Dataset, Usability and Process - Developing an Interdisciplinary, Multi-modal Data Collection Tool and Platform for a Rare Disease

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: sinead.impey@adaptcentre.ie, jonathan.turner@adaptcentre.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, emmet.morrin@adaptcentre.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, macfarlro@tcd.ie, meldrumd@tcd.ie, dmurray1@tcd.ie, hardimao@tcd.ie

***Abstract*—Large data sets are required to understand disease progression, investigate treatment options and discover potential cures in rare neurological conditions such as Amyotrophic Lateral Sclerosis (ALS). Generating large data sets for such rare neurological conditions requires the participation of multiple clinical sites. The Precision ALS project is a partnership between multiple clinical sites and industry partners across Europe that seeks to collect and analyse multi-modal data collected from participants with the disease. In this paper, we describe the development of a data collection tool that allows for the collection and integration of data collected at these multiple sites. We focus particularly on the requirements gathering method, which was divided into three pillars: Dataset, Usability and Process. The data collection tool runs on an Android tablet and is now in use enabling collection of data from across Europe for the Precision ALS project.**

Keywords-amyotrophic lateral sclerosis; motor neurone disease; agile development process; requirements gathering; data integration.

I. INTRODUCTION

Amyotrophic Lateral Sclerosis (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 neuron diseases [1][2]. Most (> 90%) cases of ALS have no known cause [3][4]; the remaining cases have a genetic cause [5][6]. Collaboration between clinicians and data scientists is required in the collection, curation and analysis, including by machine learning methods, of large multi-modal data set to understand the disease and its heterogeneity [7]. However, generating such large datasets for a rare disease like ALS can be 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 nine clinical sites, interested industry partners and technical researchers. The project aims to develop a data collection tool and an interdisciplinary data platform to gather, share and analyse multi-modal data. Following this introduction, the rest of this paper is organized as follows. Section II describes the methods used. Section III describes the early results from this work. Section IV discusses our conclusions and plans for future development. The acknowledgement and references close the article.

II. METHOD

Development of the data collection tool was in two streams: requirements gathering and refinement, and application development and refinement. Requirements gathering focused on the needs of clinicians, data collectors and analysts; application development was the domain of the technical team. Refinement of requirements and of the developed application was dependent on effective two-way communication between the clinical and technical groups. To ensure effective communication during requirements gathering, visits from members of the clinical and development team to data collection sites took place, to meet with clinicians, researchers and data collectors at each site.

ISBN: 978-1-68558-105-3

A. Requirements gathering

To ensure that the needs of the project stakeholders were met, a requirements process was developed which divided the requirements gathering for the data collection tool into three pillars as shown in Fig. 1. Pillar 1, “Dataset”, collected content requirements for the data collection tool, i.e., the fields to be collected. Pillar 2, “Usability”, focused on requirements for the workflow through the data collection tool to maximise data collection efficiency and accuracy. Pillar 3, “process”, investigated requirements relating to the flow of data into the collection tool from the various data sources that contribute to the final dataset.

1) Pillar 1: dataset

In Pillar 1 of the requirements gathering, fields to be used in the data collection were identified. Expert sources used for identification of these fields included clinicians specialising in ALS, data analysts, and individual partner sites who had been collecting data on ALS patients for some years. Following identification of fields needed, options within those fields were identified for inclusion on a paper worksheet. This worksheet was used as a prototype for the tool. For example, participants were to be questioned about their history of taking ALS symptomatic medications; a list of possible medications was included in the data collection tool as options from which the data collector could select appropriately.

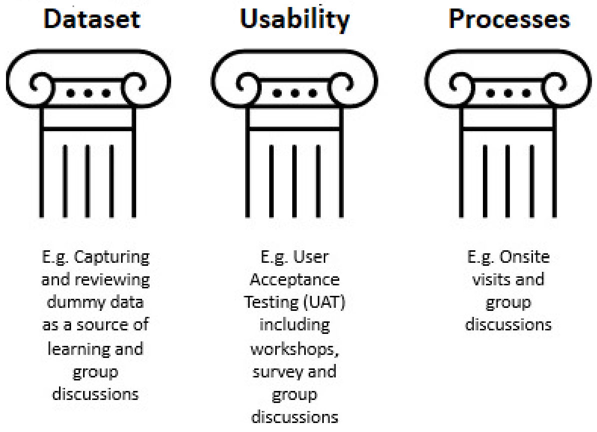

Dataset
Usability
Processes
E.g. Capturing and reviewing dummy data as a source of learning and group discussions
E.g. User Acceptance Testing (UAT) including workshops, survey and group discussions
E.g. Onsite visits and group discussions

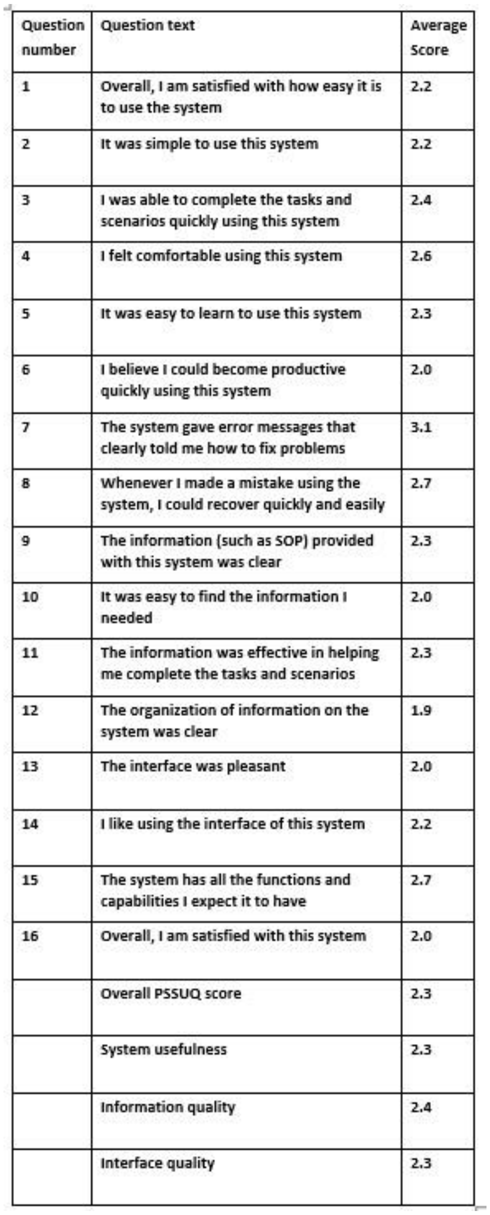
Figure 1. The three pillars of requirements gathering.

2) Pillar 2: usability

The data collection tool is intended to be used in a variety of settings, including face-to-face with participants and/or their carer(s), in telephone interviews, or for data entry from existing records. In each scenario, the tool must be usable efficiently and accurately by the data collector. With no existing data collection process in place, usability of the tool was focused on clear presentation of data items and ease of data entry.

Representatives from each site, who were planning to use the tool for data collection, were given a complete walkthrough of the collection tool, at a single meeting involving all data collection sites. 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. Following this familiarization process, a user survey was completed by each site member, 13 users in total, using version 3 of the Post-Study System Usability Questionnaire (PSSUQ) [9]. It is intended that the tool will be further refined over the period of its development and use, with changes in usability measurabl-e by repeat use of the questionnaire. Results of the baseline PSSUQ are shown in Table 1. Possible scores range from 1 (best) to 7 (worst).

TABLE I. RESULTS OF THE BASELINE 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

3) *Pillar 3: process*

The process pillar is concerned with understanding the current processes and associated actors, both human and technological involved in data collection. The purpose of this is to gain an understanding of how the proposed technology may impact the current processes or where changes are likely. 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. Existing data collection processes at data collection sites were examined for roles of actors involved in the collection process, in particular: the role of data collectors; registration of participants; clinical coding systems used; information systems used; and remote monitoring of participants. The collected data is 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. Oversight of these processes was coordinated at the top level of the project via a team that included data protection experts.

The three pillars, although discussed separately, in practice were reviewed together. To do this, two project researchers visited each site. This allowed the researchers to discuss and compare findings. 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. Initial discussion focused on the data collection process proposed by the site which included location of such data as medical records, current studies or discipline-specific databases.

To understand the variables contained in the worksheet and the interface design, the data collector was asked to take part in a mock interview. For this interview, one researcher acted as the participant and the other researcher took notes. The data collector at the study site used the collection tool to capture participant responses 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.

Information gathered during the site visits was discussed between the researchers post interview and recorded. These were brought to the wider project group for a decision or further discussion. It is expected that some of these findings could be incorporated into standard operating procedures.

.

B. *Application development*

Development of the data collection tool was carried out in-house in the ADAPT Centre [10], 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. It was decided that dedicated devices would be provided to the sites for data collection, which would be managed remotely This 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 application was developed by the in-house team using the Kotlin programming language [11] in Android Studio [12]. The data collection form structure is configured using a metadata driven approach, allowing easy updates without the need to modify the application code itself. Development followed a lightweight Agile [13] approach with regular prototype releases to project stakeholders.

III. Results

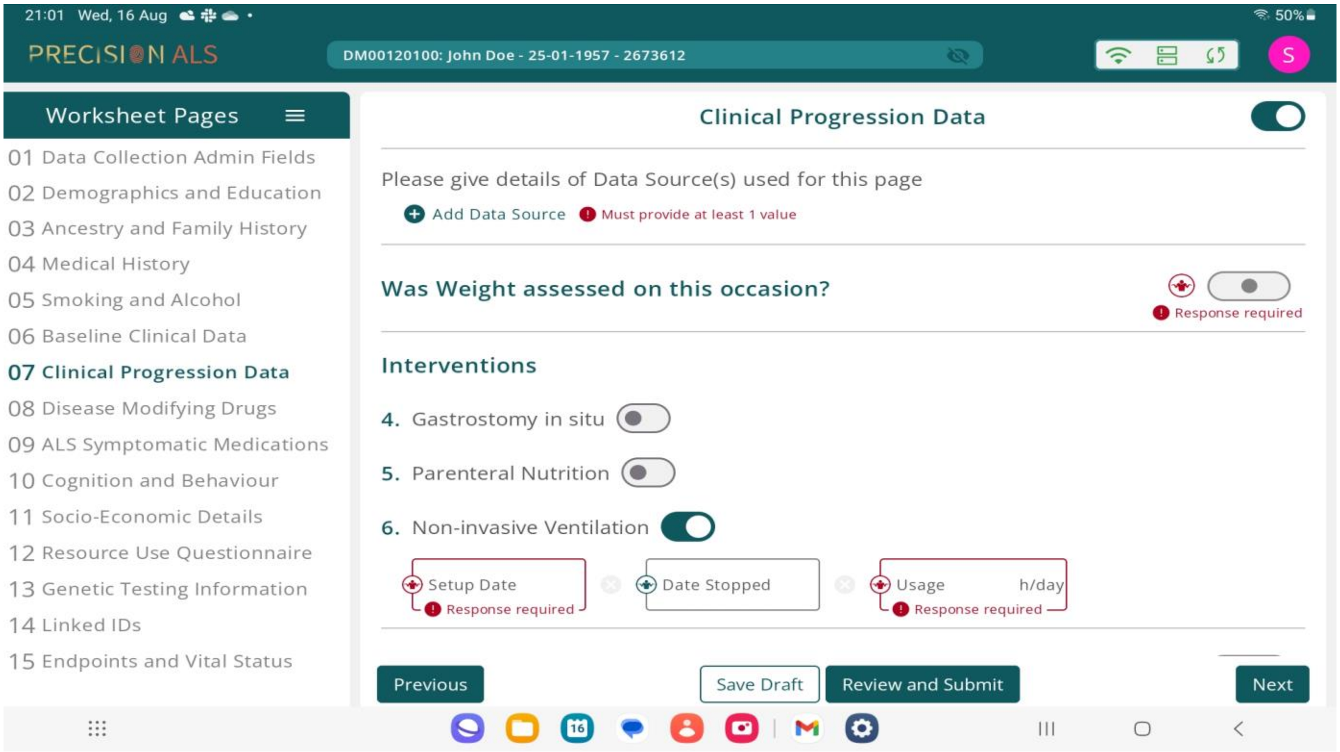
Use of the three pillars for requirements gathering was successful. From Pillar 1, dataset, an agreed data set was identified. Pillar 2, usability, identified functions to improve engagement; results of the user survey are shown in Table 1. Pillar 3, process, identified the people and systems currently used at data collection sites and how these actors could be replicated in the collection tool. A sample data collection page is shown in Fig. 2. This also lists, on the left, the full set of pages available in the collection tool. The data collection tool contains 15 pages, each focusing on a particular division of data to be collected, for example ‘Smoking and Alcohol’ use, or ‘Socio-Economic Details’. This allows for some pages to be skipped when not appropriate, 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.

The developed version of the data collection tool was first used in a clinical setting in August 2023, at one site. Further sites started data collection in September 2023.

IV. Conclusions and Future Work

A set of requirements for the development of a data collection tool can be constructed from the requirements of different groups of interested parties (clinicians, analysts, industry partners), with the success of the developed tool dependent on regular communication between these parties and the technical development team. The tool can incorporate requirements from existing data collection practices at individual partner sites and new requirements elicited as part of the requirements gathering process.

Future work has three main strands. Development and refinement of the data collection tool will continue, with the knowledge and feedback gained from its use in the field informing this. Following the successful collection of data from participants in the P-ALS research project, development of the data platform infrastructure will commence using a similar development process, i.e., requirements gathering from clinicians, researchers, data


21:01 Wed, 16 Aug
50%
PRECISION ALS
S
Worksheet Pages
01 Data Collection Admin Fields
02 Demographics and Education
03 Ancestry and Family History
04 Medical History
05 Smoking and Alcohol
06 Baseline Clinical Data
07 Clinical Progression Data
08 Disease Modifying Drugs
09 ALS Symptomatic Medications
10 Cognition and Behaviour
11 Socio-Economic Details
12 Resource Use Questionnaire
13 Genetic Testing Information
14 Linked IDs
15 Endpoints and Vital Status
Add Data Source
Previous
Save Draft
Review and Submit
Next
16

DM00120100: John Doe - 25-01-1957 - 2673612

Clinical Progression Data

Please give details of Data Source(s) used for this page

Must provide at least 1 value

Was Weight assessed on this occasion?

Response required

Interventions

4. Gastrostomy in situ

5. Parenteral Nutrition

6. Non-invasive Ventilation

Setup Date — Response required

Date Stopped

Usage h/day — Response required

Figure 2. An example page from the data collection tool.

development team of the data platform structure required.

Data platform infrastructure development will include ensuring that data from multiple modalities, e.g., medical imaging, wearable devices, can be imported into the platform and made available for analysis. Development of the platform will be described in future -publications. Finally, once data of sufficient quantity and quality is available, analyses can be performed on the data. Work is underway to ensure that academic analysts and industry partners have the opportunity at an early stage to describe questions that they may wish to answer, to ensure that the data platform enables the ability to answer these questions. These questions include the costs of the disease, both to society and to families; if incidence of ALS is higher for particular occupations; and whether time to disease progression events can be predicted from early information on an individual.

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.

REFERENCES

[1] 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.

[2] 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.

[3] National Institute of Neurological Disorders and Stroke. *Amyotrophic Lateral Sclerosis (ALS) Fact Sheet*. [Online]. Available from: www.ninds.nih.gov [Accessed 2023.08.21]

[4] ALS Association. *Understanding ALS*. [Online]. Available from: https://www.als.org/understanding-als [Accessed 2023.08.21]

[5] 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

[6] MedlinePlus. *Amyotrophic lateral sclerosis*. [Online]. Available from: https://medlineplus.gov/genetics/condition/amyotrophic-lateral-sclerosis [Accessed 2023.08.21]

[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: www.precisionals.ie [Accessed 2023.08.21]

[9] 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

[10] Adapt Research Centre. *ADAPT: The Global Centre of Excellence for Digital Content and Media Innovation*. [Online]. Available from https://www.adaptcentre.ie/ [Accessed 2023.08.31]

[11] JetBrains. *Kotlin*. [Online]. Available from: https://kotlinlang.org/ [Accessed 2023.08.21]

[12] Google LLC. *Android Studio*. [Online]. Available from: https://developer.android.com/studio [Accessed 2023.08.21]

[13] Agile Alliance. *What is Agile?* [Online]. Available from: https://www.agilealliance.org/agile101 [Accessed 2023.08.21]