programme testing introductory and guidance materials are available for new staters in one area, and this should include the recordings of any training and guidance sessions held. The MHHS programme and all test participants may have benefited from a central tracker of resource availability across all teams, for PPs to reference. Test participants should plan resourcing to always ensure cover throughout the programme test plan, however, for smaller organisations, or in the case where organisation events occur that interrupt availability, it would be helpful that all test participants are aware and can plan accordingly and maintain momentum.

SIT Industry Retrospective Session Theme Synopsis: Test Approach & Structure (3 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **Migration -** Testing should seek to run processes in the order they would be run at go-live, i.e. starting with Migration (British Gas).	The MHHS E2E test approach had originally assumed that the testing process would follow the order that would occur in production implementation, i.e. migration, then functional activity, however due to delays in the Migration design and development this was not possible and the programme changed approach and initialised test data in a post-migration state as a means commence functional testing, therefore beginning to benefit from exercising the MHHS solution that was available and ready to test. Migration was introduced later when developed by participants. This agile approach, whilst not ideal, helped maintain overall programme momentum and timescales. It is recommended that future programmes seek to follow a live like scenario where possible, but when issues occur that block areas of functionality, they should also explore options that could continue to obtain testing value, maintain pace and progress.
Improvement Learning	**PPs stated [paraphrased]:** • **NFT -** Measurement of NFT success criteria wasn't ideal, one sheet per cohort led to too much granularity and losing sight of the bigger picture. Recommend reviewing test success measurement methods for future programmes (Kraken). • **NFT** - Recommend splitting testing more by test area to keep focus, rather than broad all-in-one approach, recommend future testing could segment activities to reduce getting "lost in the weeds" (Kraken).	NFT success criteria was very clear for Theme 1 and Theme 3, however Theme 2 encountered more issues, due to the significant challenge in trying to reconcile outcomes of concurrent activities across a significant number of PPs. An exact reconciliation approach was landed on which was determined to be the most-timely and least resource intensive given the limitations of execution window and team size. This would encounter difficulties where processes did not run as expected. This happened for various reasons which then lead to a necessary subsequent extended investigative period to verify outcomes. Future programmes should carefully consider the optimal approach to verifying outcomes of concurrent NFT activities involving multiple PPs and processes up to and including scaling up the NFT team size, and to allow for manual verification - working with PPs directly in the case of issues. While splitting NF testing by test area could indeed be helpful, if project timelines allow, there will eventually inevitably have to be 'Production-like' testing where all expected concurrent activities are being processed as they would be in live. Allowing for preceding phases of more targeted testing by specific function/area could serve to mitigate the risks before entering into this necessary final phase, but it is recommended that additional time and resources for such activities would need to be allotted in any Programme plan.

SIT Industry Retrospective Session Theme Synopsis: Test Approach & Structure (4 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **NFT** - Clear goals on what was expected from NFT should have been more defined and communicated from the outset (IMServ). • **NFT** - the 'Big Bang' approach made reconciling failures difficult and forced PPs into granular investigations (MHHS Programme). • **NFT** - there was a situation where multiple participants were often investigating the same issues in parallel, creating duplication of effort (MHHS Programme). • **NFT** - there was a lack of clarity & confusion regarding when Secured Active (Core central activities) was running in the test environment, the DITL pack should have consistently included the Gate Closure information for each testing stage (PDA).	NFT success criteria was very clear for Theme 1 and Theme 3, however Theme 2 encountered more issues, due to the significant challenge in trying to reconcile outcomes of concurrent activities across a significant number of PPs. An exact reconciliation approach was landed on which was determined to be the most-timely and least resource intensive given the limitations of execution window and team size. This would encountered difficulties when processes did not run as expected. This happened for various reasons which then lead to a subsequent extended investigative period to verify outcomes. Future programmes should consider the optimal approach to verifying outcomes of concurrent activities involving multiple PPs and processes, up to and including scaling up the team size to allow for more manual verification and working with PPs directly in the case of issues. It is recommended that DITL specifications are as consistent as possible across all programme test stages, noting that there will obviously be some necessary and appropriate differences due to the approach and objectives of a given test stage.
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **All Stages -** there were benefits and advantages to being a SIT Participant, in that there was an early understanding of the design and to a degree influencing of design direction, non-SIT participants that haven't been engaged whilst the design was being developed are likely to have a lot of queries, and with the design now baselined, they may have challenges they need to overcome purely in their own systems (PDA).	Market participant organisations in future industry programmes should consider the potential advantages offered by early participation in the design phase of the programme in order to build knowledge and influence outcomes, and then with participation in a SIT phase, as this offers their teams the opportunity to learn how the solution operates in readiness for a competitive advantage in production.

SIT Industry Retrospective Session Theme Synopsis: Test Cases (Design, Review & Maintenance) (1 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **SIT F Settlement** - Settlement test cases were very complex and PPs may have benefitted from a more granular walkthrough of the test cases ahead of executing them. This initially delayed PP execution but improved gradually due to the support of the MHHS programme SME involvement, and the implementation of kick off sessions and daily stand ups, which maintained the momentum, however sometimes the bandwidth of the SME could create bottlenecks (Kraken & IMServ). • **All Stages** - we found the programme to be pragmatic and receptive to feedback on test cases, particularly with the script issues that we saw and that was appreciated. We felt that we could enter a dialogue with the program and work out the best way forward (Kraken).	There is no getting away from the fact that test cases and test execution in such a complex industry programme will present challenges for all participant teams involved. Central SME, preparatory, and ongoing execution support helps test participant teams to overcome this, and programme and central party teams should plan to have these types of resources in place, and highly available to test participants, to enable the greatest chance of test outcome success. There was an understanding from the outset in MHHS SIT that even with a high degree of review, there would likely be a need for participants to raise queries and defects against test cases once participant test teams came to execution, and future industry programmes should plan for this in terms of process and resource capacity. There was also the appreciation that in some cases programme requirements, for genuine reasons may not be possible for some participants to exercise in a test environment and that a policy and process for handling dispensations was required. Additionally unexpected events could occur in a test environment, whilst all should be investigated and understood, if they were not a defect and did not undermine the core objectives of a test case, reasonable policies and processes for assessing and permitting tests passing with 'observations' or 'workarounds' are a pragmatic solution. It is recommended that future industry programmes should consider, with the right controls in place, implementing these policies.

SIT Industry Retrospective Session Theme Synopsis: Test Cases (Design, Review & Maintenance) (2 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - there was insufficient time allowed in the plan for the test case review process due to the volume of change in response to review comments, this meant it was not possible for a second round of review. Sufficient time should be built into plans that will allow for multiple review cycles as was the case on the Faster Switching programme, which led to a perceived greater quality of test cases and less subsequent TC defects raised during execution (St Clements). • **All Stages** - need to ensure that sufficient time is built into the plan for internal review ahead of industry review, and to then expect multiple iterations of industry review and build this into the plan. During the industry review period particular attention needs to be given to summarising and communicating the changes from version to version of test cases, including red-lining, at times there was significant change that wasn't easily identified by PPs, for example where the market roles involved had changed, this led to not realising TCs were in scope for a PP and missing the review requirement (ESG). • **All Stages** - reviewing test cases once loaded in the test tool was more practical and easier to carry out in the end state format, it would have been a preference to get these loaded into test tool earlier ahead of execution (PDA). • **All Stages** - baselining of test cases should be achieved before testing commences to reduce fluctuations on scope numbers due to descoping of tests and functionality during a testing phase. More work upfront to rationalise and scrutinise test cases and scope based on design and the production viewpoint may have enabled this (British Gas).	It is recommended that industry programmes of the size and complexity of MHHS acknowledge that central test case design and production will be a significant effort, both in terms of the initial creation but also the internal and industry reviews required to validate the test cases, and that ample resource capacity be available to support this activity both centrally, and within industry participant teams. Realistic timescales should be incorporated that include multiple review cycles and adequate comment response activity periods. It is recommended that the structuring of the programme requirements and design should be driven with a conscious intent to be testable and traceable i.e. without compound or implied requirements, as this could reduce the time and effort associated with test case creation. The of potential AI driven test case creation solutions should also be explored. Sufficient stability within the requirements and design should be established before embarking on test scenario and test case creation, but where this is not possible particular attention and care should be taken to version control test case changes during the review period, and always during a review period any changes should be accompanied with clear ‘red-lining’ and summaries of the version changes being made available to review groups. Test case structure and design should account for the compatibility with the programme’s chosen test tool, to ensure ease of test case loading, maintenance and requirements traceability, to avoid these things becoming cumbersome and a burden of effort. Investment in these things increases the potential to be able to more confidently baseline test cases and scope prior to the commencement of test execution, thus mitigating the risk of scope fluctuation and reducing the risk of test case defects impacting capacity and throughput during testing.

SIT Industry Retrospective Session Theme Synopsis: Test Cases (Design, Review & Maintenance) (3 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - test case complexity was very high, and length of test cases was too long, meaning that if there was a issue / defect encountered near the end, the whole test needed to be restarted. It is recommended that test cases, if possible, be broken down into smaller process chunks or alternatively implement a modular test case design where steps are called from other test cases (ESG). • **SIT F - Settlement** - the large settlement tests involved multiple MPANs, this meant that the whole test failed if only 1 MPAN type had failed, which led to some confusion surrounding what had been achieved and where there were current issues / defects. It would have been preferred to break these down into separate identical tests by MPAN type or segment (e.g. Unmetered), whilst this would have created more tests to execute it would have avoided this confusion, enabling PPs and stakeholders to see with more clarity where confidence had been achieved, current issues needed resolution and areas which were yet to be exercised (Kraken & PDA & ESG). • **SIT F - Settlement** - settlement test cases didn't refer to the exact settlement reports or settlement runs specifically in the tests, in addition if there was incoming or losing supply in the CoS scenario, the exact supply in that situation wasn't specified and that meant participants had to do quite a lot of working out of what settlement report was being referred to at any particular point. In essence with such a complex and critical area there needed to be more specific information and guidance available within the test to enable successful execution. During testing the programme requested guidance videos be created by Helix on how to do the calculations which were exceptionally helpful, however PPs felt that these should have been available from day one due to the complexity of the test cases (Kraken & ESG). • **SIT F - Settlement** - SIT F Settlement - there were very limited negative test scenarios in scope for Settlements tests, some negative test defects were encountered by chance (IMServ).	The MHHS design artefact baseline is extensive and complex. It covers more than 21 interconnected business processes and related method statements, that span numerous industry codes, systems, and market participant roles. There is also a large set of requirements, both explicit and implied, linked to each step in these business process journeys. To keep the design artefacts manageable, the processes were described at a level that assumed a certain amount of industry knowledge from the reader. Additionally, the programme could not know the internal workings of all industry participant systems. As a result, the design narrative focused on describing what events should occur and the expected outcomes, rather than detailing every internal system process. These factors influenced how MHHS test scenarios and test cases were developed. In some cases, these journeys were very lengthy and complex, sometimes involving between 200 and 500 individual steps. There were often many requirements linked to many steps, numerous hand over points between participants and only a limited amount of detailed guidance was included to support people carrying out the tests. During test execution, the effect of such long and complex test cases made it more difficult to establish where confidence had been achieved or where defects and issues were congregating and blocking downstream processing, and this proved a challenge in the early stages of testing. In addition, it made it much more difficult to achieve the successful completion of a single test case, as this could take weeks of activity and multiple re-starts, sometimes nearing the very end of the test script. At a test case unit level, this led to the appearance that there had been much less progress and confidence built than was the case, and to a degree served to undermine confidence in the achievability of the test schedule. Acknowledging this, the programme then began to track test step achievement as a measure of progress, thus more clearly demonstrating day-to-day testing throughput, and then developed a test estimation model which allocated points to different size bandings of test cases, which underpinned more confident test execution velocity estimates and tracking. It is recommended that future industry programmes learn from this and seek, if possible, to break down test cases into a more modular design, with manageable chunks being the subject of test cases, at least in the earlier stages of test, which could then build to longer E2E tests in later stages when sufficient stability and confidence has been achieved. Furthermore, the need to walkthrough test cases with participant test teams prior to execution may help reduce the number of restarts due to execution mistakes, so time and resource capacity would need to be planned for this.

SIT Industry Retrospective Session Theme Synopsis: Test Cases (Design, Review & Maintenance) (4 of 4)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **Ops** - Some Ops Test scripts were incorrect; consequently, test case guidance was issued for some test cases that contradicted what the test case objective was all about. That raised queries and confusion in terms of what PPs were trying to do and whether to follow the guidance or whether to follow the test case, because there wasn't alignment between the two (IMServ). • **Ops** - in some cases Ops tests were very similar to what was being executed in NFT e.g. with DIP outages and certificate issues, essentially duplicating in 2 test stages. Greater attention to cross stage scope could avoid this (Kraken).	There were a relatively high number of test case defects discovered during the SIT Ops execution stage, which highlights the importance of allowing sufficient time and cycles of review in the preparatory stages. In response to this situation, the programme sought pragmatic solutions, via supplementary guidance for test cases, as means to maintain progress and pace whilst test case corrections were being addressed. This approach carried the risk of confusing participants further. It is recommended that future programmes invest the time in the test case review process to mitigate this situation, and if possible, a) include test case walkthroughs and b) if guidance is issued, that the associated communications are clear and strongly re-iterated to participants whilst in place. While some NFT and Ops testing activities did appear to be replicated in terms of steps conducted, the difference that must be considered are the objectives of each test and what outcomes were being validated. If close alignment can be achieved between Ops and NFT streams, then perhaps single test runs could satisfy both OAT and NFT objectives, but it should be noted this would then also create a scheduling dependency between these two workstreams.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Data (1 of 3)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - Having data load validation and analysis as a separate stage at the start of testing would have helped flush out data issues earlier, with validation points ensuring data alignment across all participants and central systems prior to e2e testing. Proving the data using legacy world processes, via legacy regression tests, may have reduced the number of test data related issues which were encountered during testing, but a lack of time in the plan prevented this, so in future it is recommended that a separate early test stage for data validation/analysis be incorporated (St Clements). • **All Stages** - Loading of data was carried out by PPs but needed to be done repeatedly as new versions of data had been published in SFTP and needed to be downloaded, tested and re-loaded. Causing some challenges for PPs due to automated data load. Revisions of data load was sometimes not communicated to PPs, so they had been unaware. Baseline data before asking PPs to load, but if not possible at least version control, and provide clearer comms on data loading revisions (ESG & IMServ).	The programme data strategy was based on the precedent from previous energy industry programmes, where all SIT participants were coordinated to take a production data cut within their systems on an aligned date. This had to be done significantly ahead of time, and prior to when test scenario and test cases had matured. When the scenarios and cases were more developed the programme data team then analysed central systems data and identified MPANs that were suitable to be used for test cases and mapped these in readiness for allocation to SIT participants. Once the cohort model was created this approach proved to be an issue as the parties within each cohort didn't have aligned data sets within the data back up they had taken. In addition, many PPs didn't have the range of MPANs within their back ups to provide test coverage – e.g. a Supplier only had smart meters in their portfolio and no advanced meters. Whilst the loads for cycle 1 and 2 had issues, significant improvements were made as lessons were learned in terms of the quality and scope of the data provided. This included increasing the complexity of the test data provided to include meter data, which was in direct response to PPs requesting the programme provide this data rather than the PPs having to create the data as per the original data plan. Future programmes should not assume PPs back ups are suitable and seek greater assurance to mitigate this.

SIT Industry Retrospective Session Theme Synopsis: Data (2 of 3)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	• **All Stages** - Suggested one central "source of truth" for MPANs across cohorts to avoid duplicate investigations Consider centralised MPAN/test data tracking in future (Kraken). • **All Stages** – the MPAN tracker was not easy to use (lots of filters / freezing panes, due to being used by all Cohorts simultaneously), it should be robust and probably a password protected sheet with more control would make more sense for future programmes (ESG). • **NFT** - Test data issues (incorrect / problematic MPANs) caused confusion and required time-consuming investigations. Improve data quality and visibility upfront; central ownership of test data (Kraken).	The MHHS programme used an MS Excel spreadsheet as the centralised solution for managing test data selection and check out / check in by PPs. A solution was developed to replace the Excel spreadsheet feedback related to the MPAN Tracker in Excel was that PPs were happy with the solution and changing to a new solution could lead to additional complexity. Different participants had different requirements for data and also had different levels of understanding of the data design. The data was controlled centrally within the programme and provided via the programme for loading by SIT participants. Future industry programmes could review how greater levels of data visibility are provided to testing PPs and how verification of data loaded by individual PPs could be verified as being correct via reference to a central data repository.

SIT Industry Retrospective Session Theme Synopsis: Data (3 of 3)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **All Stages** - Data load and data quality improved during SIT by using more realistic data for testing. More consistent and accurate data enabled us to test SIT functionality rather than face data issues. (St Clements). • **SIT F - Settlement** - PPs felt that they didn't have enough MPANs assigned for Settlements testing. That's not necessarily an issue of the settlements itself, it was just the data that was provided, initially in the 'Settling Normally' test case PPs were getting used to the settlement process and the patterns, but when other test settlement tests were executed that impacted MPANs it caused all sorts of confusion and there was no spare MPANs to use instead. So, it's just that generic comment that more data would have been better, to be able to play with may have alleviated this. (PDA). • **SIT F - Settlement** - PPs found every so often that when they were reconciling the data that we were looking at was being impacted by MPANs used elsewhere and that caused some problems because it meant that PPs couldn't reconcile with the results that they would expect. A different approach probably needs to be taken that enables PPs to ring fence the data more easily (Kraken).	Using production data in SIT was crucial to success, it is recommended that future industry programmes adopt this approach for these reasons: 1) Some defects would not have been identified if pure test data was used, real production data issues identified areas of design that required revision. 2) Further refinements to data were developed across all stages of SIT which resulted in fewer data related blockers to testing as SIT cycles progressed. 3) Production data usage, including data taken from PPs production systems data backups, ensured that data was correctly aligned across all systems. During Settlements there was a constraint within Helix test systems that limited MPAN Settlement reconciliation to a maximum total of 400 MPANs. This limited the flexibility of MPANs available to be used by all SIT participants, and due to compressed timescales creating a need to run different Settlement test scenarios in parallel, this carried an acknowledged risk that the impact on MPANs could confuse the reconciliation results, which materialised. Enabling a greater set of MPANs in this testing would have allowed for greater flexibility to ring fence MPANs to a greater degree and reduce the effect of confusing reconciliation results.

SIT Industry Retrospective Session Theme Synopsis: Defect Management (1 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** - the structure of the daily defect calls made them quick, simple and easy, and were well organised and consistent in the way they were always run. There was a real positive benefit of having a separate defect management MS Teams channel to engage on when needed (PDA). • **All Stages** - Defect calls went very well by the end of it, and it could be seen how these improved as they went along. (St Clements). • **All Stages** - the defect management and the release management was fantastic in the way it was dealt with, not least as there were so many defects. There was a great improvement once the anonymity policy was dropped, as PPs could see each other's defects and there weren't child records clouding the picture (St Clements). • **All Stages** - defect reporting was very good and well received (British Gas).	MHHS benefited from a clearly defined defect and release management approach, which was consistently administered through all SIT stages. MHHS maintained a culture that if testing participants encountered unexpected outcomes in testing, that decisions should not be made at a local level as to whether a valid defect had been encountered or not, but instead a defect record should be created so that this could be centrally assessed to identify the cause and the correct course action to resolve, be it a fix, clarification or rejection. The daily in-flow of defects and management of each defect through to resolution required suitable resource capability (both SMEs and process managers) and significant capacity within the programme team and all testing participants, and involved a high degree of engagement, communication and collaboration across all teams to make the process a success. It is recommended that future programmes (both centrally and within industry teams) invest in the right skills and mobilise dedicated resource capacity to meet the high demands required to investigate, manage and resolve defects, and that this is built into resource plans to ensure success. As is also recommended in the **Test Tooling** section, particular attention should be focused on how the defect module is set up and configured within the programme selected test tool, with the design of the defect workflow controls and MI/Reporting requirements clearly defined very early on, so that a tool selection can be made that enables the customisation required to meet the complex needs of any similar industry programme. It is also recommended that in the case of MI/Reporting requirements and any fields in the defect module that drive reporting, be widely agreed ahead of test execution, as it will become increasingly difficult to implement any new reporting requirements once testing has commenced and the defect database has begun to be populated.

SIT Industry Retrospective Session Theme Synopsis: Defect Management (2 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - We found that a lot of defects ended up with us because MPRS is very central to everything. Initially, a lot of defects came straight to us when in fact when we looked at the design we worked out that it wasn't our issue, but things regularly just got sent (St Clements). • **All Stages** - The defect calls did work well, however one improvement would have been to get an agenda issued before the meetings, because there were occasions where we were expecting responses about defects and they were missed i.e. they had been mentioned the day before. We also observed other times there weren't any time scales, or perceived urgency on some of the defects, whilst to a degree this was dictated by releases and how quickly people could triage internally, more SLA accountability in the meetings may have helped move things along (ESG).	The MHHS defect triage process first assessed that defects met the minimum requirements for quality of information, payloads and logs as supporting material to investigate. The programme triage team, which consisted of key design and testing SMEs, would assess the information and make decisions on which resolver group to assign the defect to. The central triage team did not have access to internal design or SME knowledge of central systems, so on occasions their judgement calls could be challenged, this improved over time as knowledge naturally increased, however future industry programmes may consider whether central party SMEs should be more closely integrated in the central triage function, in order that defects could be triaged more effectively. Ideally, good practice involves providing meeting agendas, and actions should be made available after / ahead of each meeting, however this sits within a broader topic of meetings overhead, and the resource capacity required to support this, covered under the **Meetings** theme. Defect resolution SLAs were documented in the Defect Management Plan, however these could only ever be seen as guidelines, as no organisation was commercially or contractually obliged to meet them. Future programmes might seek to either create contractual incentives and obligations, where possible, or seek a regulatory directive that could be used to enforce SLA compliance.
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **All Stages** - Early in the process, defects were closed without the raiser's knowledge. So, there were times when we weren't aware an issue had been successfully resolved (St Clements). • **All Stages** - Defect emails were going to the MHHS emails that had been created; that we didn't have access to (ESG).	The process for closing defects was changed and improved early on during testing by the defect management team, whereby the policy became that defects were only closed with agreement by the raiser. The recommendation is that future programmes should consciously decide on the policy for how defects are closed, and by who's authority this decision is made, and then be consistent in the application of the policy throughout the programme testing. By a quirk of licencing, MHHS SIT participant MS users were created for the purpose of ADO and MS Teams access, and in the set-up process were given assigned email addresses, however these email addresses weren't given email accounts, and this wasn't initially widely understood and led to confusion. The programme had always intended on using MS Teams as the primary channel for SIT communications and this proved to be a very effective tool throughout.

SIT Industry Retrospective Session Theme Synopsis: Collaboration (1 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** – in our cohort, we had brand new people in teams who had never all worked together, and you have to find ways for those people to all work together, get to understand each other, and become friends with each other and be able to talk at the end of the day. If you don't foster that, you don't have a cohort that can work together and that took time establish before the cohort could function effectively. Different cohorts teams would have had different dynamics and it should be acknowledged that there needs to be some time assumed that will be required for people to build relationships and to establish ways of working together (E.on Next). • **NFT** - Programme team were always on hand and approachable during NFT (Kraken & IMServ). • **Ops** - Coordination support, especially during the Service Management testing was good. The stand ups worked well in OPS, and all other test stages (IMServ).	MHHS sought to embed a culture of collaboration in the testing programme from the outset and implemented various approaches to foster this. During SIT, testing participants were organised into 8 Main 'Cohorts' each comprising the MHHS Market Roles required to exercise the design, E2E business processes and central systems within in the new MHHS arrangements. To enable Change of Agent and Change of Supplier theme tests, the main Cohorts were paired to enable E2E MPAN gaining and losing scenarios. Central Parties (Helix, RECCo, DIP, DCC and Electralink) were members of each Cohort, with their MHHS requirements, designs and development being exercised by the testing of each of the 'Main' and 'Paired' Cohorts E2E tests. Each Cohort was assigned a central test coordinator (4 in total) covering each main cohort and the associated cohort pairing. The programme test coordinators were the primary point of support contact for test participants on a day-to-day basis, facilitating stand ups, maintaining test case assignments in ADO, test case handovers and driving an ongoing support dialogue with the cohort members, advising on defects and priorities, and maintaining alignment across cohorts in the testing effort and DITL operations. A wider set of testing SME and support resources were also in place centrally, covering; design, data, defects, releases and environments. This was all pivotal to the testing success, and it is recommended as a model for future industry programmes and should form part of resource planning. Cohorts were also encouraged to build their intra-cohort relationships and ways of working in a way that suited the dynamics of their groups, that fostered a responsibility to self-manage as much as possible. During the formation of cohorts, introductory meetings were held and in the early stages leading roles in those teams tended to emerge, in most cases within the supplier market role. For Central Parties, theirs was the challenge of supporting the testing of all Cohorts. The programme maintained regular daily contact with this group separately, to ensure that any issues or constraints with their support could be understood and dealt with fairly, and responses effectively managed and communicated across cohort testing. Ongoing daily contact within an array of stand-up and programme test meetings, and specific Cohort MS Teams channels (some with specific purposes e.g. data, or defects), ensured the ongoing dialogue of all teams. The programme consciously ensured that all voices were heard and all contributions were dealt with fairly. To reinforce this, a code of respectful behaviours that was expected of all participants was documented and all parties were seen to adhere to this throughout. It is recommended that future programmes consider the criticality of the human dynamics and team building in a successful delivery and consider their approaches to enable this.

SIT Industry Retrospective Session Theme Synopsis: Collaboration (2 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **SIT F Settlement** – the delays in response from Helix SMEs impacted progress (ESG). • **NFT / Ops** - we felt that there could have been a little bit more coordination between Ops and NFT, as these stages were at times in a running concurrently in the same environment. Testing got a bit messy, because running Ops involved DIP outages at times that impacted NFT testing, then Ops testing needed to be paused for a fair bit of time whilst NFT took priority and so there were consequent delays. (Kraken). • **Ops** - More communication of when there were defects impacting across cohorts. i.e. as the knowledge improved there was a consultation with other cohorts. This didn't go well at the beginning. If a defect was found, the tests were blocked in other cohorts, but there were occasions where we had already gone past those steps, and the test was then blocked which meant we consequently had to rebuild / re-run the test case within ADO, which could have been avoided by pausing the test , had we been aware. It a recommendation that more cross cohort consultation occurs where defects are found by one cohort that may also impact others. (ESG).	It is recommended that central party organisations should consider how to best utilise their SME resource, by enabling sufficient capacity within their broader teams so that those SMEs can be more accessible to the programme and testing participants in real time, as this will ensure greater chances of success for all participants. While some testing activities in SIT Ops and NFT did appear to be replicated in terms of steps conducted, the difference that must be considered are the objectives of each test and what outcomes were being validated. If close alignment can be achieved between Ops and NFT streams then perhaps single test runs could satisfy both OAT and NFT objectives, but this would then also create a scheduling dependency between these two workstreams. In many cases issues caused by blocking OAT and NFT activities were down to the use of a single shared environment. Ideally a dedicated environment could be provided for each activity or else NFT and OAT should be executed sequentially and not in parallel, but this approach comes with its own associated challenges and risks – see **Environment Management**. During MHHS testing it was critical that all parties had a good understanding of areas of stability, and where open defects were having an impact, and could then plan where to progress and avoid accordingly. A key recommendation for future programmes is to avoid any kind of anonymity policy, as when in place on MHHS this served to decrease transparency and constrain this dynamic capability, and in cases this caused some wasted tested effort.
Observation (Potential Improvement Opportunity)	N/A	N/As

SIT Industry Retrospective Session Theme Synopsis: Test Preparation & Execution Timelines (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** - Testing had to start at some point and although a little more prep time was needed, testing had to start to uncover the issues during test and iron out teething issues, there is only so much planning that can be done and felt we started testing at the right time (E.on Next). • **All Stages** - we shouldn't have gone into SIT as early as we did while the design was still not fully stable, and things were still being developed. With more stability and a little more understanding the transition into SIT could have been smoother and may have resulted in faster execution timescales, but the collaboration between teams was very good and that carried it through very well, the fact that teams were very engaged and everybody was approachable and easy to get hold of most of the time really helped (St Clements).	The initial proposed timescales for the MHHS programme were inherited, the programme then developed the planning and estimation of the timescales required for development, PIT, SIT test preparation and execution under CR022, factoring in learnings from DCC and Faster Switching, and in particular with the latter, via an analysis exercise that contrasted the design scope and complexity versus MHHS to understand the relative scale. These findings were extrapolated, combined with planning assumptions about where efficiencies could be built in from the experiences of those previous programmes, and along with industry consultation this formed the MHHS programme test plan and schedule. PIT completion and SIT readiness assurance controls were built into the plan, then tracked and met, underpinned by Participant self-declarations of readiness to commence, and on this basis the governance decision was taken to commence SIT testing. As is stated elsewhere in this document, in retrospect the programme may have benefited from more time to prepare and industry review SIT test scenarios and test cases. Test case related issues would have very likely impacted test execution throughput due to the time and focus this demanded from all parties to understand and resolve, however this is difficult to quantify, but it is recommended that future industry programmes emphasise the importance of planning realistic timescales to mitigate this factor. That said, the relatively low percentage of severe defects in SIT that resulted in code fixes required, does suggest that participant systems were generally ready to be exercised, and all SIT participants benefitted from commencing testing at the time the programme did, as this served to establish that confidence. The lessons the MHHS programme learned regarding the structuring of the design and the relative complexity and sizing of tests cases, the chunking up work into more manageable sprints and the associated modelling of test execution, as described elsewhere in this document, are all factors that future programmes should take account of in streamlining their timescale planning and approach.
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - We felt a key lesson learned should be to allow more time for preparation, because due to lack of time, due to rushing in to SIT before everything was in place, we felt this created issues in execution that could have been avoided with more time for preparation, so the design wasn't complete, test cases weren't reviewed sufficiently, and evidence requirements weren't specified well enough. It is understood everybody's trying to do agile, but is agile really relevant within a programme of this size(?!), agile should ideally be used in much smaller projects rather than huge programmes of work like MHHS. We gave a time scale of how long it would take to deliver, to develop and deliver code and we were not given the length of time that we required to get all our ducks in a row prior to starting SIT, and given more time there would have been a lot less issues, and SIT would have been over a lot quicker (St Clements). • **All Stages** - Setting smaller targets to begin with e.g. sprints, gives people more incremental confidence, which was difficult to obtain in the early stages where there was so much testing to cover. Assumptions of how much can be achieved in early stages of testing should be realistic (E.on Next).	
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **NFT** - Time constraints (limited environment access window) added pressure to participants and investigations (MHHS Programme)	See the **Environment Management** section.

SIT Industry Retrospective Session Theme Synopsis: Test Scope (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **Ops** - Ops Testing gave us a good view of the ServiceNow. It also gave us a good view of the DIP portal to see how that was going to work. It gave us a good insight of the reporting and how it was all going to hang together as DCPs that are going to be responsible for a number of clients. That was quite useful (ESG).	Market participant organisations in future industry programmes should consider the potential advantages offered by participation in a SIT phase, as this offers their teams the opportunity to learn how the solution operates in readiness for a competitive advantage in production.
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** – it was a challenge to understand the rationale behind descoped tests and getting the information related documented and provided to industry in a timely manner to enable the PP to internally digest, plan testing and assess risk profiles. Recommend more consultation with PPs on test de-scoping (British Gas). • **NFT** - It would have been good to have clear and unambiguous Non-functional requirements from the outset (IMServ). • **Ops** - stand up meetings seemed chaotic at times, because of the number of tests being questioned and then ultimately being descoped was numerous in this test stage. In some cases, tests were confirmed to be in scope, but after further discussions they were descoped. It meant there were several iterations of the Ops Plan, which impacted our ability to do resource allocations effectively (ESG). • **Ops** - In the original Ops E2E guidance, the assessment did not have supplier and MS roles in OPs testing and so we planned accordingly. Then after a couple of version changes these roles were in scope, without any calls or consultation, which left us little time to prepare and resource (ESG).	The MHHS design artefact baseline is extensive and complex. It covers more than 21 interconnected business processes and related method statements, that span numerous industry codes, systems, and market participant roles. There is also a large set of requirements, both explicit and implied, linked to each step in these business process journeys, that span numerous industry codes, systems, and market participant roles. There is also a large set of requirements linked to each step in these business process journeys. Test case journeys were consequently very lengthy and complex, sometimes involving between 200 and 500 individual steps, and there were often many requirements linked to many steps. Test cases went through internal and industry reviews cycles, and during the creation, review and early stages of test execution, elements of the design had still not reached full maturity or stability. These factors meant that there were defects raised against test cases once participant test teams came to execution, which in turn required test cases to be revisited and corrected. Through this process and the associated learnings regarding requirements coverage rationalisation, there was a consequential impact on the test scope baseline. Where this occurred, any scope changes were documented, with the associated rationale and socialised with test participants via SIT Working Groups and then approved at the SIT Advisory Group. To mitigate this scenario it is recommended that future programmes appreciate the importance of; a) creating a design and requirement traceability structure that is explicit and driven to be testable b), there is sufficient stability within the requirements and design established before embarking on test scenario and test case creation c), there is sufficient and realistic timescales for test scenario and case creation and reviews built into plans (including multiple review cycles and adequate comment response activity periods) d), at all times during a review period any changes should be accompanied with clear 'red-lining' and summaries of the version changes being made available to review groups, and e), that there is an appropriate industry consultation and governance mechanism to monitor and validate any scope and traceability changes.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Communication (1 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **NFT** - Communication around NFT was clear, approachable, with consistent touchpoints (Kraken).	Multiple communications channels were set up for MHHS industry SIT testing. Primarily communications were initiated within the programme governance framework, with the monthly and ad-hoc working groups (i.e. SIT WG, Migration WG, Data WG, and Environments WG) which were open to all programme participants to attend and contribute to, and which fed recommendations for key decision to be made at the SIT Advisory group. Alongside were the following regular communications mechanisms that were established: • The Programme Participant Coordination (PPC) Team regularly kept participants up to date during preparation and execution via the weekly 'Clock' newsletter, or with ad-hoc comms via email or on the programme website / collaboration base where announcements or notifications needed to be communicated • A testing mailbox where test participants could contact the programme test team with queries or issues • Programme test team ad-hoc bi-lateral test participant meetings • PPs had a regular bi-lateral meeting and comms channel to raise any issues or queries to PPC team • DITL and guidance sessions ahead of each testing stage or major test stage cycle • The programme implemented a framework of daily meetings and stand-ups (see **Meetings** section) • Multiple MS Teams channels built around the test stage and Cohort structure, with specific sub-channels for data, defects, environments and releases (see **Collaboration** section) • Daily comprehensive test and defect status reports to test participants during each testing stage • Industry Testing Retrospectives following Cycles 2 and 3 • Detailed interval test status management reports following sprints and cycles • Regular status updates from the Programme Test Team at wider programme forums i.e. FTIG, QWG, PSG All these channels were crucial for maintaining a dialogue with test participants and programme stakeholders throughout the programme journey. The complexity involved in the programme required that at every step this level of communication was required to convey key information, seek industry consultation, foster collaboration and avoid confusion, provide clear tracking of progress and issues, identify course corrective options, obtain consensus decisions on how to progress, and maintain the level of assurance required on an industry-wide programme of this scale. Future industry central programmes of this size and complexity should not underestimate the resource capacity and capability required to operate this level of engagement and plan accordingly. Similarly, programme participants should assume the necessary resource capacity to be able to consume and respond to the level of information flow required to stay on track.

SIT Industry Retrospective Session Theme Synopsis: Communication (2 of 2)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - Response times on testing mailbox did not meet expectations and sometimes did not get responses in a timely manner (IMServ & British Gas). • **Ops** - We saw lots of duplicate comms regarding releases coming out across different cohort MS Teams channels with the same message. It was coming through multiple times in multiple places and probably quite a significant overhead in the program to issue that same comms out eight times into each cohort channel (IMServ).	The MHHS programme underestimated the level of correspondence that would be required with participants, primarily during the preparation stages of programme testing and found the demand from participants for the programme test team to answer technical queries and questions was reliant on too few SMEs who were also covering many other tasks. It is recommended that future industry programmes expect this demand and have dedicated resources in place that can handle this day-to-day correspondence, are able to provide a consistent messaging to participants when faced with common questions, and a mechanism for efficiently connecting with SMEs when queries are more complex. Due to cohort structure, and a legacy of the earlier anonymity policy and how this had influenced the foundations SIT structure, it did result in some inefficiencies and overheads in communications. The programme erred on side of caution to ensure all participants received necessary comms, but this could mean that central parties and some participants that operated in multiple cohorts or market roles could unfortunately receive a dearth of duplicate communications. A recommendation for future programme could be to have a single broadcast only channel for common updates like releases, deployments, downtimes, P1 defects etc…
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Programme Knowledge (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - St Clements ended up answering a lot of design questions, and a lot of questions about test cases, which it's not just our design, it's the broader design. The design was there, but people were not clear where to find the information and then consequently raised defects which could have been avoided, plus anonymity and the segregation of cohorts also appeared to compound the issue. Improvements in knowledge building were seen as things developed, but perhaps more time should have been spent pre-SIT starting ensuring that knowledge was in place within teams (St Clements). • **All Stages** - whilst we had all the design documents, we didn't have the design documents or internal knowledge for MPRS (MHHS Programme). • **All Stages** - the design is incredibly complex and because everything was done so quickly, people didn't have a chance to understand before SIT, however, SIT participants have now learnt a great deal through the experience, but even so in Regression we were still getting some of the same errors relating to knowledge gaps, so the concern is that once we do go live, there will continue to noise and traffic due to lack of design knowledge and how the MHHS solution actually operates, and this will not be a smooth process (St Clements).	At the beginning of testing, as is natural, it took time for the programme team, participant test teams, supporting SMEs and testing support functions to get to grips with the design & solution functionality, and the test programme ways of working. The programme set up guidance & kick off sessions ahead of testing stages and cycles, and for key complex areas of test, notably Settlements. A design knowledge repository was built and published on the programme collaboration base and was maintained throughout SIT and this proved to be key resource for participants. In some cases, Central Parties provided key guidance documents, 'how to' videos and MS Teams sessions which greatly assisted, but was often provided only when it became clear that participants were having challenges in certain areas. Whilst to a degree, test team knowledge building will be organic, future programmes may want to consider accelerating this process specifically for central systems and services that all participants use, for example via early SME-led walkthroughs for central systems and shared internal documentation, so this can be used as a reference during test case creation and defect triage. Central organisations should also consider how to best utilise their SME resource, enabling sufficient capacity within their broader teams, so that those SMEs can be more accessible to the programme and testing participants in real time, as combined, this may be a means to reduce the number of queries and defects being raised, assigned and responded to during testing. It's also recommended that a tagged, searchable repository for the design is created, and design artefact orientation sessions held. In addition, non-central test participants should consider investing in an SME component within their teams, that can maintain their local knowledge base and support other members of their test execution team with accessing the available knowledge resources, as this could reduce the number of queries and defects being raised by their teams and consequently increase testing throughput.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Test Tooling (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** - ADO Test case management had a centralised view including the split across Cohorts and then collaboration between participants supporting each other was very helpful (IMServ). • **All Stages** - The overall control of the sprints using ADO was very good, so it was easy to see what was going on (British Gas).	It is recommended that future industry testing programmes invest heavily upfront in defining the central test tool requirements in terms of functionality, configurability, scalability, test case loading and maintenance, requirements traceability, reporting and MI capability, user administration and controls, and licencing costs, and then compare multiple solutions in the marketplace to assess which one will best meet the specific needs of the programme.
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - The test case structure in ADO was not good for Ops, as we had all the test cases in one master view, unlike in functional where this was structured around the test plan, sprint and cohort, and so in Ops that didn't necessarily work as well in terms of being able to track what the last run state of the test case was because when one test showed up as passed, it was showing as passed irrespective of Cohort; or rather it was it was kind of cohort agnostic in terms of the test result. So it wasn't necessarily clear that every cohort had passed it, which in most cases wasn't the case. For example 3.2 was split by participant, but again, because there was still only one instance of the test case itself, you still had that overlaying problem (IMServ).	Along with the general requirements of the selected test tool it is also important to consider the structure of test case execution and defect management modules within the tool. Planning how to structure test stages, master test case repositories, test case identification and segregation, test plans, testing of specific processes or functional areas, cohorts of participants, sprints, test case handover between participants, requirements status, the management of dispensations, or tests that passed with observations or workarounds, etc… This is crucial to the successful coordination of testing, visibility of status and potentially dictates reporting / MI and therefore should be carefully considered, designed and communicated to test participants prior to execution. Once established this should remain consistently applied throughout the programme testing so as to avoid any confusion amongst testing participant teams.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Design (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - At the outset, a misalignment arose between the interface design and the Swagger specifications, leading to challenges (ESG). .	To mitigate similar issues in future projects, it is recommended to establish and communicate from the very beginning, which artefact serves as the authoritative source for delivery and design. This approach will ensure that all Participants are aligned and working from a single, clearly defined reference, thereby reducing the likelihood of duplicated efforts and miscommunication.
Observation (Potential Improvement Opportunity)	**PPs stated [paraphrased]:** • **All Stages** - the design as it currently stands is now frozen and everything is in Code. Now the design documents won't be being updated anymore, and this may be a challenge for organisations who have not been involved in the SIT stage. (MHHS Programme).	Market participant organisations in future industry programmes should consider the potential advantages offered by early participation in the design phase of the programme in order to build knowledge and influence outcomes, and then with participation in a SIT phase, as this offers their teams the opportunity to learn how the solution operates in readiness for a competitive advantage in production.

SIT Industry Retrospective Session Theme Synopsis: Meetings (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **SIT F - Settlement** - Settlement testing had the clearest guidance with the daily stand-up sessions (PDA).	The MHHS programme understood the need to maintain regular direct contact with test participants leading up to and during the test execution process, this initially enabled key information to be conveyed, but was also a crucial factor in establishing the working relationships between the programme and amongst testing participants within their cohort groupings, and continued to be a vital means to discuss, review, respond and dynamically coordinate in a constantly changing status position throughout testing. During SIT Functional and Migration, daily stand up meetings were held for each of the 8 cohorts, and each of the 5 paired Cohort groupings, along with the daily defect triage and update meetings, a stand up specifically for Central Parties who were supporting all cohorts, and internal programme stand ups. This meant the daily fixed schedule of test related meetings facilitated by the programme totalled 20 per day, with this number only increasing when SIT NFT and Ops came online in parallel in the later stages. We were conscious that individual test participants would be committed to their own internal daily meetings, and participants were also attending the broader array of MHHS governance and working group forums in the diary. In the earlier cycles of testing, Cohorts were more segregated to preserve the participant anonymity, as this policy was removed after Cycle 2 it enabled greater opportunity for ad-hoc meetings to be set up that involved all test participants and this increased the ability to foster collaboration, however not all participants were confident to contribute in forums with much larger groups of attendees, and the existing meetings were established and Cohorts were working together well, so these remained to ensure that continuity. Future industry programmes should carefully consider their approach to testing meetings; how many are required, how they are structured (with clear objectives), who attends and what the frequency should be, and it should not be underestimated how much time will need to be committed to supporting this overhead by both by central programme resources and test participants. Striking the right balance on the level of engagement via meetings will be a key success factor, and programmes should be conscious that some participants may fall into multiple groupings and so this should be catered for to ease that burden. In conclusion, regular consistent contact was seen to provide the necessary 'drum beat' for maintaining momentum, too little contact runs the risk of participants losing direction and potentially falling behind plan, future programmes of similar size and complexity should assume a high degree of meeting overhead, and all participants should consider this overhead when planning the optimal dedicated resource capacity required to cover all programme testing demands.
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - One thing though to be careful of, is we found that the number of daily meetings was high. Especially for us, this is slightly a Kraken problem because we were across 2 cohorts, but even so, if you account for Functional and Migration stand-ups, joint cohort stand-ups, daily internal meetings, daily defect meeting, it became quite a lot. So, we'd suggest next time just having alternating days maybe or doing just ad-hoc ones when there was plenty of content to cover and rely on other communication channels like MS Teams (Kraken).	
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Reporting and MI (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	**PPs stated [paraphrased]:** • **All Stages** - The daily reports were extremely thorough and detailed and clear in terms of who was where and what every Cohort was doing (PDA). • **All Stages** - Once we lost the anonymity, we could see what was going on in other places. And could get a sense of what was happening ourselves. (British Gas).	It is recommended that future programmes identify and define test reporting and MI requirements at the earliest point when test and defect management tool selection and configuration is being defined, as it will be more difficult to respond to changing MI requirements later down the line during the testing process, where bespoke or manual reporting solutions are also likely to be more difficult to implement, costly and time consuming to maintain. It is recommended where possible to avoid anonymity within reporting and MI, as this can obscure the true status of testing, or patterns / impacts of central issues, and therefore potentially reduces the effectiveness of MI to inform programme and test participant decisions on progress and course correction, for instance in areas where confidence has been established or defects are potentially impacting multiple similar participants.
Improvement Learning	N/A	N/A
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Test Assurance (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - Test assurance responses to test cases that have been executed a long time ago proves difficult to review and re-evidence after a long-elapsed time. Rather than a big bang approach later in the testing process, earlier and ongoing assurance closer to real time and issues would enable issues to be addressed sooner and with greater ease (IMServ). • **All Stages** - We observed there was sometimes inconsistencies in what Central Party evidence was requested, which later identified by Test Assurance in some cases 12 months down the line. If there had been time been allowed to get the requests initially consistent and correct, this would have been avoided. Recommend better planning from the start on who needs to capture evidence for which test cases, so that the participants are aware from the outset and can plan and capture accordingly (St Clements).	It's recommended that future programmes do not underestimate the resource effort involved in test assurance reviews of industry participant test case execution, ranging from verifying that evidence is present and supports the outcome of the tests, to ensuring that sensitive un-redacted data has not been uploaded to the programme test tool which then needs addressing. When test assurance non-compliance has occurred, there is also a significant communication and subsequent corrective action validation effort required which all takes up time and effort for central programme resources. It is recommended that test assurance compliance expectations are clearly set out for testing participants before execution commences, and that this expectation is constantly messaged and reinforced by dedicated resources throughout the testing life cycle to ensure it is at the forefront of minds, equally testing participants should ensure sufficient resource capacity to meet this requirement and assume this to include corrective action effort. In addition, it is recommended that the central test assurance team is sufficiently resourced to enable the test assurance compliance reviews to begin immediately after test execution has commenced, thus ensuring that compliance issues can be identified and corrected as close as possible to test execution, and to avoid the scenario where historical test case compliance issues become more difficult to address once participants systems have moved on, mitigating the need to re-execute tests to resolve non-compliance. Test case steps that required evidence to be provided were clearly identified within test cases, and this is recommended, however it was identified during testing that more evidence was requested than was required to validate outcomes, which was having the effect of slowing down test execution. Test case evidence requirements were then subsequently reviewed for all test cases and reduced later in the testing process; this resulted in a great many test cases needing to be updated (central time / effort). It is recommended that particular attention be focused on limiting the amount of evidence required on tests, and to model the compound time and effort impact this will have on test participants. MHHS adopted a 2 of 8 test case evidence policy for Central Parties, which was positive to avoid an unnecessary duplication of effort burden for those parties, however this approach requires careful planning, communication to test participants and dynamic management where non-compliance has occurred, it is recommended that future programmes implement an effective management solution for this to avoid in some cases not obtaining the necessary Central Party test evidence, as was the case on MHHS.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Environment Management (1 of 1)

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - switching environments caused some challenges. Environment stability for future testing would reduce the overhead of setting up and configuring new environments during testing (Kraken).	Future industry programmes should acknowledge that the establishment of an industry wide e2e test environment is complex to stand up, configure, load with aligned data across testing participants, and takes time to stabilise sufficiently to enable effective test execution, paired with an associated code maintenance overhead and this proves costly for all participants to build and maintain. Consequently, the number of test environments planned should be carefully considered and where possible limited, either by not attempting to run too many testing activities in parallel within the testing programme, or by running testing activities sequentially and repurposing established environments throughout the life span of the programme, as this will reduce the complexity, risk and costs associated with running multiple environments in parallel.
Observation (Potential Improvement Opportunity)	N/A	N/A

SIT Industry Retrospective Session Theme Synopsis: Release Management

Learning Type	Synopsis of Attendee Comments	Summary of Recommendations for Future Industry Programmes
Positive Learning	N/A	N/A
Improvement Learning	**PPs stated [paraphrased]:** • **All Stages** - it was observed that time was lost by not releasing code during outside of testing hours (9am – 5pm), especially in those critical days when we were trying to get through a lot of tests and we were behind schedule (British Gas).	On MHHS it was not possible to secure agreements with central parties and core providers to deploy fix releases outside of business hours, not least because this was going to be a very long period of time that such a capability would be required. To limit the impact of in hours code releases, the programme sought to establish a weekly release slot on a Monday, and in the cases where ad-hoc urgent releases were required, to negotiate and align releases from multiple parties at the same time where possible to further limit necessary down time. Undoubtedly, if an out of hours deployment arrangement could be established, future programmes would benefit from an increased testing throughput, week on week, and potentially could achieve completion of test objectives within shorter elapsed timescales, however if this is not possible it is recommended that programmes should ensure that assumed downtime is built into test execution estimation models.
Observation (Potential Improvement Opportunity)	N/A	N/A

End