1a How will new data be collected or produced and/or how will existing data be re-used?

Only new data will be collected in this study (Data re-use and Provance are therefore NA). New data is required for this study as it aims to assess effectiveness and implementation of an initiative that has not been delivered to scale to date.

Methodologies or software used for data collection and analysis:

Both qualitative and quantitative data will be collected.

Data will be collected from stakeholders who include Participants, their Significant Others, Volunteer Coordinators at sites, Community Partners (Clubs & Local Sports Partnership staff) and Football Cooperative (organisational level) and Funders.

Qualitative data: This will be collected via interview and/or focus groups both in-person via Digital Dictaphone and online via the Zoom platform. Nvivo will be used to support analysis.

Quantitative data: Outcomes for all stakeholder groups, with the exception of Participants, will be assessed via Qualtrics survey.

With respect to the Participant stakeholder group, two groups will be established - Intervention Group (IG) and Control Group (CG) - to assess outcomes. The IG will have four categories; C1 (n~700, questionnaires only), C2 (n=200, questionnaires + on site measures), C3 (n=100, questionnaires, onsite measures + wearables), C4 (n=50, questionnaires, onsite measures, wearables + GPS vests). The CG will have C1 (n=200), C2 (n=100) and C3 (n=50) only.

This data will be collected at Baseline (B), 3 Months (3M) and 12 Months (12M).

For the purpose of efficiently recording and reporting the volume of data we plan to collect from the IG and CG, we have purchased a subscription for use of the Kitman Labs data portal. This will be a bespoke portal designed for the study and only members of the research team will have access to it. Data for C2-4 participants will be recorded via this App under a unique identifier (see below).

Only those participants in the C2-4 (IG & CG) will download the Kitman Labs App on their phone and the username they will be assigned for their App will be their UID. As data is recorded for each of these participants by members of the research team on site, the portal will sync with the App on their phone and they will have access to their own data in real time. Via this portal the research team can also send notifications for data collection that is not collected in-person on site.

Finally, the portal will also be set up to access data from a wrist wearable e.g. Garmin watch that a subset of participants will receive (C3). Participant can access data in real time from the wearable via the portal. Participants in C4 will wear Stat Sports GPS vests when playing, and the portal will also be set up to access data from this device.

For participants who are unable/unwilling to install the Kitman Labs data portal App, the research team will collect data manually on site via a password protected excel spreadsheet at each time point of data collection. Spreadsheet files will be sent to the PI post data collection and will be filed accordingly on OneDrive. Any copy of the file will be deleted from email.

Participants wishing to remove their data from the Kitman Labs portal can contact the research team who will then work with Kitman Labs to erase the data. They may continue to participate in the study by having their data collected manually via excel spreadsheet at per above.

Specifically, the data to be collected from the Participant group is:

- A self-reported questionnaire (Qualtrics), will be administered to C1 participants in the IG via the FC gatekeeper. The link to the questionnaire will be sent to these participants via email and the FC App. For C2-3 IG participants and CG participants, the Qualtrics questionnaire will be administered via the data collection App i.e. the research team can send participants a link to the questionnaire.
- On two weekdays and 1 weekend day at data collection time periods, participants in C2-4 will receive a notification through the data collection App asking them a) how many fruit and veg portions they ate that day and b) how much alcohol they consumed that day.
- Data will be collected on site for C2-4 participants to assess physical health outcomes at all time points. Specially, these are:
  - Blood pressure (BP, mmHg) will be measured using an automated sphygmomanometer (Heine Ltd., Munich, Germany).
  - Waist circumference (cm) will be measured over light clothing using a standard tape measure.
  - Weight (kg) will be measured using a Seca 813 electronic weighting scales (Seca Corporation, Hamburg, Germany) with participants wearing light clothing, no shoes and with empty pockets.
  - Height (m) will be measured using a portable stadiometer (Seca 220; Seca Corporation, Hamburg, Germany) and measurements will be taken to the nearest centimetre.
  - Body composition will be assessed via a total body scan. These will be conducted using a multifrequency bioelectrical impedance analyser (InBody 720; Biospace Co. Ltd, Seoul, Korea) at frequencies of 1, 5, 50, 250, 500, and 1000 kHz.
  - Aerobic fitness (VO2max) will be measured by asking participants to complete a distance of 1-mile at their own pace and recording their time to completion OR by the YoYo Intermittent Recovery test (YYIR).
  - Lower body strength will be measured via counter movement jump (CMJ) test using a Chronojump Plate (Chronojump Inc. Barcelona, Spain), which measures flight time taken as the duration between take-off and landing.
  - Cholesterol (mmol) will be measured by analysis of a finger prick blood sample using a Accutrend Plus.
- A subset of participants (C3) in both the IG and the CG will be given a wearable e.g. Garmin watch, at B and asked to wear it for the 12M period. At data collection timepoints, data related to physical activity, sleep and stress will be synced with the data collection portal via an API.
- A subset of participants in the IG (C4) will wear a STAT Sport GPS vest and heart rate (HR) monitor during 3 games at each data collection time point to assess game play demands and playing efficiency over time. Again, this data will be synced with the data collection portal via an API.

Data stored on the Kitman Lab portal will be exported to excel after each data collection time point. At the end of the data collection period when the excel files have been checked for accuracy, data will be destroyed on the Kitman Lab portal.

Data stored on Qualtrics will be exported to excel after each data collection time point. At the end of the data collection period when the excel files have been checked for accuracy, data will be destroyed Qualtrics.

In order to calculate the Social Return on Investment (SROI) for the initiative, all input and output data from the initiative will also be collected and will include financial statements and resource

inputs and details on games played and events organised. This data will be shared by the initiative via Excel spreadsheet.

Quantitative data analysis: Data will be exported from excel format to SPSS V.30 for analysis. In order to calculate the Social Return on Investment (SROI) computed stakeholder data along with all inputs and outputs data from the initiative will be inputted into the SV International Value Map (Excel).

1b What data (for example the kind, formats, and volumes), will be collected or produced?
Note: Information derived from previously existing data sources - namely output, processed, analysed data – are to be considered new data under this question.

- Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media
- Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf)
- Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used
- Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such 'preferred formats')
- Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns)

<u>Kind of Data:</u>

Qualitative Data: Qualitative methodologies will be used to develop Theories of Change for each stakeholder group that will focus on a) profiling the stakeholder group, b) the environment that supported their involvement, c) the impact of their involvement and d) the outcomes experienced as a result of their involvement. Throughout implementation, qualitative data will also be a collected to assess implementation (adoption, fidelity etc.). Only audiofiles will be saved and uploaded to OneDrive for transcription (only relevant for data collection on Zoom). Audio files will be transcribed verbatim for analysis and once checked for accuracy, they will be deleted from OneDrive. This will occur within 7-10 days posts data collection.

Quantitative Data: Quantitative data will be downloaded from Qualtrics and the Kitman Labs data collection portal in Excel spreadsheet format. Once saved, they will be uploaded to OneDrive. All 'Individual Data Event' (IDE) Excel spreadsheets from Kitman Labs will be merged to create a 'Masterfile' for analysis. All Qualtrics survey Excel spreads for C1 participants will be merged with the Masterfile. The Masterfile will then be exported to SPSS to be analysed.

<u>Data Format:</u>

Qualitative Data: Transcriptions will be coded by the stakeholder group and interview/focus group number in sequence of occurrence e.g. volunteer coordinator interview no. 1 would be coded as CVI1.pdf.

Quantitative Data: IDE spreadsheets from Kitman Labs will be coded by timepoint and site (there will be 7 IG sites and 4 CG sites) for storage before being merged into the Masterfile e.g. IDEB1.xls. IDE spreadsheets from Qualtrics will be coded for storage by timepoint, data collection method and site e.g. IDEBQualtrics1.xls. A Masterfile at each timepoint will be created and coded for storage by

timepoint and sites e.g. MasterfileB1-11.xls. The final Masterfile of all timepoints and all sites will be coded for storage to represent all timepoints and sites i.e. MasterfileB3M12M1-11.xls.

Justification of the use of file format:

Qualitative Data: A PDF format is preferred for the final transcript so that it cannot be edited further.

Quantitative Data: SPSS is the preferred format of the statisticians on the research team for data analysis and Excel the preferred format to export to SPSS.

Sharing and Long-Term Re-use of Data:

Initially, only members of the research team will have access to the data. Access will be controlled and monitored in compliance with current GDPR laws. All data will be stored on a password protected OneDrive owned by the PI. Within the 'Project Folder' specific folders for different aspects of the study will be created e.g. qualitative data and quantitative data by stakeholder group. Only those members of the research team who require access to specific folders for analysis will have the relevant access. The PI will share the folders accordingly. SETU will be the data controller for all data pertaining to this study.

In the longer-term, data will be made available on the Zenodo repository on the recommendation of and with the support of the Research Support Unit (RSU) at SETU. Currently, no member of the research team has experience of data sharing via this repository but the RSU has staff with this expertise.

Volume of Data:

Qualitative Data: Collection will involve ~10-30 per stakeholder group depending upon the stakeholder group. In total, a maximum of 120 stakeholders across all groups will be included in the study comprising of ~35 transcripts across interviews and focus groups.

Quantitative Data: Approximately 900 men will participate in the IG and 300 in the CG. Data will be collected for these 1200 men across 3 time points. The variables to be collected at each time point include:

- Survey: Reach (date of birth, ethnic origin, educational attainment, relationship, housing and employment status), how participants had heard about FC, smoking status, use of primary care services and prescription medicine, perception of health and workplace capacity. Mental well-being, structural and functional social relationships, generic health status, fruit and veg intake and alcohol consumption (n=18 variables for ~1000 men).
- On-site: Blood Pressure, Height, Weight, Body Composition, Lower Body Strength, Aerobic Fitness and Cholesterol (n=7 for 300 men)
- Wearables: Heart Rate, Physical Activity, Sleep (n=3 for 150 men)
- GPS Units: Heart Rate, Speed, Total Distance, Accelerations (n=4 for 50 men).

Therefore a total of 20,650 variables will be collected from the 1,200 participants at each time point equating to a total of 61,950 data points. Each participant will be assigned a row in the IDE Excel spreadsheet and each variable will be assigned a column. While there are a considerable number of data points to be collected, both Excel and SPSS spreadsheets are sufficient to log this data.

2a What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used
- Use community metadata standards where these are in place
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on
- Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks

All transcripts, the IDE Masterfiles and the Masterfile (MasterfileB3M12M1-11.xls) will be made available for re-use. During the project data will be stored in a folder on the PIs OneDrive entitled 'Data'. Within that folder, subfolders will be created entitled:

- Stakeholder Consultations – within which there will be a subfolder for each stakeholder group in which the transcripts for that group will be stored.
- IG Participant Outcome Measures – within which there will be a subfolder for each site that will house the IDE and Masterfile for that site.
- CG Outcome Measures – within which there will be a subfolder for each site that will house the IDE and Masterfile for that site.
- IG and CG Masterfile Outcome Measures – within which the Masterfile will be housed across all time points.

A folder entitled 'Procedures' will also be created and within that the 'Procedures Manual' for all data collection will be stored. Specifically this manual will be developed throughout the planning, data collection and analysis phases of the study and will detail the definitions of variables, units of measurement, methodology used to collect the data, data collection conditions and staff as well as analytical and other procedural information. This manual will be developed as a word document and saved for sharing as a PDF document.

2b What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies

The quality control of measures will include:

- The calibration of all equipment to be used for on-site measures.
- Employing the same staff to take the same measures at each site across all time points.
- Taking repeated measures where necessary to ensure the validity of the result i.e. lower leg strength and waist circumference.
- Removing the necessity for paper-based data collection by using Qualtrics and the Kitman Labs data collection portal. All data will be collected electronically and exported to Excel. Surveys and prompted questions will be completed by stakeholders online and onsite measures will be inputted to the Kitman Labs portal by members of the research team who will double check the input before saving data. Data from both the wearables and the GPS vests will also be automatically uploaded to the Kitman Labs portal.

- Peer review of manual data entry in the event that paper based data collection needs to happen for a particular individual and/or at a particular site.

3a How will data and metadata be stored and backed up during the research?

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended

Throughout the study the data will be securely stored on both the Universities OneDrive (specifically the PIs account) and on the data collection systems employed i.e. Qualtrics (PIs University Account) and the Kitman Labs Portal (backed up on their secure cloud storage facility). This back up will be performed at each data collection timepoint. Ethical approval has been given for this study, and as per that approval, no data will be stored, on any portable device or desktop. All data must be securely stored on the PIs OneDrive and the password protected Qualtrics and Kitman Labs cloud storage facilities.

3b How will data security and protection of sensitive date be taken care of during the research?

- Explain how the data will be recovered in the event of an incident
- Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships
- Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed
- Explain which institutional data protection policies are in place

In the event of an incident with the University's OneDrive, in addition to getting support from the University's Computer Services to access/retrieve the data, copies of the data will be retrieved from the Qualtrics and the Kitman Labs cloud based facilities. In the event of an incident in either Qualtrics or Kitman Labs, their IT specialists will work to access/retrieve the data on our behalf. Notably, data will be exported from both external cloud based facilities immediately post data collection and backed up on the University's OneDrive.

All members of the research team who will have access to the data are either staff or students of SETU. The relevant folder containing data will be shared with team members with a role in the analysis of that data. The PI will determine collaboratively who will have access and share the appropriate folder.

With respect to data protection the main risks are as follows:

- Inappropriate disclosure of personal data internally due to a lack of appropriate controls being in place. As stated above, audiofiles that would identify stakeholders will be uploaded to the PIs OneDrive immediately post data collection before leaving the data collection site and removed from the Digital Dictaphone. If on Zoom it will initially be stored on the computer and once uploaded it will be deleted from the computer. When transcribed the audiofile will be deleted from OneDrive and all personal/identifiable details will be removed thereby anonymising the transcript.

- All quantitative data will also be anonymised as participants will only be identifiable by their mothers' maiden name and date of birth (C1 participants) and their Unique Identifier (UID) (C2-4 participants). Therefore, the research team will only have access to anonymised data and no member of the research team will be able to identify or have access to contact details of members of the participant stakeholder groups. All emails to other stakeholder group members will be deleted from email accounts and stored securely on the OneDrive created for this study.
- Accidental loss of electronic equipment: As soon as the interview/focus group is recorded, it will be uploaded to the PIs One Drive and deleted from the recording device. Interviews will occur in a place with Wi-Fi to enable this to happen. Therefore, qualitative data will not be portable devices and therefore the loss of same will not breach data protection. Once interviews are transcribed verbatim, the audio recording will be deleted. This will be completed within 10 working days of data collection. All quantitative data will be recorded on site via tablets and uploaded to the data collection portal after each data collection evening. No data will be stored on the tablet and access to both the tablet and the data collection portal on the tablet will be password protected i.e. two passwords will be required to access the data via the tablet should a tablet get lost.
- Handling of paper data: It is envisaged that no data will be collected in paper format. However, should participants not complete the consent/Par-Q/surveys on their phone, hard copies will be brought to data collection evenings and completed prior to data collection. All will contain UIDs only. The collected data will be placed in a sealed envelope and stored securely on site to prevent unauthorised access. The collected data will be immediately transferred to SETU, where it will be held securely and kept in a locked filing cabinet in the Department of Sport and Exercise Science office in SETU [PIs office]. All information [consent and Par-Qs] in paper form will be scanned and uploaded to OneDrive before being destroyed i.e. confidentially shredded. Survey data will be manually inputted into the IDE spreadsheet. This will be done within 5 working days of collection.
- Emails: All emails between the members of the research team and stakeholders in this study will be removed from email accounts and stored securely on OneDrive. This practice will be done weekly to ensure stakeholder anonymity should researcher emails be breached.

Both SETUs Data Protection and Data Retention Policies are in place and have informed this DMP. Both can be found here:

Data-Protection-Policy.pdf (setu.ie)

Data-Retention-Policy.pdf (setu.ie)

4a If personal data are processed, how will compliance with legislation on personal data and security be ensured?

Ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with:

- Gain informed consent for preservation and/or sharing of personal data
- Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data)
- Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible)
- Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party)

- Explain whether there is a managed access procedure in place for authorised users of personal data

Confidentiality: The researcher will be guided by SETU's data protection and compliance officer and will adhere to the HEI's data policy. Confidentiality of the individuals will be maintained throughout the project. It is imperative that participants have confidence that the information they give will be held in confidence and that their identity will be anonymous. Participants will be informed that their data will be held in confidence. The maintenance of confidentiality is also related to anonymity and data storage, which are discussed below.

Informed consent: All participants will receive detailed information about the study. Processing of participants' data will only occur with their consent. All participants will be fully briefed on the data that will be gathered, and its purpose and utility will be explained. Participant anonymity will be ensured when reporting results. Information will be given to the participant stakeholder group and informed consent obtained at the front of the questionnaire (Qualtrics link) that will be sent to them. Information and informed consent sheets will be emailed directly to all other stakeholder groups.

Anonymity: Qualitative Data: Data from the interviews/focus groups will be transcribed verbatim and all stakeholders' names will be replaced by codes/pseudonyms. All identifying information will be removed as far as possible. Participants will be asked to provide their date of birth (DoB) and mothers' maiden name (MMN) for identification purposes and link this information to their UID. Codes/Pseudonyms will be stored on a separate OneDrive folder within the PIs account.

Quantitative Data: Participants in both the IG and CG will be given a UID. Members of the research team will ascertain their DoB and MMN for identification purposes and link this information to their UID. Only their UID, DoB and MMN will be recorded on the data collection portal i.e. the research team will not know the name or contact details of participants and will only contact them through the FC gatekeeper (C1 participants) or the data collection App on their phone (C2-4 participants). To access their participation data from the FC App, the FC initiative will also collect their DoB and MMN at registration so that the research team can share their UID with the FC initiative. The FC initiative will not have access to any of the data on the data collection portal. Sharing of the UID is only so that the FC initiative can identify the participants at each site who are participating in the study so that they can share participation data with the research team. All other stakeholder groups: At 12M these groups will be sent a questionnaire that they will be asked to complete anonymously i.e. no identifying information will be collected.

For this study, only members of the research team will have access to the raw data. No collaborators outside of SETU will have access to the data. Access will be controlled and monitored in compliance with current GDPR laws. All data will be stored on a password protected OneDrive owned by the PI; the PI will give access to relevant folders to members of the research team that require access. Both Kitman Labs and Qualtrics are also GDPR compliant. SETU will be the data controller for all data pertaining to this study.

4b How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Explain who will be the owner of the data, meaning who will have the rights to control access:

- Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which?

- Consider the use of data access and re-use licenses
- Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement
- Indicate whether intellectual property rights (for example Database Directive, sui generis rights) are affected. If so, explain which and how will they be dealt with

The IP and ownership of all of the data resides with SETU as all members of the research team are affiliated with SETU. On completion of the project which includes the dissemination of same, data will be made available as per above with no restrictions on usage.

4c What ethical issues and codes of conduct are there? How will they be taken into account?

- Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning.
- Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.

Ethical approval has been granted for this study by SETUs' Research Ethics Committee [Ref: SETU/REC/24/25/081]. The Ethics Application included a detailed 'Data Protection Impact Assessment' and as detailed above, this includes issues such as confidentiality, informed consent, anonymity, data access and storage (see above for details on all of these) and data destruction. As per SETUs' policy, all data will be destroyed after 10 years. Notably, SETU's Data Protection and Data Retention policies referenced above are in keeping with international best practice.

5a How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism)
- Outline the plan for data preservation and give information on how long the data will be retained
- Explain when the data will be made available. Indicate the expected timely release
- Explain whether exclusive use of the data will be claimed and if so, why and for how long.
- Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents
- Indicate who will be able to use the data
- If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions

The study will commence in January 2026; it may be necessary to collected data in 2027 as well as 2026 if numbers are low in 2026. Therefore, it is expected that data collection will not be complete until December 2027. Given data processing and dissemination of findings, it is expected that data will not be made available for re-use until June 2028. Exclusive use of the data will be claimed by the research team until then to allow for the publication of findings. The data, as per above, will be made available on the secure Zenodo data repository for a period of 10 years after which time it will be destroyed. All data (as per above) will be freely available to anyone with access to the Zenodo repository. See Section 2a for details on what metadata will be shared.

5b How will data for preservation be selected and where will data be preserved long-term (for example a data repository or archive)?

- Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes
- Indicate how it will be decided what data to keep. Describe the data to be preserved long-term
- Explain the foreseeable research uses (and/ or users) for the data
- Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked

According to SETU's Data Retention Policy, all data must be retained for a period of 10 years after which time it must be destroyed. The data will be stored on the Zenodo repository which has been checked for metadata standards and costs. It is probable that others seeking to understand the effects of community based physical activity/sport/football initiatives on health and/or study SROI will access the data for meta-analysis. See Section 2a for details on what metadata will be shared.

5c What methods or software tools are needed to access and use data?

- Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data.
- Indicate whether data will be shared via a repository request handled directly, or whether another mechanism will be used

As per Sections 1b and 2a above, files will be saved as PDF documents, Excel and SPSS spreadsheets. Therefore anyone wishing to access the metadata will need this software to do so. Data can then be shared via a repository request handled directly, or if someone contacts the PI directly.

5d How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

- Explain how the data might be re-used in other contexts. Persistent identifiers (PIDs) should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and re-use
- Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier

As per Section 5b above, it is probable that the data may be used by others seeking to understand the effects of community based physical activity/sport/football initiatives on health and/or study SROI will access the data for meta-analysis. It is our understanding that the Zenodo repository will assign a persistent identifier and manage same as per the service they offer. They will work with a DOI registration agency e.g. CrossRef that will provide both the infrastructure and maintain their persistence and resolvability. Each dataset will be accompanied by standardized metadata (such as title, authors, publication date, repository, keywords), which will submitted to the DOI registration agency. In this project, all datasets (see Section 2a above) will be assigned a unique DOI. The PI and two members of the research team will receive training on DOI usage and oversee DOI compliance via regular auditing to verify all published datasets have DOIs, metadata completeness and DOI resolution.

6a Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

- Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible
- For collaborative projects, explain the co-ordination of data management responsibilities across partners:
  - Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and if necessary revised
  -Consider regular updates of the DMP

The PI, Dr Paula Carroll, will have overall responsibility for data management. Members of the research team will play a key role in certain data management activities;

- Data capture; Dr Rory Shepard (Post Doctoral Researcher), Research Assistant yet to be appointed, Chief Ibrahim (PhD student) and Panagiotis Papageorgio (PhD student).
- Metadata production; Dr Rory Shepard and Dr Paula Carroll
- Data quality, Dr Rory Shepard, Research Assistant yet to be appointed, Chief Ibrahim (PhD student) and Panagiotis Papageorgio (PhD student).
- Storage and backup; Dr Rory Shepard, Chief Ibrahim (PhD student) and Panagiotis Papageorgio (PhD student).
- Data archiving; Dr Rory Shepard & Dr Paula Carroll
- Data sharing; Dr Rory Shepard & Dr Paula Carroll

While this is not a collaborative project with other partners, the DMP will be reviewed quarterly by Dr Paula Carroll & Dr Rory Shepard and updated accordingly. Ultimately, Dr Paula Carroll is responsible for implementing the DMP.

6b What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-Usable)?

- Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in
- Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges
- Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered

The goal of data management and stewardship is to ensure that data management and storage will be ethically and legally compliant in line with GDPR and to share data and study documentation in a way that reaches as many people as possible. The FAIR principles will be achieved as follows:

- Findability - A description of the project will be made available on the 'Centre for Health Behavior' webpage on SETU's recently launched new website detailing the who, what, when, where, and why of the study along with acknowledging the source of funding. The description will use commonly accepted terms and phrases (MeSH) that matches how other studies refer to the same information. Specifically, this study will be registered as a clinical trial and in addition, details of the study will be available on that website for public consumption as will all outcomes of the study.
- Accessibility - An audit of the volume of data to be generated in the study will be conducted. As per above, it is envisaged that this will include qualitative and quantitative data from a variety of sources. In order to effectively audit data on an ongoing basis it will need to be well organised and adequately documented. A project folder will be created on the PIs SETU

OneDrive account; subfolders pertaining to each research objective will be created to house relevant data. All data will be electronically recorded and anonymised. Once the data has been audited, organised and

- anonymised, the research team will identify the datasets suitable for sharing as per Section 2a above. Thereafter, details of the study will be shared online via 1) the Centre for Health Behavoiur Research group webpage where we will post information about the study, 2) publishing publicly accessible supplemental files with journal articles and 3) the use of the Zenodo repository and 4) making data available upon request to PC.
- Interoperability and Reusability - Information will be prepared in a manner that is easily shared between systems and merged with information from other sources as per Section 1b above.

Costs to prepare data have been included in staff costs (RS & PC) who are responsible for data management as per Section 6a above. Data storage will be on OneDrive and Qualtrics are accounted for under University agreements with same. A specific contract has been drawn between Kitman Labs and SETU for the use of the portal for the duration of the study and the cost of this service has been accounted for in the overall consumables costs for the project. Costs for the Zenodo repository are also covered under a University agreement. Therefore all data preparatory and storage costs have been accounted for to ensure the DMP can be implemented as proposed.