Developing a Process for Gathering and Managing Multi-site User Feedback for a Data Collection Project in a Rare Disease Setting. Using Design Science to Explicate a Knowledge Contribution

Sinéad Impey, Jonathan Turner, Frances Gibbons, Anthony Bolger, Gaye Stephens, Lucy Hederman, Ciara O'Meara, Ferran De La Varga, John Kommala, Matthew Nicholson, Daniel Farrell, Emmet Morrin, Miriam Galvin
ADAPT Centre
Trinity College Dublin
Dublin, Ireland
e-mail: simpey@tcd.ie, jonathan.turner@tudublin.ie, frances.gibbons@adaptcentre.ie, anthony.bolger@adaptcentre.ie, gaye.stephens@tcd.ie, hederman@tcd.ie, ciara.omeara@adaptcentre.ie, ferran.delavarga@adaptcentre.ie, john.kommala@adaptcentre.ie, matthew.nicholson@adaptcentre.ie, daniel.farrell@adaptcentre.ie, morrinem@tcd.ie, galvinmi@tcd.ie

Mark Heverin, Éanna Mac Domhnaill, Robert McFarlane, Dara Meldrum, Deirdre Murray, Orla Hardiman
Academic Unit of Neurology
Trinity College Dublin
Dublin, Ireland
e-mail: mark.heverin@tcd.ie, amacdomh@tcd.ie, mcfarlro@tcd.ie, meldrumd@tcd.ie, dmurray1@tcd.ie, hardimao@tcd.ie

Abstract*— Large datasets are required to understand disease progression, investigate treatment options and discover potential cures in rare neurological conditions such as Amyotrophic Lateral Sclerosis. Generating large datasets for such rare conditions requires the participation of multiple specialist clinical sites. The Precision ALS project is a partnership between multiple specialist Amyotrophic Lateral Sclerosis clinical sites (n=10) and industry partners across Europe that seeks to collect and analyse multi-modal data collected from participants with the disease. The project is managed by an expert group comprised of clinical, technical, business, legal and interdisciplinary specialists. Adopting a design science approach, the project created an artefact – a data collection tool. This tool is used by data collectors at each specialist clinical site to capture a range of patient information, including biological and socio-economic data. Applying an iterative approach, the initial user requirements were based on extensive collaborative projects undertaken by clinical sites, which formed the basis of a bespoke worksheet. Additional modifications were introduced through project group discussions and engaging members from one of the clinical specialist sites. The lessons learned from this initial work were formulated into a knowledge contribution in the form of a process for integrating pre-existing paper-based data collection processes, additional requirement gathering and managing user feedback. The process is a way to identify and manage the early digital representation of data gathering paper worksheets, along with the volume and variety of unstructured feedback generated during a diverse range of multi-site user engagements. A central component of the process is three themes that were adapted from the initial feedback and presented as three pillars: Digital Worksheet; Usability; and Process/actors. These three pillars were embedded in a four-step information pipeline (capture, record, review and target), which provided a structure to ensure that the management of this feedback was transparent and auditable from source to decision. This paper describes the knowledge contribution of the user requirements gathering process explicated through a design science approach targeted towards the data collection tool resulting from this process.

Keywords- amyotrophic lateral sclerosis; design science; knowledge contributions; requirements gathering; data collection.

I. INTRODUCTION

This paper expands on work presented at HealthInfo23 on the development of a data collection tool (DCT) for Amyotrophic Lateral Sclerosis (ALS) [1]. ALS is an incurable progressive neurodegenerative disease responsible for up to 10,000 deaths per year in Europe; it is the most common form of the motor neurone diseases [2][3]. ALS is a complex genetic disease. Approximately 15% of patients have a family history of ALS or frontotemporal dementia. Of these, approximately 70% are associated with known genetic variants, the commonest being mutations in C9orf72, SOD1, FUS and TDP 43 [4][5]. Of the remaining 85%, the pathogenic mechanisms are poorly understood, with at least 30 at-risk genes associated with ALS [6]. To further elucidate that disease heterogeneity, and to identify subgroups of patients with shared clinical and pathogenic features that would be suited to targeted drug therapies,

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

collaboration between clinicians and data scientists is required in the collection, curation and analysis, including by machine learning methods [7]. However, generating such large datasets for a rare disease like ALS is challenging due to the low numbers of affected individuals. In response to this challenge, the Precision ALS (P-ALS) project [7][8] was initiated as a partnership between ten clinical sites in Europe, all part of the TRICALS [9] network of 48 specialised ALS sites, interested industry partners, and technical and clinical researchers.

A complexity of combining data from many sites is in ensuring consistency in data collection. A solution chosen for the data collection strand of the P-ALS project is to use a common data collection tool (DCT) at each site. The tool contains a common set of questions across all collection sites and, where appropriate, a defined list of valid responses. These agreed "core data elements" were developed in paper format at TRICALS sites in the context of extensive previous collaborations over a 15-year period. These were provided to the ADAPT Centre [10] as a structured paper 'worksheet' to be used as the basis for the development of the DCT. The research adopted a lightweight Agile approach [11].

This paper describes research undertaken as part of the initial development stages of the DCT in the P-ALS project. The purpose of this research was to explicate the processes required to refine and stress test the suitability of the worksheet, to evaluate usability of the DCT, and to support the TRICALS sites collect data.

Following this introduction, the paper is organised as follows. Section II presents background to the research. Section III describes the methodology – elaborated action design research (eADR). Section IV describes application of eADR activities in this work. Section V presents some of the feedback generated from user engagements. Section VI discusses our conclusions and plans for future work. The acknowledgment and references close the article.

II. BACKGROUND

Wicked problems are complex, multifaceted issues that lack clear definitions and have a high degree of ambiguity [12]. To address the wicked problem described in this paper, an elaborated action design research (eADR) approach [13] was adopted. This provided a structured approach to conducting research in information systems and related fields that not only produced a design artefact but also design knowledge. A core component of eADR is the combination of action research and design science [14]. This combination results in an ensemble of knowledge contributions.

eADR is a four-stage method – diagnosis, design, implementation, and evolution – that emerged from design science roots [14][15] and is illustrated in Figure 1. For this paper, only the first stage of eADR, diagnosis, is described.

According to Mullarkey and Hevner [14], diagnosis is problem-centered. Each stage of the eADR can take several cycles to complete, depending on research requirements. Each eADR stage comprises five activities: Problem Formulation/Action Planning (P); Artefact Creation (A); Evaluation (E); Reflection (R); and Formalisation of Learning (L). For each stage, an artefact is created to address the problem formulated. Artefact creation is an important part of design science research [14].

The knowledge contributions of this work are a DCT artefact and a 'Three Pillar Information Pipeline' that embeds a process for gathering and managing feedback from multiple sources that could potentially be targeted towards user requirements. This paper reports on the development and evaluation of the information pipeline explicated through a design science approach and targeted towards the data collection tool resulting from this process.

III. METHODOLOGY

The P-ALS principal investigator (PI), the leading clinical academic at the National Specialist ALS Centre, is based in Ireland. The DCT project sub-group comprised a clinical subject expert group from Beaumont National Centre & Academic Neurology Trinity College Dublin and a technical development group from Trinity College Dublin. Some members were jointly members of both groups; to differentiate these members, they will be referred to as the 'interdisciplinary' group in this paper. In addition to the project group, and the individual sites, there was a data collectors' group. Data collectors are responsible for using the DCT at each partner site to capture participant information. There is at least one designated data collector at each of the ten sites. The legal and financial expertise necessary to ensure the success of the project was provided by a separate group.

Development of the DCT was in two concurrent streams. The first stream was concerned with gathering technical requirements through review of the previously agreed TRICALS "master worksheet" along with additional discussions and technical workshops with the project group and one partner site. Acting as a test site, it was chosen as it was linked to the project PI and was geographically convenient for the majority of the group. These requirements were subsequently reviewed and refined at the other sites (n=9) during the project. Requirements refinement focused on the needs of clinicians, data collectors and analysts. These requirements informed the iterative development of the DCT by the technical team.

The second stream was concerned with the refinement of the previously generated worksheet to ensure data collected was accurate and capable of being collated into a single dataset. It was important to ensure data collected was capable of being collated into a single dataset, to ensure future cross-site analysis of collected data.

A number of workshops occurred throughout the initial stage of the project to gather additional requirements for both the DCT and the worksheet. These workshops were facilitated by and involved P-ALS team members from the ten data collection sites. Feedback was recorded and reviewed after each workshop.

An early challenge that emerged during the research was the volume of unstructured information that had been implicit in the originally generated TRICALS paper-based collection processes. This was reflected in the structural heterogeneity across sites involved in the project (n=10) along with the variety of stakeholders across a range of

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE I. APPLICATION OF EADR TO THIS RESEARCH

Activity	Performed in this research
Problem formulation/ action planning (P)	Managing the potential high volume and variety of feedback captured from a range of stakeholders including ten different geographical sites that could translate into user requirements so that nothing is overlooked could be challenging.
Artefact creation (A)	An information pipeline to manage feedback so that it is auditable and ensures feedback reaches its target.
Evaluation (E)	The pipeline was developed based on group discussions and engagement with one site. To evaluate its utility, it was used as the additional sites.
Formalisation of learning (L)	Lessons learned were formulised into an illustration of the evaluated pipeline.
Reflection (R)	Reflection was continuous throughout the process during project group discussions.

interests including design/development of the DCT and analysis of data captured. The challenge of gathering additional user requirements from ten individual sites, and with a range of stakeholders, displays the characteristics of a 'wicked' problem: it is 'ill-formulated', information is 'confusing', with many stakeholders often holding conflicting perspectives [14][15].

Design of a bespoke digital data collection tool, based on the TRICALS worksheet, was carried out in-house in the ADAPT Centre, with the development team based at Trinity College Dublin. This team met regularly with clinicians and data analysts to ensure that the developed tool met their needs. A tablet-based application approach was chosen to ensure portability and to enable operation without a working internet connection. This enabled data collection to be performed using the tablet at locations without wi-fi internet connection, which could include some participants' homes. The data collection form structure implemented in the tablet follows the worksheet structure and content developed by TRICALS sites and is configured using a metadata-driven approach, where the user interface is described by metadata rather than hard-coded, allowing easy updates without the need to modify the application code itself. Development followed a lightweight Agile [11] approach with regular prototype releases to project stakeholders. The application is deployed via a mobile device management solution to minimise security and device management concerns. Using Android with Mobile Device Management software provides the mechanism to distribute private apps and client certificates and allows for restricted and secure access to the server. Android provides a more open and accessible development platform than Apple and iOS, which does not provide a distribution mechanism for the small scale required.

The four stages of eADR as applied to this research are shown in Figure 1. Table I describes each step of an eADR stage and describes how each step was carried out in the Diagnosis stage.

IV. APPLICATION OF EADR ACTIVITIES

For each stage of the eADR five activities are performed. This section discusses the activities performed during this stage (Diagnosis) and the outcome.

A. Problem Formulation (P)

The first activity surrounded problem formulation/action planning (P). Problems are formulated by reviewing the learning from previous stages [14]. As Diagnosis was the first stage performed in this research, the problem was developed based on early group discussions with the P-ALS group members and following initial project workshops. From this work, it was observed that due to the high number of stakeholders (including ten participating sites from different European countries), a potentially large volume of information could be generated and transformed into user requirements.

B. Artefact Creation (A)

Artefacts are created to address the problem formulated. Artefact creation is an important part of design science research [14]. Throughout the creation process, the abstract nature of the artefact created will develop. To manage the feedback generated during workshops and discussions, an artefact in the form of an information pipeline was proposed. The purpose of this artefact was to ensure that all feedback was recorded and reviewed, and to facilitate the efficient flow of information from the source to an appropriate target

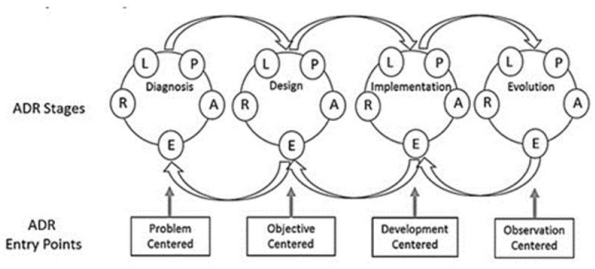

Figure 1. The four stages of elaborated action design research [14].

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

so that a decision could be made. To do this, a proposed route from information identification to categorisation to reaching an appropriate target was constructed by the project group. To categorise the information in the artefact, outputs from the early work completed (paper worksheets, workshops and group discussions) and feedback from one of the ten sites was analysed and three broad themes emerged – feedback relating to the original worksheet, to the usability of the technology, and to the data collection process including actors involved in those processes. These themes were reviewed by the project group and refined into 'Three Pillars': Digital Worksheet, Usability, and Process/actors. Each pillar represents a broad area that was explored to identify user requirements.

Pillar 1, "Digital Worksheet", captured feedback related to the early digital format of the TRICALS master worksheet. These could be technical or clinical. Pillar 2, "Usability", focused on user experience requirements. Pillar 3, "Process/actors", investigated requirements relating to the flow of data into the collection tool from the various data sources that contribute to the final dataset, and the actors required to perform these tasks.

In the pipeline, all feedback received was given a unique reference code and assigned to a pillar, and the nature, provenance and outcome of the submission was recorded. This research uses the term 'Four-step Information Pipeline' to describe the pipeline in its entirety.

C. Evaluation (E)

The eADR process has roots in action design research (ADR). A central component of ADR is to build, implement and evaluate a relevant artefact [15]. Building and implementing in eADR are presented as individual stages (design and implementation, respectively), whereas evaluation is the activity of all stages. The purpose of the evaluation is to ensure artefacts created meet their intended purpose. The 'Four-step Information Pipeline' artefact was created based on project group discussions and workshops but only captured feedback from one of the ten sites. To ensure its suitability across all sites, the pipeline was used to manage feedback from the online and in person site engagements at all sites.

User feedback was identified in three ways. Firstly, prior to the in-person site visits, data collectors at each site were given an evaluation version of the DCT and instructed to enter dummy data. These data were reviewed by the project team for any anomalies. Secondly, lessons learned from previous TRICALS projects in which data were gathered using paper-based format were applied. These previously captured datasets (comprising clinical research from 22,000 patients) is referred to as extant data. Finally, in-person visits were arranged following online engagement meetings with each site. Through the initial online site engagement meetings, variations in the data collected or the interpretation of the question was captured and included in the information pipeline for review by the project group. These discussions also determined how ready each site was to 'go live', so that, if necessary, a suitable remedy or interim measure could be initiated so that the site could move toward collecting data in digital format using the collection tool. A digital document was used to record relevant information using a set of headings that emerged as the project progressed. The final set of headings are shown in the following text:

Pillar 1 (Digital Worksheet) – highlighted any known challenges related to the question set (or worksheet) that needed to be addressed before a site could go live.

Pillar 2 (Usability) – highlighted any known challenges related to the usability of the DCT that needed to be address before a site could go live.

Pillar 3 (Process/Actors) - identified a data collection process for the site and actors required to collect data. Under this heading any challenges to either the process or actors that could be a barrier to go live were identified, addressed or a suitable interim step was proposed.

Ethics - under this heading it was established if ethics was completed and approved, in progress, or not started.

Governance - under this heading it was established if other legal agreements required (such as data sharing agreement, privacy statement) were completed and approved, in progress, or not started.

Collection Tool Provided - information captured under this heading related to whether a site had received a DCT and that all the necessary security measures were in place.

Data collector in situ/contact details - this noted if a data collector was in post at the site. Potential answers were Yes, in post; Post filled awaiting staff member; No. Reasons for 'no' could include that the job had to be advertised or the candidate had not taken up the position. It was important that the site visits included meeting with the data collector as a large part of the site visit surrounded providing training.

A 'traffic-light' system was employed to indicate site readiness. When the outcomes for all headings were positive (i.e., all 'Yes'), the site was deemed 'Green' and an in-person visit to the site was arranged. If readiness was imminent the site was deemed 'Amber' and the project group focused on addressing known challenges. Some sites required extensive clarifications prior to formal ethics approval, so were deemed 'Red' as no data collection could be initiated. The Site Readiness for Go Live document was frequently reviewed to track progress for each site.

Using this Site Readiness for Go-Live document allowed The DCT project sub-group members to access a high-level overview of all sites and quickly identify challenges that individual sites were having.

The site visit following a 'Green' decision followed a protocol that was constructed to ensure a standardised approach across all sites. This protocol reflected the process used at the first site. This protocol included the length of time of the visit and who would be attending. Attending each site visit were two members of the project team, who met with, at minimum, a data collector from the site visited. Each site was encouraged to include as many stakeholders as they wished in these visits, but it was imperative that the data collectors were available. The visit took place over two days to ensure that the site was not over-burdened, being mindful that clinical areas are busy environments.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

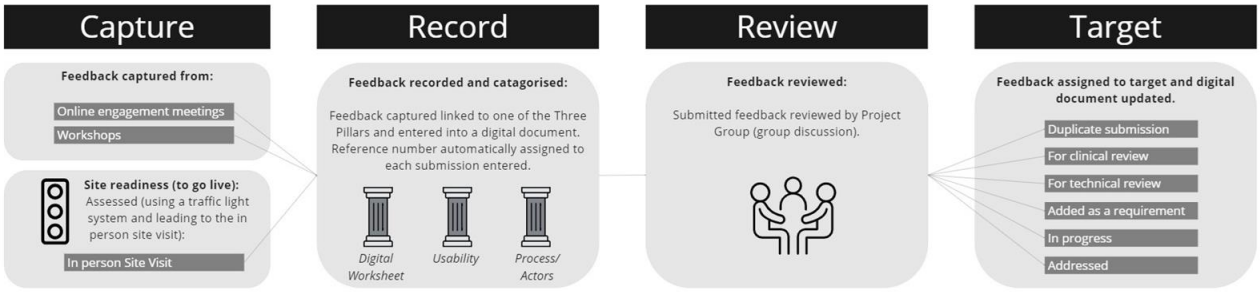

Figure 2. Four-step information pipeline.

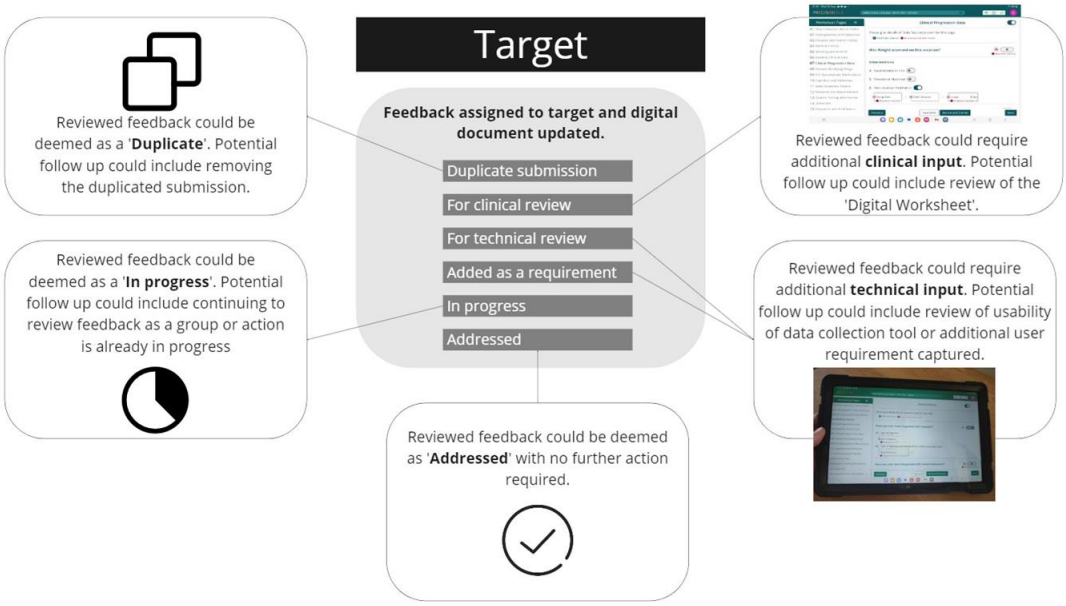

Figure 3. Potential follow ups once user feedback has been assigned to a target

It was envisaged that once each site visit had been completed satisfactorily, the site could 'go live' with participant data collection, so it was important that no barrier remained unaddressed. Feedback captured during the visit was entered into the 'Four-step Information Pipeline'.

A major focus of each site visit was a mock data collection interview. This interview was conducted by one researcher who acted as the patient participant and the data collector, while the other researcher took notes.

The purpose of this interview was fourfold:

1. Understanding: The researcher and data collector ensure all questions are fully understood so that the appropriate data could be collected accurately (related to pillar 1).
2. Training: As a training exercise to ensure the data collector can use the DCT (related to pillar 2) and is comfortable performing an interview with a real patient.
3. Discuss potential data collection process and actors required (related to pillar 3)
4. Generate new user requirements: The interview is observed by a second researcher who takes notes.

During the interview, the data collector at the study site used the DCT to capture the interview responses (from the researcher) and was encouraged to discuss their thoughts on the interpretation of the question, what type of answer they might expect and how the response could best be captured in the tool. For example, one aim of the mock interview was to ensure that the data collector is comfortable with the wording of the questions and understands the rationale and meaning behind them and to provide training if any queries arose. This will help to ensure the data collected is relevant and that the collector understood the rationale for the questions. Along with this training, the purpose of this exercise was to check the contents of the digital worksheet

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

to identify any site-specific issues that could limit the ability of the project to combine data from the ten individual sites.

The three pillars were used to structure the site visit discussion. While the three pillars are presented as separate entities, during the site visits it became apparent that there can be an overlap in discussion. For instance, during the mock interview the discussion might include the type of data (relates to pillar 1), how it could be captured in the data collection tool (related to pillar 2), the potential source of the data and how/who collected it (related to pillar 3).

The notes gathered during the site visits were first reviewed by the research group members involved in the mock data collection interview to ensure they accurately represented what was discussed. At the end of the visit, the team further reviewed all the feedback gathered to ensure all points generated were captured. This validated feedback was then submitted to a digital spreadsheet and further reviewed in a post-site-visit meeting with the project manager, clinical group and technical group for a decision (to include as a user requirement) or for further discussion.

In addition to discussions, during the mock data collection interview, the data collector was observed to see how they engaged with the DCT. This observation included noting when they reverted to use of paper to record information for later data entry.

D. Formalisation of learning (L)

According to Mullarkey and Hevner [14], the lessons learned during the development of an artefact should be formalised into design knowledge. The pipeline proposed has four parts to managing feedback. These are capture, record, review and assign to a target. Ways to capture feedback included workshops, online engagement meetings and in person site visits. The in-person visit adopted a mock data collection interview to review the usability of the tool and questions it contained, along with reviewing the data collection process used at each site. The feedback was recorded in a digital document and automatically assigned a reference number and from this it was reviewed by the project group and assigned to a target (for instance, 'for clinical review' or 'for technical review').

To date five out of ten site visits have been completed with ongoing online site engagement. No major changes were required to the pipeline, apart from refinements to wording. For example, initially the worksheet pillar was referred to as 'dataset/worksheet'. 'Dataset' was removed as it was thought to the refer to any data collected rather than the question set. A further change was the identification of additional targets for the reviewed feedback. These describe who the feedback was directed to. Initially it was envisioned that the targets would align with the pillars. However, from reviewing feedback these categories did not fully capture the type of follow-up the feedback required. Therefore, each reviewed feedback remained assigned to one of the three pillars but was assigned a query type. These are:

Duplicate submission – this referred to any submission that was currently represented in the recorded feedback.

For clinical review – refers to any submission that required additional review by the clinical members of the group before a decision can be made.

For technical review - refers to any submission that required additional review by the technical members of the group before a decision can be made.

Added as a requirement – referred to any submission that was accepted as a user requirement during the review without any further exploratory or information required.

In progress – describes a submission that has been recorded but an answer is not possible during the current review session. This could be for a number of reason, such as, additional expertise or information is required to reach a decision.

Addressed – This denotes any submissions that are addressed and an outcome recorded.

The need for these query types arose as some queries crossed more than one pillar. It is expected that some of these findings could be incorporated into a set of standard operating procedures to give operational guidance to each site. Based on the outcomes of the evaluation, an illustration, Figure 2, of the pipeline was produced to depict the process from start to finish across the four steps. The final artefact created is a pipeline that is proposed as a way to manage feedback generated in a multi-site healthcare technology development project. The aim of the pipeline is to manage user requirements from initial feedback to final decision while providing an auditable record of all decisions. Figure 3 shows the potential follow-up once user feedback has been assigned to a target.

E. Reflection (R)

In the eADR activities, Mullarkey and Hevner [14] propose a reflection activity that is performed in every stage. The purpose of reflection is to identify learnings that can be incorporated into the developing artefact. In this research, reflection took the form of project group discussions that occurred during research. Members of the project group were clinical, technical and interdisciplinary. An overview of how the activities were applied is shown in Table I.

V. FEEDBACK GENERATED

The next section describes the feedback gathered during online engagement meetings and site visits. It is proposed that this information could be useful for system designers developing data collection tools for a rare disease setting. The text is structured using the 3 pillars.

A. Pillar 1: Digital Worksheet

"Digital Worksheet" relates to the set of questions to be used in the collection tool. The questions used in the digital worksheet were based on successfully conducted collaborative clinical research by the TRICALS sites, funded by European grants (e.g., Joint Programme in Neurodegeneration (JPND), FP7 and Horizon) group.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

The aim of this pillar, therefore, was to further review and refine the paper worksheet, to generate a digital version and to identify any potential harmonisation challenges. This work was undertaken to ensure consistency of data collection across the sites and presented an opportunity to ensure the consistency of meaning of the fields contained within the worksheet. Feedback from the site visits raised a number of common queries relating to understanding the digital worksheet and individual questions within the worksheet. These queries included wording issues, for example, whether a question relating to children was intended to capture the number of biological children only, or any and all dependent children including those adopted or from a new partner.

To address this finding, a document with answers to frequently asked questions relating to the digital worksheet was constructed. Once evaluated, the content could be integrated into the DCT. To access this content, data collectors can click on 'information buttons' associated with a question to quickly access information relating to a specific question if required.

This work also evaluated the consistency of meaning of the fields contained within the digital worksheet, to identify the potential range of answers and if standard application programming interfaces (APIs) could be incorporated into the data collection tool.

In addition, the work identified a range of harmonisation challenges.

1) Numerical data challenges: Units, particularly currencies; preference for numerical values being recorded as precise values or banding.

2) Social/cultural challenges: Societal mores may vary in how acceptable it is to reveal personal information, such as income, in a research study.

3) Documentation practices: Assumptions may be made at individual sites that their way of working is documented with sufficient precision, for example, when recording a patient's forced vital capacity (FVC), the maximum amount of air that the patient can exhale, the patient can be standing or sitting for the measurement. Different sites use different positions. The measurement is then recorded as 'FVC sitting', 'FVC lying' or simply 'FVC' (without the patient's position being recorded).

As each harmonisation challenge was identified, it was shared among all data collectors (across all sites) for feedback and to see if it was a site-specific challenge or global.

Feedback gathered during this part of the research included noting variations in how each variable was interpreted by each site, being mindful that this is a multi-national project; that each variable is relevant for data analysis or that there is an analysis impact of the omission of the variable; and what, if any, challenges exist in collating multisite data. Apart from identifying any data variable issues, reviewing the question set also helped ensure data collectors were comfortable with their understanding of the wording of the questions and the rationale behind them. It is envisioned that this will help improve the quality and consistency of the data collected once routine data collection commences.

B. Pillar 2: Usability

The second pillar was concerned with usability of the DCT. Usability of the tool was focused on clear presentation of required data items and ease of data entry. Prior to the site visit, online site engagement meetings, with representatives from each site, were given a complete overview of project and walkthrough the DCT. Users then had the opportunity to explore the tool independently, with members of the development team available to answer any questions and to capture verbal feedback.

To record a baseline measure of usability of the DCT, 13 users from across all sites completed a user survey. The tool used was version 3 of the Post-Study System Usability Questionnaire (PSSUQ) [16]. The PSSUQ is a 16-point questionnaire that is used to measure users' satisfaction with a product. Questions include, for example 'Overall, I am satisfied with how easy it is to use the system' or 'It was simple to use the system', with users giving each statement a score ranging from 1 ('Strongly agree') to 7 ('Strongly disagree'). The PSSUQ produces an overall score together with three subs-scores for system usefulness, information quality and interface quality.

Results of the baseline PSSUQ are shown in Table II. Possible scores range from 1 (best) to 7 (worst). For each of the three sub-scale scores (system usefulness, interface quality and information quality) and for the overall score, the tool performed well, exceeding the mean scores calculated by Sauro and Lewis [17]. It is intended that the tool will be further refined over the period of its development and use, with changes in usability measurable by repeat use of the questionnaire.

An early usability issue noted was that dropdown menus containing a high number of potential answers, particularly medications, interfered with users' experience of the tool as the dropdown covered the majority of the screen.

C. Pillar 3: Processes (including actors)

Technologies are influenced by the environment within which they are situated and vice versa. This is referred to as reciprocal shaping by Sein et al. [15]. Understanding the context is therefore very important. A problem faced by the development team was that the collection tool was to be located in ten sites from different European countries or cities, all working in different healthcare systems. The final pillar allowed the tool to be observed being used at the sites, to gain a better understanding of how data collectors planned to capture participant data using the collection tool and who were the actors needed to produce, review or record these data. Understanding data collection processes and relevant existing workflows allowed the project to generate further user requirements.

In addition to processes and actors, under this pillar, current data collection systems and any other artefacts used at the site to collect data were identified and the impact on data collection reviewed. The purpose of this was to

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

understand how these could fit with the P-ALS DCT to minimise the impact of the data collection process on clinical work at the data collection sites. This pillar ensures that the data collection process required by the tool does not impose inefficient or impractical working practices on the data collection sites.

From this an understanding of how the tool would be used at the sites was developed. For example, initially it was thought that the primary source of patient information would be an interview. However, following the site visits it was noted that common to all sites was a collection process whereby a portion of the information would be transcribed from existing clinical documents (paper and electronic) into the tool. The remainder would be captured in an interview at a later time. This led to a requirement that data entered in one location be available to view when the data collector went to add additional information (from the interview).

The data collection tool contains 15 pages. Each page focuses on a particular division of data to be collected, for example 'Smoking and Alcohol' use, or 'Socio-Economic Details'. Table III lists the 15 pages and describes their contents. Table IV gives an example of the contents of one page in more detail. This allows for some pages to be skipped when not required in a particular data collection encounter, e.g., during a repeat data collection encounter when the focus of the data collector is on fields that may have changed since the last encounter, such as clinical progression or resource use. A change resulting from the feedback was the rearrangement of the pages. This resulted in pages containing questions that would be collected during an encounter to be grouped together. A sample data collection page is shown in Figure 4.

Another change identified was that, as not all questions would be asked in each participant encounter, users required a way to know what had been asked or what needs to be asked (during the current data collection encounter). A requirement around carrying over previous data is currently in development to address this.

Future iterations of the collection tool aim to extend data collection beyond the clinical encounter and patient interview to include data from wearables, carers, and other modalities. While no unified data collection process was found that suited all sites, the arrangement of the worksheet (now question set) was also amended to better match the flow of information. For example, information that is likely to be updated frequently, such as clinical progression, was grouped together, whereas socio-economic information is captured at specific intervals so was presented separately.

Along with the question set and usability review, existing data collection processes at the site were examined through discussions with the data collection. This included understanding the role of data collectors; registration of participants; clinical coding systems used; information systems used; and remote monitoring of participants. The data on participants are personal and sensitive, and each collection site was required to follow their local processes for ethics approval, data protection impact assessment and to sign a data transfer agreement. It was for each site to ensure that they comply with their local data protection laws.

TABLE II. SCORES FROM PSSUQ

Question number	Question text	Average Score
1	Overall, I am satisfied with how easy it is to use the system	2.2
2	It was simple to use this system	2.2
3	I was able to complete the tasks and scenarios quickly using this system	2.4
4	I felt comfortable using this system	2.6
5	It was easy to learn to use this system	2.3
6	I believe I could become productive quickly using this system	2.0
7	The system gave error messages that clearly told me how to fix problems	3.1
8	Whenever I made a mistake using the system, I could recover quickly and easily	2.7
9	The information (such as SOP) provided with this system was clear	2.3
10	It was easy to find the information I needed	2.0
11	The information was effective in helping me complete the tasks and scenarios	2.3
12	The organization of information on the system was clear	1.9
13	The interface was pleasant	2.0
14	I like using the interface of this system	2.2
15	The system has all the functions and capabilities I expect it to have	2.7
16	Overall, I am satisfied with this system	2.0
	Overall PSSUQ score	2.3
	System usefulness	2.3
	Information quality	2.4
	Interface quality	2.3

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE III. SECTIONS IN THE DCT

Section number	Section name	Description
00	Index Capsule	Personal identifying information, such as name, address, date of birth, that should be kept private. Not available during routine use of the data collection tool.
01	Data Collection Admin Fields	Details of the data collection, such as the date of collection, data source (e.g., patient interview or healthcare record), data collector's ID.
02	Demographics and Education	Demographic information of the participant including details of their education history and number of children.
03	Ancestry and Family History	Ethnicity and geographic origin of participant, parents and grandparents; medical history of relatives.
04	Medical History	Including specifically history of diabetes, cholesterol, psychiatric conditions, and medications taken for these
05	Smoking and Alcohol	History of smoking and alcohol consumption
06	Baseline Clinical Data	Details of ALS onset, first symptoms and diagnosis
07	Clinical Progression Data	Data recorded at each clinic visit. Includes weight, whether gastrostomy is in situ, FVC score, use of NIV, ALSFRS-R score and sub-scores.
08	Disease Modifying Drugs	History of taking disease-modifying medication (primarily Riluzole, Edaravone or Tofersen)
09	ALS Symptomatic Medication	History of taking symptomatic medications or therapeutic interventions
10	Cognition and Behaviour	Dates and scores of cognitive screening (primarily ECAS), behavioural screening, dementia status
11	Socio-Economic Details	Including information on household income, employment history and occupation, and state benefits
12	Resource Use	Information on care received, including community-based care, social care, palliative care, counselling services, hospital care, aids and appliances received, and additional costs due to ALS
13	Genetic Testing Information	History of genetic testing, including which genes were tested for, dates of tests and results of tests.
14	Linked IDs	the key unique local IDs that allow the Precision ALS participants information to be linked to all other research information.
15	Endpoints and Vital Status	Patient vital status, date and place of death, dates of permanent mechanical ventilation and tracheostomy if applicable.

TABLE IV. FIELDS WITHIN SECTION 10 OF THE DCT

Question number	Question	Notes	Frequency
01	Was cognitive screening performed?	'Yes' or 'No'	Recurrent
02	Which assessment was used?	'ECAS' or 'Other, please specify'	Recurrent
03	Date of cognitive screening		Recurrent
04	ECAS ALS_SPECIFIC_Total Score	0-100	Recurrent
05	ECAS ALS Specific Classification	'Normal' or 'Abnormal'	Recurrent
06	ECAS Total Score	0-136	Recurrent
07	Cognitive Screening Status	'Normal' or 'Abnormal'	Recurrent
08	Was behavioural screening performed?	'Yes' or 'No'	Recurrent
09	Which assessment was used?	ECAS, BBI, Apathy Scale, FTDQ. Other, please specify	Recurrent
10	Date of behavioural screening		Recurrent
11	Behavioural status	Mild impairment, Moderate impairment, Severe impairment , No Impairment	Recurrent
12	Does the person have dementia?	'Yes' or 'No'	Once-off
13	Source of Diagnostic Dementia Information	Pre-existing Dx before visiting clinic , Neurologists notes, ECAS , Full battery , neuropsychological testing, Other, Please Specify	Once-off
14	Date of diagnosis		Once-off
15	Type of dementia	FTD (Frontotemporal Dementia), Amnestic Dementia, Other, Please Specify	Once-off

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

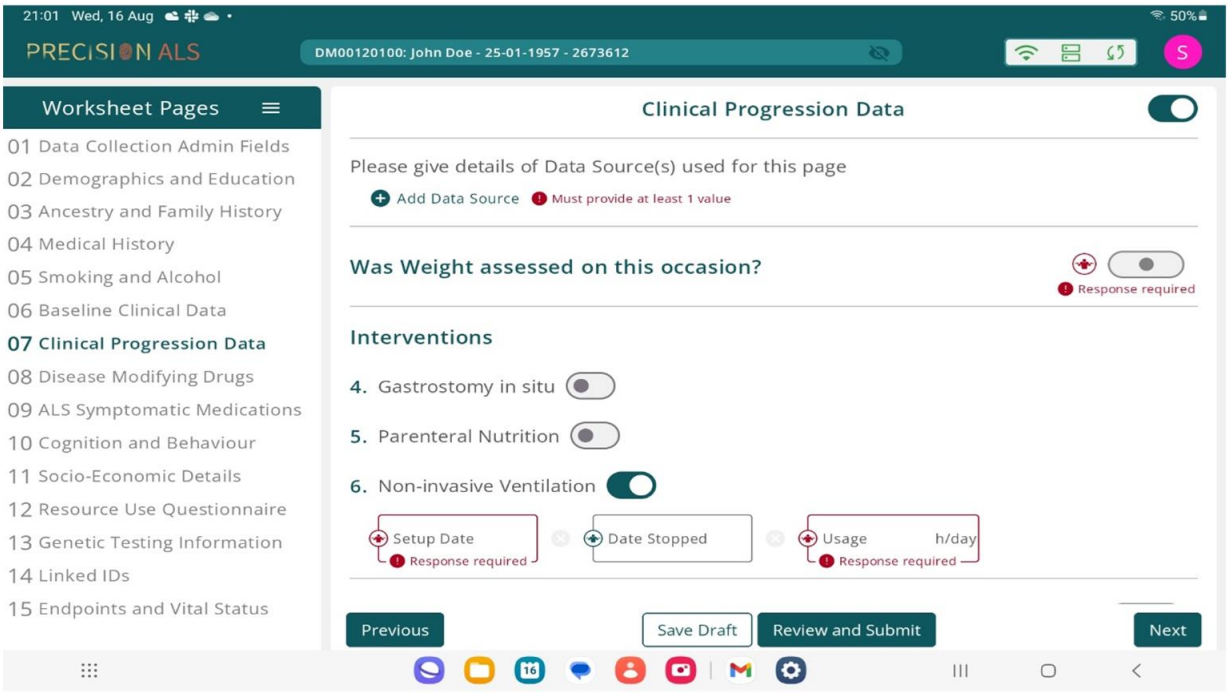

Figure 4. An example page from the data collection tool

VI. CONCLUSIONS AND FUTURE WORK

User requirements for the development of a data collection tool were constructed from the feedback from different groups including ten clinical sites. The work described in this paper addresses the wicked problem of managing the potential high volume and variety of feedback captured from a range of stakeholders including ten different geographical sites that could translate into user requirements so that nothing is overlooked. To address this problem, a knowledge contribution in the form of a process for an auditable information pipeline was developed to manage this feedback so that it reached its target, the DCT. The initial pipeline was developed based on project group discussions and engagement with an initial test site.

The pipeline proposes four parts to managing feedback. These are capture, record, review and assign to a target. Ways to capture feedback included workshops, online engagement meetings and in-person site visits. The in-person visit adopted a mock data collection interview to review the usability of the tool and questions it contained, along with reviewing the data collection process used at each site. The feedback was recorded in a digital document and automatically assigned a reference number and from this it was reviewed by the project group and assigned to a target (for instance, for clinical review or for technical review).

Feedback is categorised into three pillars: digital worksheet, usability, and process/actors. Each pillar represents a broad area that was explored to identify user requirements. Pillar 1, "Digital Worksheet", captured feedback related to the question set. These could be technical or clinical. Pillar 2, "Usability", focused on user experience requirements. Pillar 3, "Process/actors", investigated requirements relating to the flow of data into the collection tool from the various data sources that contribute to the final dataset, and the actors required to perform these tasks. The evaluation process did not lead to identification of further pillars, and we believe that the three-pillar model has proved to be both helpful and robust.

From this work, challenges were identified that could potentially limit the collected data to be combined from many sites into a single dataset. For example, under pillar 1 (digital worksheet) wording and harmonisation were identified. The latter included numerical, social/cultural and documentation challenges. Development and refinement of the information pipeline will continue at the remaining sites.

ACKNOWLEDGMENT

This work is funded by the Science Foundation Ireland grant SP20/SP/8953. OH's work is also funded from grants 16/RC/3948 and 13/RC/2106_P2.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

REFERENCES

[1] S. Impey, J. Turner, F. Gibbons, A. Bolger, G. Stephens, et al. "Dataset, Usability and Process - Developing an Interdisciplinary, Multi-modal Data Collection Tool and Platform for a Rare Disease," HEALTHINFO 2023, The Eighth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing. Valencia, 13th-17th November 2023. ISBN: 978-1-68558-105-3. Available from https://www.thinkmind.org/articles/ healthinfo_2023_1_50_80022.pdf

[2] M. Ryan, M. Heverin, M. A. Doherty, N. Davis, E. M. Corr, et al. "Determining the incidence of familiality in ALS: a study of temporal trends in Ireland from 1994 to 2016," Neurology Genetics. 2018;4:e239.

[3] C. A. Johnston, B. R. Stanton, M. R. Turner, R. Gray, A. H. Blunt, et al. "Amyotrophic lateral sclerosis in an urban setting: a population-based study of inner city London". Journal of Neurology. 2006;253:1642–3.

[4] S.A. Goutman, O. Hardiman, A. Al-Chalabi, A. Chió, M. G. Savelieff, et al. "Recent advances in the diagnosis and prognosis of amyotrophic lateral sclerosis". The Lancet. Neurology. 21 (5): 480–493, May 2022, doi:10.1016/S1474-4422(21)00465-8. PMC 9513753. PMID 35334233

[5] MedlinePlus. *Amyotrophic lateral sclerosis*. [Online]. Available from: https://medlineplus.gov/genetics/condition/amyotrophic-lateral-sclerosis [Accessed 2024.05.31]

[6] P. Masrori and P. Van Damme. "Amyotrophic lateral sclerosis: a clinical review," European Journal of Neurology, 2020 Oct;27(10):1918-1929. doi: 10.1111/ene.14393. Epub 2020 Jul 7. PMID: 32526057; PMCID: PMC7540334.

[7] R. McFarlane, M. Galvin, M. Heverin, É. Mac Domhnaill, D. Murray, et al, "PRECISION ALS—an integrated pan European patient data platform for ALS," Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2023. 24:5-6, 389-393, DOI: 10.1080/21678421.2023.2215838

[8] Precision ALS. *Precision ALS.* [Online]. Available from: https://www.precisionals.ie [Accessed 2024.05.31]

[9] TRICALS. *TRICALS: The Highway Towards A Cure [Online]*. Available from https://www.tricals.org [Accessed 2024.05.31]

[10] Adapt Research Centre. *ADAPT: The Global Centre of Excellence for Digital Content and Media Innovation.* [Online]. Available from https://www.adaptcentre.ie/ [Accessed 2024.05.31]

[11] Agile Alliance. *What is Agile?* [Online]. Available from: https://www.agilealliance.org/agile101[Accessed 2024.05.31]

[12] C. Churchman. "Wicked problems." Management Science, 1967:14(4), B141–142

[13] H.W.J. Rittel and M.M. Webber. "Dilemmas in a General Theory of Planning." Policy Sciences, 1973:4 (2): 155-169, doi:10.1007/bf01405730

[14] M. Mullarkey, A. R. Hevner. "An elaborated action design research process model." European Journal of Information Systems, 2019:28(1):6-20. doi: 10.1080/0960085X.2018.1451811

[15] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. "Action Design Research." MIS Quarterly 35, no. 1 (2011): 37–56. https://doi.org/10.2307/23043488.

[16] J.R. Lewis. "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use". International Journal of Human–Computer Interaction. 1995. 7:1, 57-78, DOI: 10.1080/10447319509526110

[17] J. Sauro and J.R. Lewis, Quantifying the user experience: Practical statistics for user research, 2nd ed. Cambridge, MA: Morgan-Kaufmann, 2016.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org