

Green paper Consultation on recommendations for developing Population  
Research UK

Published by HDR UK as part of the Population Research Development programme  
on behalf of ESRC, MRC and Wellcome

July 2021\_v5 for publication

# Contents

<b>Contents</b> .....	<b>2</b>
<b>Executive summary</b> .....	<b>3</b>
<b>Introduction</b> .....	<b>4</b>
Responding to this consultation .....	4
<b>Needs and opportunities</b> .....	<b>5</b>
<b>The vision for PRUK</b> .....	<b>10</b>
The anticipated impacts of creating PRUK.....	10
Delivering the vision of PRUK.....	11
<b>Recommendations for delivering PRUK</b> .....	<b>12</b>
R1: Enhance discovery capabilities for LPS .....	12
R2: Streamline data access processes and mechanisms .....	13
R3: Facilitate linkage between LPS data and health, administrative and further types of data .....	14
R4: Extend strategic working that enhances interoperability between LPS.....	15
R5: Sustain public trust and transparency in LPS data sharing.....	16
R6: Maximise the connectedness of LPS with the research, innovation and policymaking ecosystem .....	17
<b>Delivering PRUK through a coordinated network of infrastructures and capabilities</b> .....	<b>18</b>
<b>Conclusion and next steps</b> .....	<b>20</b>
<b>Appendix 1: Design and Dialogue consultation outputs</b> .....	<b>21</b>
<b>Appendix 2: Summary of consultation questions</b> .....	<b>22</b>

## Executive summary

---

The UK's Longitudinal Population Studies (LPS) are invaluable resources that researchers can use to address many of societies' biggest challenges, such as the Covid-19 pandemic, climate change, inequalities and the ageing population. However, the full potential of the UK's LPS has yet to be realised due to long-standing issues in some areas around data discoverability, access, linkage, and cross-discipline collaboration. This Green Paper details these issues and sets out recommendations for how a new initiative, Population Research UK (PRUK), can be developed with the LPS community in response.

**The vision for Population Research UK (PRUK) is to maximise the use, innovation and benefit from the UK's rich collection of LPS across social and economic, and biomedical science. By bringing studies and data together and tackling common challenges, PRUK will enhance collaboration and alignment across LPS and allow researchers to address high-impact research questions making the best use of LPS data resources. This can enable a greater understanding of the complex interplay between biological, social, economic and environmental determinants of health, and social and economic outcomes.**

HDR UK are scoping, on behalf of the Economic and Social Research Council (ESRC), the Medical Research Council (MRC) and Wellcome, how this vision can be achieved. This paper sets out high-level six recommendations for potential activities for PRUK. These are:

1. **Enhance the discoverability** of UK LPS by implementing metadata and discovery standards more broadly, increasing dataset visibility, and providing tools and support for discovery and navigation.
2. **Coordinate and streamline access to LPS data** by aligning data access processes and by setting a system-wide approach to provide a coordinated network of data-sharing platforms and safe havens that streamline access and use of more LPS data.
3. **Facilitate linkage between LPS data and health, administrative and other types of data** through the providing centralised linkage as a service and jointly solving outstanding challenges in creating data linkage.
4. **Expand opportunities for aligned activity and enhanced interoperability of LPS** by resourcing greater multi-study activities and support capacity development and skills within study teams.
5. **Sustain trust and transparency in LPS data sharing** through involvement and engagement with the public and study participants.
6. **Maximise the connections between UK LPS and the research, innovation and policy ecosystem** through enabling the community to function as a broader collective.

It is proposed that PRUK is delivered through drawing together a collection of existing infrastructures and initiatives that support LPS, enhanced for their functionality and in their alignment, coordinated through the development of a new LPS Data Resource Hub. This would ensure PRUK is placed to build upon existing infrastructures and support broad inclusion of disciplines and expertise.

The outputs from this scoping, shaped by the LPS community through responses received to this document, will inform a Prospectus for PRUK that will be taken forward by ESRC, MRC and Wellcome.

# Introduction

---

The UK supports an unparalleled collection of Longitudinal Population Studies (LPS) that provide a wealth of long-term information from participants, describing both them and the society in which they live. To maximise the scientific potential of this data now and for the future, three key funders, the Economic and Social Research Council (ESRC), the Medical Research Council (MRC) and Wellcome, have come together to respond to the needs of the LPS community, building on independent strategic reviews of longitudinal studies by these funders<sup>1</sup>. Subsequently, Health Data Research UK (HDR UK) have been commissioned to undertake a development programme in conjunction with the LPS community and wider stakeholders, to understand the priority needs and what support and infrastructure a new initiative could offer, under the title Population Research UK (PRUK).

HDR UK are running a 'Design and Dialogue' consultation with individuals across the UK LPS research landscape to provide an understanding of where more action is needed to maximise LPS data use. This has so far consisted of 67 interviews with invited individuals from within and outside academia, across disciplines and career points, and with experience of running LPS and using LPS' data. The consultation was supported with an online survey, to which over 200 individuals responded. A separate scoping review was commissioned to establish best practice and recommendations for engagement and involvement with the public and study participants in PRUK. These inputs are summarised through this document and links provided in Appendix 1. The programme has also been guided by an expert Partnership Group drawn from membership of the LPS community and the public.

## Responding to this consultation

This Green Paper sets out the vision for how PRUK could work to maximise LPS data use, based on the consultation described above and Appendix 1. This Paper first outlines the context of needs and opportunities in LPS data use (page 5) within which PRUK will be developed drawing on the Design and Dialogue consultation, the vision of the benefits (page 10) that could arise from the creation of PRUK, and the recommendations (page 12) for the activities and possible model for delivering PRUK.

This paper **sets out questions and invites feedback** from the wider LPS community to refine and further develop this proposal of how LPS impact can be maximised. There are consultation questions throughout the document which can be responded to via an online survey linked here.

[PRUK Green paper consultation survey](#)

After, or alongside, reading this document you are invited to submit your comments and responses to the questions via the survey link. Individual and group responses are invited. The online survey will open on 12 July and remain open until 13 August. The outputs of the consultation will be used to inform HDR UK's final recommendations for PRUK to ESRC, MRC and Wellcome in September 2021.

---

<sup>1</sup> MRC Maximising the value of UK population cohorts (2014); ESRC Longitudinal Studies Strategic Review (2017); Wellcome Longitudinal Population Studies Strategy (2017)

# Needs and opportunities

The UK has made significant investments in primary LPS that recruit and collect data from participants and further initiatives that support LPS use by enhancing the discovery, access or use of data<sup>2</sup>. However, more can be done to utilise the totality of this investment. The Design and Dialogue consultation with the wider LPS community (Appendix 1) crystallised that several long-standing challenges in maximising the value and use of LPS data remain in some areas, not only in data discovery, access and linkage, but also in fostering the cross-cutting understanding and collaboration that many pressing research questions require.

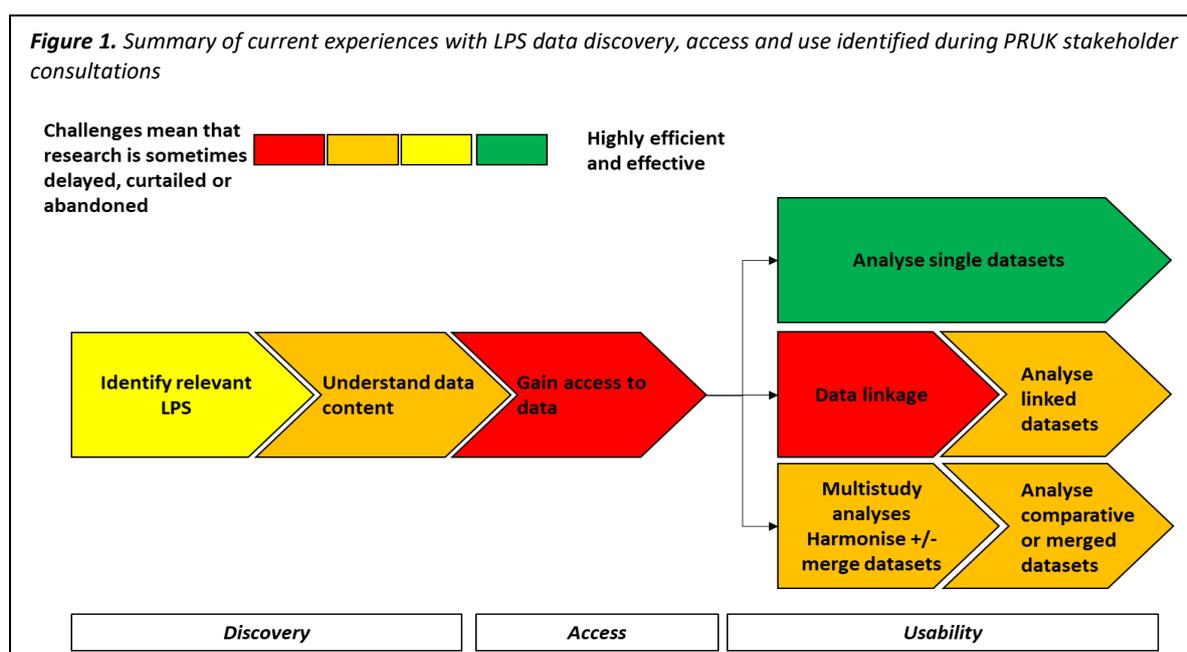
The consultation identified the **need to improve access and discoverability** of some LPS data (see Figure 1 and Appendix 1). These issues are not new, and experiences are variable. For some individual studies, including those available from the UK Data Service (a requirement of all studies funded by ESRC), effective discovery and access practices are already in place. However, remaining challenges cited in the consultation included discovering LPS with relevant populations and data, navigating sometimes **opaque and lengthy access procedures**, and the variability of **easily searchable and understandable documentation** of what data is available and how to use it appropriately.

## Consultation -illustrative responses (see appendix 1 for details)

*'PRUK could be the home for more open science and interdisciplinary research – I think people would buy that.'*

*'PRUK could provide standards, guidance and best practice around information assurance and access.'*

*'There is a need for a metadata strategy across the PRUK.'*



<sup>2</sup> Examples include: Cohort and Longitudinal Studies Enhancement Resource (CLOSER), Dementias Platform UK, European Bioinformatics Institute European Genome-Phenome Archive, Office for National Statistics Secure Research Service, UK Data Service and its Secure Lab

There are several routes to making data discoverable via dedicated discovery platforms, LPS websites or publications to identify potential data sources. However, none currently provide a complete view of all LPS resources, and the metadata available on many is collected in bespoke formats that do not support interoperability and reuse across platforms. There is a need to **apply best practice for data access and discoverability more broadly and consistently** across the whole breadth of socioeconomic-health scientific spectrum of LPS and to **broaden the number of LPS engaged in discovery and access initiatives**. To implement this, the additional resource requirements that it may place on some LPS study teams will need to be addressed. There are also opportunities and a desire to support **feasibility assessment against research question before commencing a data access request process** (e.g., understanding the profile and completeness of data of interest)<sup>3</sup>.

There are further opportunities to facilitate access to LPS data. Data is commonly distributed to users by one of several mechanisms determined by the inherent sensitivity of the data to protecting the confidentiality and disclosure risk of participants. These main mechanisms are (1) direct download under licence (as from the UK Data Service), (2) an extract of requested data prepared by the study team, (3) access to data in a Trusted Research Environment (TREs, sometimes referred to as a Data Safe Haven), which provides a system for allowing researchers remote, controlled access whilst protecting the privacy of individuals whose data they hold. A widely adopted 5 Safes Framework offers a set of principles for ensuring appropriate governance and security for data access<sup>4</sup>.

An LPS may use some or all these mechanisms for providing data access. Each data access model has strengths and weaknesses in terms of speed of access, cost, security and resource requirements, and this also varies across studies and data platforms. Making data available via a safe haven model may deliver advantages for users and studies, for example, it could reduce the need for data minimisation (whereby a user is provided with a subset of variables that are required against a hypothesis), allowing access to whole datasets which otherwise may be restricted. However, safe havens come at a higher cost per user than remote download and some have not always been viewed as user centric. In addition, maintaining datasets in across multiple data sharing platforms and TREs is labour intensive for LPS.

*'What is lacking is a central place where you can see what data might be available from what studies that can help with your research.'*

*'We're creating more and more data platforms but the data comes out in the same messy way!'*

*'Could PRUK have sandpit functionally? – so you've got the real interrogation ability to see what the data has in it'*

*'We need a small number of big archives that hold data'*

*'We need a platform with a number of cohorts, designed for worldwide access, with multimodal data, that's what everyone wants.'*

*'The role of PRUK should be to remove the concern that a project won't get done because the data can't be accessed. This means a single platform with streamlined access.'*

---

<sup>3</sup> Examples include: CLOSER Discovery that is a catalogue of variable level information across 10 LPS, including valid and invalid cases, and characteristics of the data response). HDR UK have launched a cohort discovery tool that allows the identification of individuals matching chosen characteristics within datasets. DataShield is a tool that allows remote and non-disclosive analysis of sensitive research data, which could support discovery. Synthetic data is a further solution with promise.

<sup>4</sup> [HDRUK Trusted Research Environments Green Paper](#)

TREs are an active area of infrastructure development (for instance a feature of the current government Goldacre Review<sup>5</sup>), including activities to work towards greater rationalisation, interoperability, standards, user experience and functionalities, and federation across TREs. The consultation identified a new initiative should prioritise **streamlining access to data**. There is **potentially a need for new data-sharing platforms and TREs in the future** to make more studies and types of LPS data available, but there is also a need to **optimise the use of current data-sharing platforms and understand the cumulative resource requirements that data sharing initiatives put on studies** and how to make them **sustainable, and standardising processes in data access**. It is essential that where access is already seen to be efficient, for instance, the access to some LPS datasets via download under licence on the UK Data Service, no additional barriers should be created.

There is a growing demand for **LPS data linked to a broad range of administrative, health, and other data types** to address many contemporary research questions (for examples, see Figure 2, Figure 3). Advocating for linkage and facilitating data linkage were identified as the highest priority activities in the PRUK online survey<sup>6</sup>. Data linkage may also help understand methodological challenges such as responder bias and missingness in survey data. Despite their potential linkages are seen as a high resource activity and may fail for a multiple of reasons. Challenges in creating linkages to routine administrative data were often perceived as due to reluctance on the part of data custodians, but work of organisations such as Administrative Data Research UK (ADR UK) and others, have potentially started to break down these barriers providing an opportunity to accelerate linkage programmes. Cost of creating some linkages and restrictive onward sharing permissions, awareness of available linked data, and knowledge to work with and analyse linked data were also seen as challenges to be addressed.

**Figure 2.** Frequently suggested types of data linkages with LPS (Source: Q21 – PRUK online survey Appendix 1)

<b>Crime and Justice</b>	<b>Digital footprint</b> <ul style="list-style-type: none"> <li>● Consumer data</li> <li>● Browsing history</li> <li>● Social media</li> </ul>	<b>Education</b> <ul style="list-style-type: none"> <li>● Educational achievement</li> <li>● University leavers</li> </ul>	<b>Environmental</b> <ul style="list-style-type: none"> <li>● Geospatial</li> <li>● Exposures</li> <li>● Climate</li> </ul>
<b>Health</b> <ul style="list-style-type: none"> <li>● Primary Care</li> <li>● Secondary Care</li> <li>● Mental Health</li> </ul>	<b>Financial and economic</b> <ul style="list-style-type: none"> <li>● Welfare</li> <li>● Income, Tax</li> <li>● Occupation</li> </ul>	<b>Local authority</b> <ul style="list-style-type: none"> <li>● Housing</li> <li>● Social care use</li> </ul>	<b>Wearables</b> <ul style="list-style-type: none"> <li>● Fitbits, watches</li> <li>● Apps</li> </ul>

<sup>5</sup> Goldacre review: [New review into use of health data for research and analysis - GOV.UK \(www.gov.uk\)](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/444444/new-review-into-use-of-health-data-for-research-and-analysis.pdf)

<sup>6</sup> Q22, Q23 (see Appendix 1)

*‘There are already so many points of access, like UKDS, ONS, SAIL, CLOSER – we don’t need another one!’*

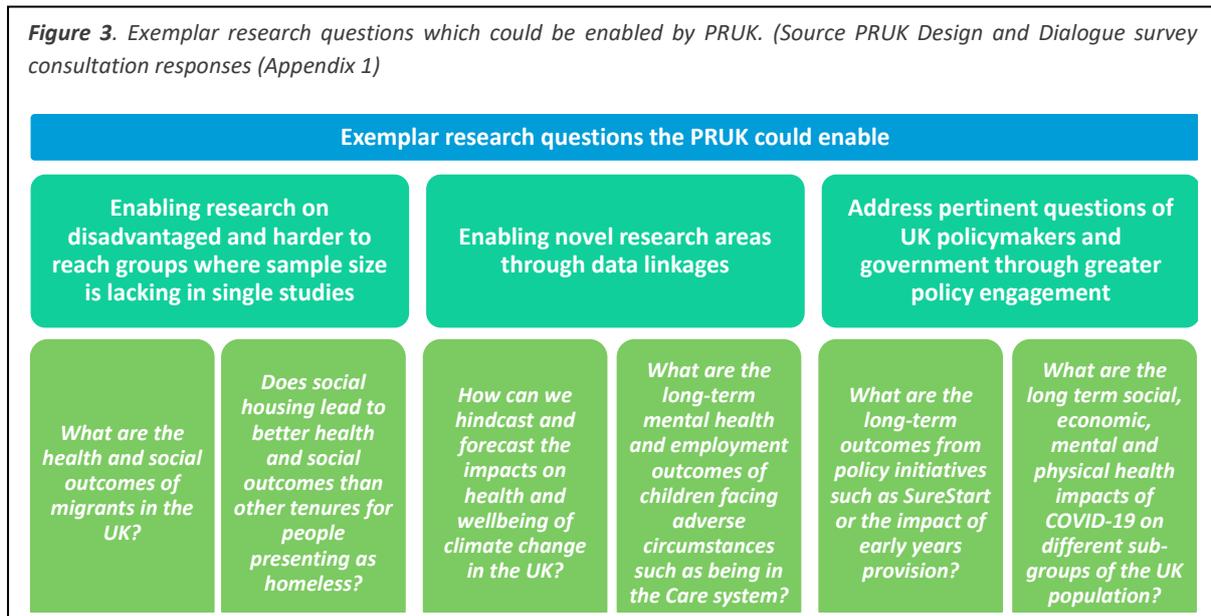
*‘Our data is downloaded 1000s of times via UKDS, so other activities need to be proportional in the resources they need.’*

*‘There’s a cultural problem with linkage. It is not really enabled. We have the permissions in place but there isn’t the inclination or resource – there’s not enough drive to make it happen.’*

*‘Working as a larger group of studies, you would have much more economic power to address linkage than an individual study.’*

*‘PRUK should become a resource for researchers to access linked data.’*

**Figure 3.** Exemplar research questions which could be enabled by PRUK. (Source PRUK Design and Dialogue survey consultation responses (Appendix 1))



There is a significant need to reduce barriers to **bringing LPS datasets together** to allow combined or comparative analyses. These opportunities can vastly increase the scope of potential research questions LPS data can answer, e.g., by enabling analysis across the life course and geographies. **Pooling LPS datasets** could allow findings to be validated across different populations but, in some instances, can increase study power to study rarer events or minority groups of interest in the population (see Figure 3). Although, aggregating studies may be a point of scientific difference between disciplines. The plurality of independent data access processes and arrangements can make it challenging to work across multiple studies, and once access is provided there is sometimes requirement for considerable resources in harmonising variables within datasets. There is an opportunity to **further support harmonisation, both prospectively and retrospectively**, of data collected between studies as an approach to support greater use of the UK LPS data resources.

The consultation illustrated the opportunity to foster greater **understanding and cross-cutting collaboration across the fragmented LPS community**. LPS data spans economic and social, biological and environmental disciplines, each operating within its own research culture, with its own policies and procedures. In recent years dedicated teams behind individual LPS have been working together more, catalysed by the workstreams and initiatives of CLOSER, with the Covid-19 pandemic triggering further multi-study initiatives across disciplines, such as the National Core Studies Longitudinal Health and Wellbeing programme and the Wellcome Covid-19 questionnaire. Continuing greater collaboration on LPS issues and opportunities for greater understanding and cross-pollination across disciplines requires ongoing investment in infrastructure and initiatives.

*‘We need to get to a point where we have a trusted framework of ethics approval that is less variable from committee to committee and able to handle multi-cohort consent.’*

*‘I think in social sciences we find the idea joining studies together quite alien, because, again, because what’s the population basis that you’re doing that on?’*

*‘PRUK should create a roadmap for interoperability within and between disciplinary domains, whilst encouraging further innovation.’*

There is a requirement to further support capacity, skill and knowledge development broadly across users and those that manage and deliver LPS. The consultation highlighted quantitative analytical training needs for users of longitudinal studies, especially if increasing availability of linked datasets and novel forms of data were priorities for a new initiative. There are already organisations engaged in providing training and resources, so any a new initiative may commit to work with other initiatives to make greater progress<sup>7</sup>. A second need identified was to **retain and further the expertise of staff that manage and prepare data within studies** required to make data available for use, recognising the expertise in data management, governance, ethics and programme management.

Studies vary in their resources and size of study teams, which has implications on the balance between running their study, engaging with their participants and collecting data, and wider engagement with LPS initiatives, stakeholders, and other data users. There is potential for LPS to pool expertise and, where beneficial, to **function as a collective**, which may generate efficiencies and accelerate progress. Opportunities may include **collective engagement with policymakers (in government, local or third sector)** to provide a more direct route for policy priorities to drive research (see Figure 3).

There is a need for **sustaining public support** toward population data sharing and linkage, which is distinct and synergistic to the engagement and involvement that individual LPS may deliver. Public attitudes toward data linkage and sharing have described the existence of ‘a social licence’ needed to legitimise the use and linkage of personal information for research<sup>8</sup>, which goes beyond formal regulation to broader expectations of society regarding the conduct and activities. Public attitudes towards data will continue to change. Therefore, there is a need for **ongoing public dialogue on social licence** and ensure that this is reflected in current and future practices. This has been recently exemplified by the public's concerns for the General Practice Data for Planning and Research scheme.

The UK is fortunate to have valued infrastructure and expertise that already supports the use discovery, access and use of LPS. Although the Design and Dialogue consultation demonstrated an opportunity advancing for data access, discoverability, linkage, and facilitation of greater cross-cutting collaboration to be addressed, the consultation reinforced that there is already expertise, impactful activity and established practices in some areas and a need for any new investment not to duplicate or reinvent activities.

*‘If you’re a platform that’s sharing data, particularly complicated data would have thought training would be a good thing.’*

*‘Investment in specialist teams (through the retention of expertise and training of new staff) to make study data accessible and discoverable across funders is essential.’*

*‘PRUK could have people who could look at the sorts of research questions government departments have, and alert them to data sources of which they might previously been unaware.’*

*‘Coming off the back of the greatest ever modern exposure to epidemiology, communication to the general population about the utility of being part of studies and allowing your data into research would be a brilliant thing to achieve’*

*‘PRUK should build on, and support, existing expertise and best practice (especially within ESRC-funded infrastructures e.g., UKDS and CLOSER) to retain an international reputation.’*

---

<sup>7</sup> [A review of quantitative analytical training needs for users of longitudinal studies, CLOSER, 2019](#)

<sup>8</sup> Elias, P. (2021). Promoting public engagement with longitudinal research: A report to the Economic and Social Research Council, Warwick Institute for Employment Research

# The vision for PRUK

---

**The vision for Population Research UK (PRUK) is to maximise the use, innovation and benefit from the UK's rich collection of LPS across social and economic, and biomedical science. By bringing studies and data together and tackling common challenges, PRUK will enhance collaboration and alignment across LPS and allow researchers to address high-impact research questions making the best use of LPS data resources. This can enable a greater understanding of the complex interplay between biological, social, economic and environmental determinants of health, and social and economic outcomes.**

A key principle to fulfil the vision for PRUK is for there to be expansion, not duplication, of valuable existing work and infrastructures combined with progressive new activities, resources and initiatives. There are many important existing initiatives that PRUK can draw on and with, although few currently cover the breadth of UK LPS across the social science and biomedical spectrum.

Key to meeting the needs of the LPS community is incorporating the full breadth of significant UK based LPS (including cohorts, panel studies and their linked biosamples), and not limited to those funded by ESRC, MRC and Wellcome. Disease-based registries being covered by other initiatives are not within the primary scope of PRUK.

## The anticipated impacts of creating PRUK

### **Broader and novel use of LPS data**

- Make it easier for researchers working outside of individual LPS to use more and broader types of LPS data to answer their research questions.
- Widen the user base of LPS data from within and outside academia.
- Enable multiple data sources to be brought together for comparative, combined and replicative analyses. For example, making it easier to bring together larger samples of minority groups or rarer health outcomes may ensure that LPS findings benefit more and wider populations.
- Accelerate the generation and use of new datasets to be applied to novel research questions, such as can be enabled by increased data linkage with routinely collected health, administrative and other forms of big data.

### **Support fresh collaborations from across the spectrum of the LPS research community and extend the collective engagement with stakeholders**

- Increase interoperability and alignment in approaches of LPS and the associated infrastructure through multi-study activities.
- Foster broader cross-cutting collaborations and benefit more widely from the data and expertise of LPS colleagues in different disciplines.

- Generate new ambitious partnerships and development opportunities with organisations and individuals from across sectors.
- Create development and progression opportunities for staff in data and study management roles that enable onward sharing and use of LPS data.

**Enable innovative and more efficient ways of using LPS data and the more rapid generation of evidence with potential for public benefit.**

- Endorse and broaden the use of sustainable and scalable infrastructure for data storage, access and distribution.
- Speed up the validation and replication of findings across datasets.
- Support a UK population research infrastructure that is primed to be more readily responsive to future emerging and new policy needs.
- Reduce the time and resources expended by users to obtain access to data leading to faster findings and fewer abandoned analyses. This should return a downstream cost benefit, through efficient use of LPS resources.
- Increase public awareness and support of the impacts of LPS research and trust in data sharing

## Delivering the vision of PRUK

For PRUK to achieve the vision and impacts outlined above, the following recommendations are set out. These represent the core, priority activities that it is proposed PRUK should deliver.

**R1: Enhance discovery capabilities for LPS**

**R2: Streamline data access processes and mechanisms**

**R3: Facilitate linkage between LPS data and health, administrative and further types of data**

**R4: Extend strategic working that enhances interoperability between LPS**

**R5: Sustain public trust and transparency in LPS data sharing**

**R6: Maximise the connectedness of LPS with the wider research, innovation and policymaking ecosystem**

It is proposed that PRUK is delivered through drawing together some existing infrastructures and initiatives that support LPS, enhanced for their functionality and resourced to support greater alignment between them. This would be coordinated through the development of a new LPS Data Resource Hub.

These recommendations and the proposed model for delivery are expanded in the following section of the Green Paper.

# Recommendations for delivering PRUK

---

## R1: Enhance discovery capabilities for LPS

There is huge diversity in the discoverability of LPS data across the UK. The role of PRUK could be to support activities that support standardised practices and increase discoverability across the full breadth of LPS. The proposed activities for PRUK are:

### **Development a metadata strategy and discovery standards across UK LPS**

- This should build on and be compatible with existing initiatives. It might incorporate (a) Recognised national/international standards to which LPS metadata should conform, taking a principle of least restriction where possible, (b) Recommendations on formats in which data dictionaries are created. (c) Increase LPS coverage on a core number of discovery platforms (focusing on those that use recognised standards, supporting interoperability and reuse).
- Provide guidance, support and leadership, and secure resources for primary LPS where needed to implement metadata standards. This will require prioritisation between and within LPS to make both legacy and future metadata about studies discoverable.

### **Create greater awareness and visibility of the totality of UK LPS data and assets**

- The UK is already fortunate to have several discovery platforms that provide both dataset and variable level discovery, (and in some cases processes for accessing data)<sup>9</sup>. Given these assets, creating a new single discovery platform to supersede these is not recommended at this time. Instead, the focus would be to increase coverage on platforms.
- In the short term, PRUK could seek to provide a high-level aggregated view of the LPS, datasets, studies and discovery platforms upon which variable level metadata can be found across the totality of UK LPS.
- Longer-term, PRUK could set out a roadmap to develop federated search capabilities across discovery platforms, built on increasing adoption of metadata standards.

### **Deployment of data utility tools to support discovery**

- There is the opportunity to scale up the use of promising approaches to understand the characteristics of a dataset and gain insight into its potential utility before a data access request. PRUK could lead in evaluating current and emerging approaches and efficient deployment of such tools into platforms and individual LPS. There will need to be support and training resources for users in the appropriate use of such tools.

### **Provide user support to navigation of the discoverability and access available to LPS in the UK**

---

<sup>9</sup> For example, metadata (at either a study and/or variable level) for LPS can be found at: CLOSER Discovery, Dementias Platform UK Cohort Discovery, European Phenome-Genome Archive, HDR UK Innovation Gateway, Maelstrom Catalogue, UK Data Service, UKRI Cohort Directory in addition to individual LPS study websites

- PRUK could develop services for users to navigate the LPS ecosystem and maximise the use of discoverability tools introduced.

### Questions for consultation

1. What would be most helpful to facilitate data discoverability across the full range of LPS and how could this be achieved?

## R2: Streamline data access processes and mechanisms

Increasing and improving the efficiency of access to the full breadth of LPS data is a crucial ambition to maximise the full use of UK LPS. The proposed activities for PRUK are:

### Development of standards for data access

- This may incorporate (a) A common and core set of information and questions asked from researchers when requesting data, (b) Development of a common set of assessment criteria for reviewing access requests, (c) Common terms of reference and operation of data access committees, (d) Recommendations for the membership of Data Access Committees, including independent and public members (e) Requirements and contents of public data access registers held by each study.

### Adoption of processes that increase and streamline access to LPS data

- Work with LPS to address barriers and support implementation of common access processes.
- Work with research funders to develop a sustainable cost model for accessing data via different mechanisms settings in partnership with studies.

### Develop a strategy for future wider use of data platforms and TREs for LPS access

- Manage a map or register of data platforms and TREs which hold LPS data, and how activities could be streamlined for both LPS teams to contribute data and for users to work within environments.
- Develop and implement a long-term strategy for creating a federated data infrastructure that supports LPS. This would recognise the need for a phased approach of working towards deeper levels of alignment federation between infrastructure. For example, alignment in approaches to recognising a 'Safe researcher' (one component of the Five Safes framework) may be a step towards federated analytics across platforms.
- Set out the requirements for investment to enhance the capabilities of current data platforms, or invest in new platforms, to provide coverage to more LPS and data types for researchers

### **Create a centralised register of LPS data access requests**

- A component of building public trust would be developing a national register of approved data access requests, including who is provided access, the purpose and duration of access. This may work through the initial development of standards for data access registers, which would be implemented at a study/access point and then aggregated across studies.

### **Questions for consultation**

2. What impact (positive or negative) do you think the implementation of common access processes will have for you or your organisation? Please explain your view.
3. For LPS teams, what actions and resources would be most helpful to increase access to the study datasets you manage?

## **R3: Facilitate linkage between LPS data and health, administrative and further types of data**

Linked data can provide new avenues of scientific enquiry, but challenges exist in their creation and use. The proposed activities for PRUK are:

### **Provide resources that facilitate greater data linkage across multiple LPS.**

- Where there is broad interest from multiple LPS to establish linkages with a dataset held by a common data custodian (for example, a government or health department), it may be more efficient to attempt to establish these linkages as a consortium of LPS. PRUK could facilitate the identification of high-value data linkages and the agreement of terms for linkages with the custodians that are acceptable from a legal, governance and ethical standpoint (in partnership with initiatives that are working with custodians of data already (for example, ADR UK).
- Develop a sustainable model of funding access to linked datasets in partnership with stakeholders and funders, promoting reuse where possible.

### **Support the increased use of linked data resources and readiness of linked data for analysis**

- There is a need to increase the awareness of available linked datasets.
- Once a linked dataset is created, many potential actions could aid preparation of the created datasets for analysis. For instance, approaches supporting multi-country analyses and comparisons to harmonise differences in administrative/health routine data from each national jurisdiction. Another activity may be to support creation of derived variables from linked data sets (e.g., that may be less disclosive and available for more permissive distribution).
- There is a need to develop skills and capacity in the analysis of linked data. PRUK should play a role, along with others working in this area such as ADR UK and the National Centre for Research Methods, in developing research capability in this area.

## Questions for consultation

4. What further actions could be taken to facilitate the process of creating and using linked data?

## R4: Extend strategic working that enhances interoperability between LPS

Maximising the full breadth of expertise and resources of the UK LPS could be supported through strengthening the resources available for multi-study activities. This would further facilitate aligned practices and foster a shared appreciation of the opportunities that LPS offer. The proposed activities for PRUK are:

### **Coordinate approaches to common study challenges and development opportunities**

The consultation provided illustrative examples of activities that could be beneficial to LPS community:

- The further design of common surveys and standardised question sets supporting prospective harmonisation of data for comparative analysis.
- Standardisation and harmonisation of new modes of data collection across a portfolio of studies such as remote, digital data collection into LPS surveys, wearable data, social media and consumer data and 'omics data. For example, furthering the harmonisation work of CLOSER that has supported activities to retrospectively harmonise survey and biomarker data.
- Developing harmonised approaches to biosample collection, storage, handling and analysis. For example, a consortium could negotiate preferential terms on services and technologies for sample storage, sample analysis and/or the purchasing of technologies for data collection.
- Approaches to increasing participant engagement/reducing participant attrition amongst different groups.
- Method development, including between different approaches that might exist between biomedical and social science approaches. There may also be a need for approaches to identify individuals present in more than one LPS study to avoid confounding analyses when undertaking some multi-study analyses.

### **Capacity building**

- Support skills development and career progression of those with roles in the delivery of longitudinal LPS, such as data and linkage managers.

## Questions for consultation

5. With an aim to build capacity, how can PRUK support skills development and progression of roles in data curation, management and linkage?
6. In what other areas could strategic working between LPS be extended?

## R5: Sustain public trust and transparency in LPS data sharing

There is a need for sustaining public support toward population data sharing and linkage, which is distinct and synergistic to the engagement and involvement that individual LPS may deliver. The proposed activities for PRUK are:

### Develop and share participant involvement practices between LPS studies

- PRUK could foster an understanding of commonalities and differences between studies in evolving concepts such as public benefit and approaches to consent, and resource or facilitate activities. This could also extend to understanding LPS participants perceptions of data sharing with third parties, including for-profit organisations, and how any discrepancies or concerns can be overcome. This could be conducted in partnership with other active initiatives in this area, such as ADR UK, HDR UK and Understanding Patient Data.
- PRUK could work with individual studies and other initiatives to gather and share best practices in their public and study participant engagement/involvement (PPIE) activities (for example, through the provision of training, resources, evaluation tools).

### Build and maintain public trust

- PRUK could support studies to create best practices of transparency throughout the entire life cycle of data use; this should include promoting good practice in providing publicly available information and inclusion of public and/or study participant representatives on all panels that govern data access, security and usage. Registers of data access use (included in recommendation 2) and could a practical action.

### Implement a public-facing communication strategy

- Data sharing and use will continue to be in the public spotlight (often using unconsented data). There is an important advocacy role that PRUK could play on behalf of the LPS community to ensure that concerns or mitigating measures in the use of unconsented datasets are not unnecessarily transferred to data collected with consent from LPS participants.
- PRUK could adopt an important communications role to improve public understanding and support of longitudinal and population research.

## Questions for consultation

7. Does the work proposed for PRUK in this area provide distinct value over the engagement and involvement of LPS engaging with their participants directly?

## R6: Maximise the connectedness of LPS with the research, innovation and policymaking ecosystem

Few LPS have the resources and bandwidth to engage with the plurality of potential organisations and individuals that could partner and bring value or impact to a study. Engagement with some stakeholders as a collective, which shows the potential scale and connectedness of UK LPS, might lead to new opportunities for research and partnership. The proposed activities for PRUK are:

### Stakeholder engagement

- There is an opportunity for the LPS community to have a focal point in engaging with industry, government and third sector as potential new users and commissioners of LPS analyses, where appropriate.
- Expand fora that bring together policymakers and researchers as a conduit for policy priorities to drive research questions (for example, greater engagement with the Areas of Research Interest published by individual government departments)<sup>10</sup>.

### Provide a collective perspective of UK LPS into the R&D landscape

- PRUK could provide a close point of interaction with funders of LPS to articulate the requirements of the research community to achieve greater sharing, use and impact of LPS data.
- PRUK could give a strong collective voice of the LPS community into the developments on the broader research and innovation landscape, for example in the development of the UK's wider digital and data architecture.

## Questions for consultation

8. Are there other roles that PRUK should play in building connectedness across and beyond LPS?

---

<sup>10</sup> [Areas of research interest - GOV.UK \(www.gov.uk\)](https://www.gov.uk/government/collections/areas-of-research-interest)

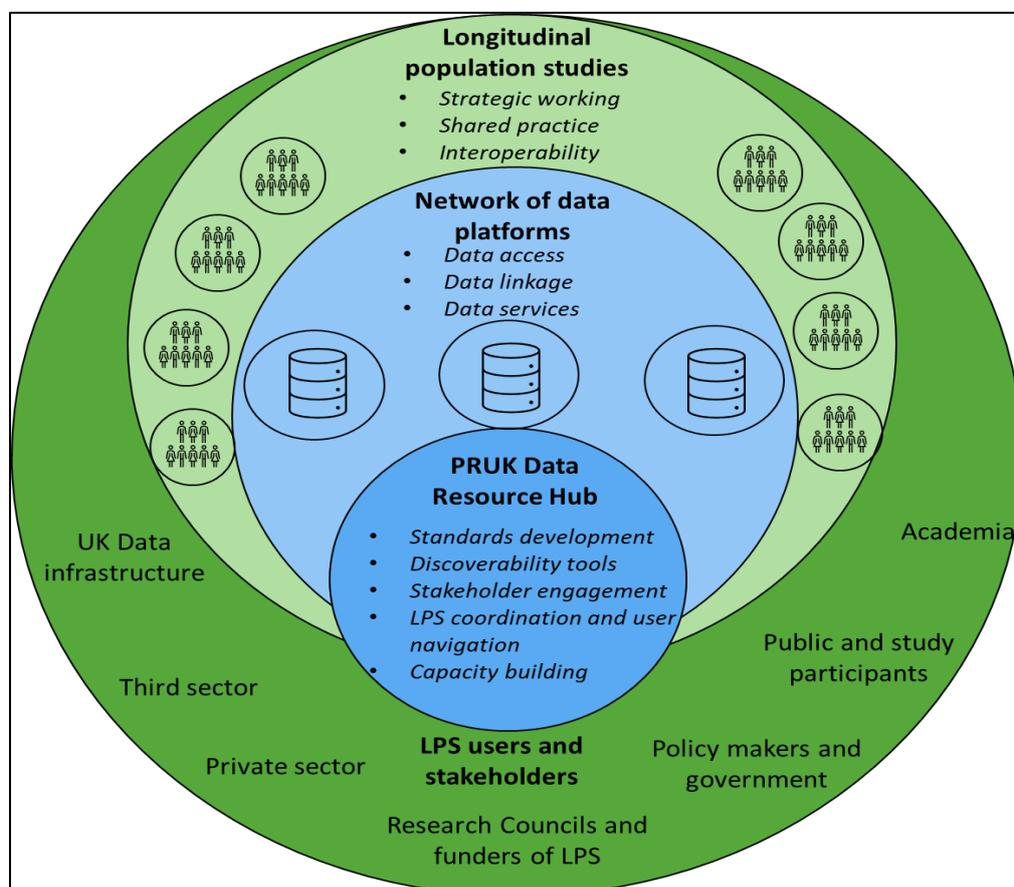
# Delivering PRUK through a coordinated network of infrastructures and capabilities

It is proposed that the activities of PRUK (described through recommendations 1-6 above) are delivered through a coordinated network of infrastructures, which will be drawn from existing and new activities. This would ensure PRUK is well placed to build upon existing initiatives and support the broad inclusion of disciplines and expertise.

Principle components of PRUK would be (see Figure 4):

- **A PRUK Data Resource Hub** that provides leadership and resources in advancing standards and engaging stakeholders across the LPS ecosystem.
- **A connected network of data platforms** that provide services and streamlined access to different types of LPS data and studies
- **Comprehensive and inclusive engagement with individual LPS** to drive forward strategic work between studies and implementation of shared practices

*Figure 4: A proposed network model for PRUK. PRUK would be delivered through a Data Resource Hub and an aligned network of LPS data platforms (shown in blue). Strategic resources will support interoperability and strategic alignment between LPS. This model would unite the strengths of UK LPS and the engagement in the wider research, innovation and policy ecosystem*



## **PRUK Data Resource Hub**

PRUK would have a centralised capability (although not necessarily located at a single location or institution) to facilitate a collective approach to the main activities set out for PRUK. The PRUK Data Resource Hub could be a new collaboration or an adaptation of an existing initiative. This would include creating standards in both discovery (in recommendation 1) and access (recommendation 2) and supporting common practices and standards where there is a need to bring together disparate approaches. In addition, the Data Resource Hub could coordinate collaboration across studies (recommendation 4), wider stakeholders (recommendation 5) and a community-wide approach to involvement and engaging the public in longitudinal population research (recommendation 6).

## **A coordinated network of LPS data platforms and infrastructures**

Access to data and data services would be supported through a coordinated network of LPS data platforms and infrastructures. Initially, this would be drawn from within existing infrastructures with potential expansion (including potential new platforms) in the future. Components of this network would bring specialism in the types of data and services that they provide whilst working towards alignment in their operation. Additional resources, where required, would support the enhancement of the data platforms towards alignment in their practices and federated approach. There would need to be a role of PRUK working between LPS and the new environments to support, streamline and resource where needed, engagement with the network.

Part of the Network activities would be to create and make available LPS data linked to health, administrative, and other data types (recommendation 3). PRUK would need to provide the necessary infrastructure to facilitate a centralised data linkage programme and the mechanism for its distribution and reuse. This would require the commissioning and funding of a unit to negotiate and manage data linkages, and a TRE to provide access to linked data resources.

Proposals for the leadership models and governance of PRUK are under development.

## **Longitudinal Population Studies**

The inclusive involvement of LPS is critical to the success of PRUK. It is recognised that additional resources to deliver the auxiliary activities in adopting standards and common practices may be required for some studies to engage. Availability of funding to support increased multi-study strategic working will facilitate novel interoperability between studies (recommendation 4), potentially allowing wider diffusion and adoption of practices through the PRUK Data Resource Hub.

The recommendations set out are intended to increase the use of LPS data and, ultimately, the impact that arises from these rich resources.

### **Questions for consultation**

9. How might a leadership and governance structure be created for PRUK that facilitates a collective approach across the community to advance the recommendations set out?

## Conclusion and next steps

---

There is an opportunity to substantially increase the potential of the UK's LPS by improving the discoverability of the data, streamlining processes for access and linkage to health, administrative and environmental data, and fostering interdisciplinary collaborations.

This Green Paper proposes a series of recommendations for PRUK activities and raises several questions for views which you can respond via the online survey:

[PRUK Green paper consultation survey](#)

The consultation outputs will be used to inform HDR UKs final recommendations for PRUK to ESRC, MRC and Wellcome in September 2021. These funders will finalise the specification for any new investment, for bidders to develop innovative proposals against the specification and for the commissioned team(s) to deliver the initiative to maximise the use of LPS data.

Thank you for your engagement and time taken to respond.

### Questions for consultation

10. What activities proposed for PRUK would be most beneficial to you or your organisation?
11. What would be the barriers to you, or your organisation, being engaged with PRUK? How could they be overcome?
12. Are there any priorities that you consider are important to advancing the LPS field that are not covered in the recommendations listed?

# Appendix 1: Design and Dialogue consultation outputs

---

Annex 1: PRUK Design and Dialogue consultation summary

Annex 2: Public and Participant Involvement in Population Research UK (PRUK): A scoping review of different approaches and recommendations

Are available via at [Population Research UK - HDR UK](#)

# Appendix 2: Summary of consultation questions

---

You can respond to the consultation via [PRUK Green paper consultation survey](#)

## **R1: Enhance discovery capabilities for LPS**

Q1: What would be most helpful to facilitate data discoverability across the full range of LPS and how could this be achieved?

## **R2: Streamline data access processes and mechanisms**

Q2: What impact (positive or negative) do you think the implementation of common access processes will have for you or your organisation? Please explain your view.

Q3: For LPS teams, what actions and resources would be most helpful to increase access to the study datasets you manage?

## **R3: Facilitate linkage between LPS data and health, administrative and further types of data**

Q4: What further actions could be taken to facilitate the process of creating and using linked data?

## **R4: Extend strategic working that enhances interoperability between LPS**

Q5: With an aim to build capacity, how can PRUK support skills development and progression of roles in data curation, management and linkage?

Q6: In what other areas could strategic working between LPS be extended?

## **R5: Sustain public trust and transparency in LPS data sharing**

Q7: Does the work proposed for PRUK in this area provide distinct value over the engagement and involvement of LPS engaging with their participants directly?

## **R6: Maximise the connectedness of LPS with the research, innovation and policymaking ecosystem**

Q8: Are there other roles that PRUK should play in building connectedness across and beyond LPS?

## **Delivering PRUK through a coordinated network**

Q9: How might a leadership and governance structure be created for PRUK that facilitates a collective approach across the community to advance the recommendations set out?

## **Conclusions**

Q10: What activities proposed PRUK would be most beneficial to you or your organisation?

Q11: What would be the barriers to you, or your organisation, being engaged with PRUK? How could they be overcome?

Q12: Are there any further priorities that you consider important to advancing the LPS field that are not covered in the recommendations listed?