

What needs to be considered when delivering a data-based research project involving multiple UK pension providers

—
MARCH 2022



This paper sets out the three essential areas that need to be considered when working across multiple providers and their data to successfully deliver an amalgamated database of member data. Our ambition is that others involved in similar projects can benefit from the experience and learnings. The three areas to consider are:

- **Governance;**
- **Data; and**
- **Teamwork.**

The paper explores each of these areas highlighting specific learnings. These learnings reflect the experience of B&CE provider of The People's Pension, Legal & General, Nest Insight the public benefit research and innovation centre set up by Nest, Now: Pension, Pensions Policy Institute, Smart Pension in delivering "The Pensions Data Project".

Summary - What to consider when merging personal data from multiple providers

Based on the work to date from The Pensions Data Project, anyone wanting to undertake a similar project, involving the merging of personal data from different providers needs to consider the following points as they go through the process.

Governance

- The project objectives and outcomes need to be clearly defined and documented at the start.
- Clear legal documents outlining roles, responsibilities and protections need to be established, agreed to, and formalised at the start of the process.
- A neutral organisation can play a key governance role helping to facilitate and coordinate activity as well as acting as an 'honest broker' outside the competitive relationships between providers.
- A formal, and clearly documented, governance framework is essential when working across several organisations, especially competitors.
- Privacy policies and notices of the providers must permit data to be used for research purposes.
- Smooth and timely decisions require a clear map of the processes and gatekeepers within each involved organisation.
- Understanding and agreeing where liability sits among and between the various participants is essential.
- It must be clear who will have access to the merged data.

Data

- A Data Protection Impact Assessment must be completed where a project involves processing personal data. It will also help to identify possible risks and issues for the project.
- A trusted third party will need to be engaged, to act as data aggregator, receiving and aggregating data from the providers and enabling researchers to analyse the merged data securely.

- Develop and agree a set of specific research questions to be addressed, and use this to develop clear, explicit data scope so that all parties are working to the same parameters.
- Any data scope should include explicit instructions and examples of how data should be laid out and formatted.
- Test, test and test again the data definitions and formatting with an initial trial project that is more limited in scope than the ultimate goal, including using both dummy and real data.
- Establish the impact of a provider's privacy policy on format of data including:
 - o Whether data needs to be pseudonymised or anonymised.
 - o The process for pseudonymisation / anonymisation.
 - o Resolve the process and protocols for merging and storing the aggregated data.
 - o Ensure suitable consideration is given on how data will be matched.
- Determine a secure and compliant way to transfer data.
- Agree and document for how long and where the raw and merged data is held.

Teamwork

- Agree clear, common project goals, objectives and outcomes, which are defined and documented and are in line with competition law safeguards.
- Allow time and space to build trust and confidence and to establish ways of working between the project participants, by running a number of initial trial projects that are more limited in scope than the ultimate goal.
- Be patient. Be prepared to go through multiple iterations to achieve the desired outcome.
- Ensure that the necessary resources are in place in all participants, ahead of being needed, specifically from the data, legal and policy team.
- Learn from the process, learn from others and share lessons.
- There is no such thing as over communication.

Key considerations for data-based projects

Currently in the UK there is no central longitudinal research base of people's total retirement savings. This exists in other countries (such as the USA) and is a powerful tool for evidencing how individual citizens' retirement savings, aggregated across their different schemes and providers, are evolving over time, thus helping to generate evidence-based policy for both pension providers and the UK Government.

A research collaboration involving B&CE provider of The People's Pension, Legal & General, Nest Insight the public benefit research and innovation centre set up by Nest, Now: Pension, Pensions Policy Institute, Smart Pension have been working on "The Pensions Data Project", a project to amalgamate data across multiple providers in order to get insight into people's savings patterns, providing the industry with a central longitudinal research database of people's total retirement savings. The current work (a Private Beta) focuses on aggregating across the master trust members with the five previously named providers.

In the longer term, there may be an opportunity to provide other industry-wide insights on the adequacy of individuals' total retirement savings and their resulting incomes by linking to:

- Other trust based and contract based Defined Contribution (DC) data;
- Defined Benefit (DB) schemes;
- Decumulation data; and
- Other data sets, such as the Annual Survey for Hours and Earnings.

The Private Beta is deliberately more limited in scope. It is working as a proof-of-concept for these longer-term ambitions, as well as delivering an important set of initial findings. It seeks to answer the following types of questions:

- o How many pension pots do people have?
- o How often do they move provider?
- o How much is their pension wealth?
- o How does this change over time?

This paper sets out the areas that need to be considered when setting up a project of this type, which could be applicable to any project which involves merging of personal data across different providers.

What aspect of governance need to be considered?

Areas that need to be considered include:

- Clear and agreed project objectives and outcomes;
- The need for a neutral body to be involved;
- An agreed governance structure;
- Clear roles, responsibilities and protections;
- Suitability of privacy notices to enable the research;
- Understanding of each organisation's decision-making processes;
- Agreement about where liability sits among and between the participants; and
- Who has access to the merged database?

The project's objectives and outcomes need to be clearly defined and documented at the start

Having a common understanding and agreement to the project's objectives and expected outcomes is a fundamental foundation for its successful delivery as it influences multiple aspects throughout the duration, for example:

- Managing the project and discussions on any 'scope creep';
- Communication with stakeholders; and
- For use in legal and other documents, such as the Data Protection Impact Assessment (DPIA).

For The Pensions Data Project its objectives and outcomes, including a description of the required analysis, were agreed and formalised in a number of key documents. Having the documents facilitated later completion of other necessary papers, such as:

- DPIAs for the project and for participating organisations;
- Trustee packs; and
- Contracts with the data processor and / or other external stakeholders.

A neutral organisation plays a key role, balancing potential competitive activities from the providers for projects like this

For The Pensions Data Project, it was agreed that the PPI would play a key governance role helping to facilitate and coordinate activity as well as acting as an 'honest broker' outside of the competitive

relationships between providers. Specifically, the PPI's role has been to manage the project, working across the providers to achieve the agreed goals, liaising with the data processor (a third party) as well as interacting with the organisations who have helped fund the Private Beta; Association of British Insurers (ABI), Department for Work and Pensions (DWP), the Pensions and Lifetime Saving Association (PLSA) and The Pensions Regulator (TPR).

Formal, documented governance is essential when working across several organisations, especially competitors

For The Pensions Data Project, a formal "Project Board" was established, with supporting formal documents, to clearly outline:

- The purpose and objectives of the project, including the analysis to be undertaken;
- The role and responsibilities of all participants in the Private Beta phase; and
- Protocols required (including what could / could not be done).

The Project Board is derived from the participating providers contributing data to the project and chaired by the PPI, as the neutral impartial participant.

Formal documents produced and agreed by the Project Board included:

- **Terms of Reference;** to clearly define the purpose and structure of the project, role of members meeting protocol including on what basis decisions will be made.
- **Memorandum of Understanding;** an agreement between all parties which defines the scope and purpose of the project.
- **Code of Conduct;** provides all Board Members with an understanding of what is required of their role.
- **Risk register;** to identify potential risks and possible mitigations in order to stay on top of potential issues that could derail intended outcomes.
- **Conflicts of interest register;** to keep track of any Board Member interests, which may be in conflict with the principals of the project and to ensure that all those involved in running the project are aware of them.

Clear legal documents outlining roles, responsibilities and protections need to be established, agreed to and formalised at the start of the process

Despite having governance procedures in place, where there is no single entity with direct legal responsibility for the project, as is the case for The Pensions Data Project, further legal documentation is required. These documents clearly outline:

- the actions required of each of the parties (including what data to provide and in what format etc.).
- the responsibilities of each party and what action would be required if any of these were breached.

This is to ensure adequate safeguards are in place to protect each provider and the personal data of their members, in line with applicable Data Protection Legislation. For The Pensions Data Project, the following documentation was produced:

- A joint agreement agreeing roles, requirements and identifying where liability sits;
- A letter of engagement with the third-party data processor documenting work to be undertaken, processes being used to undertake the work, establishing liability and ownership of data; and
- Documents clearly describing the process, the data to be processed, the analysis to be undertaken, IT infrastructure and assurances, as well as security processes in place.

Ensure that privacy policies and notices enable the research to be conducted

Privacy statements from each of the different providers needs to be reviewed to determine what each provider has communicated to its members about how their personal data will be processed and what for. This is particularly important when it comes to workplace pension data as in general, the individual data subjects have not consented to specific uses of their data. This is because they are generally enrolled by their employers, rather than proactively opting to enrol. Specifically, the following needs to be established:

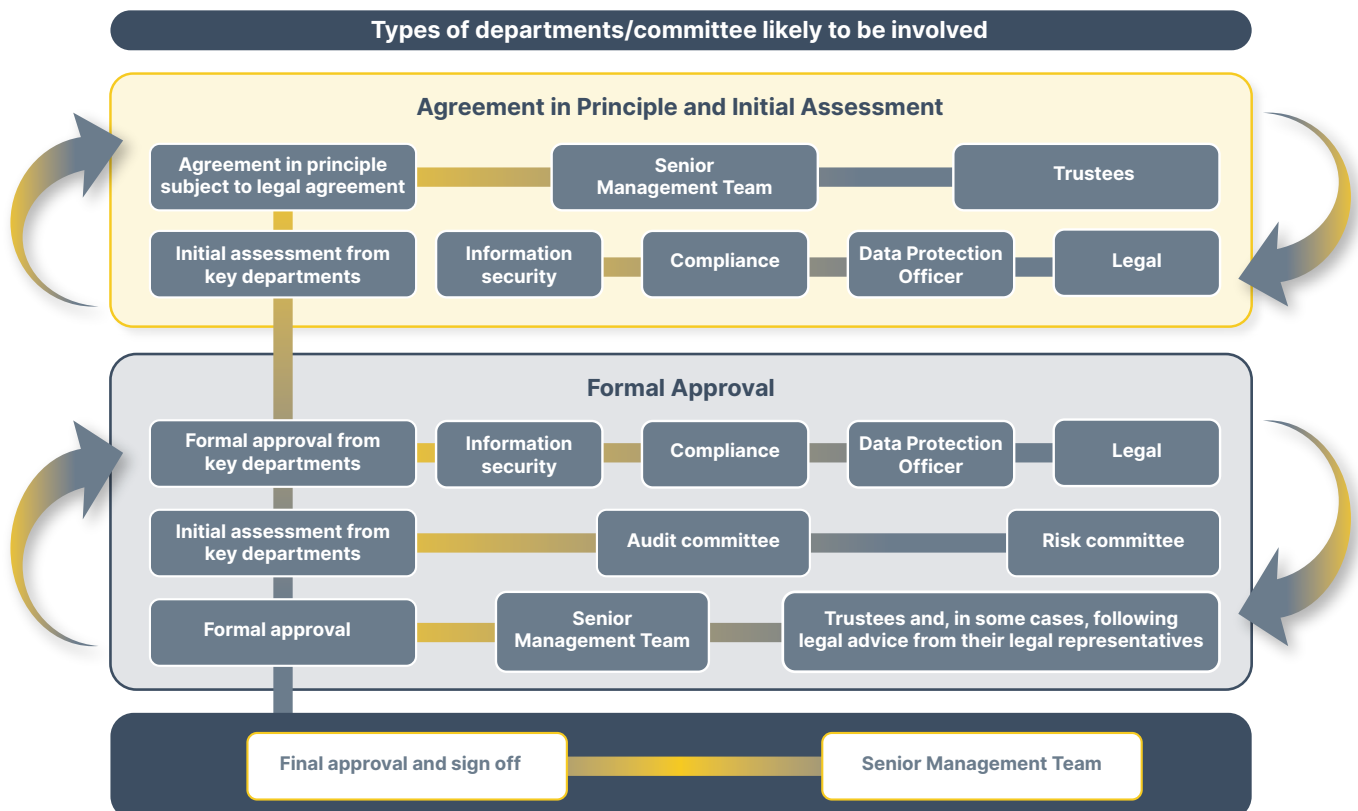
- For what reasons can their personal data be processed and used?
- How long their personal data can be held for?
- Who their data may be shared with?

If the existing privacy statements do not facilitate the anticipated processes (how data will be transferred, merged and analysed, for example) then either the process being used has to be changed or the privacy notice needs to be changed (and the associated time lapse that this generates is built into the project timeline).

Smooth and timely decisions require a clear map of the processes and gatekeepers, within each involved organisation

The Pensions Data Project is the first of its kind to merge pension scheme data from multiple providers, and therefore no precedent has been set within most of the providers either externally or internally as to what agreements and processes need to be in place and who within each organisation needs to be involved. Therefore, one of the first requirements of each board member was to establish the decision-making and approval process for such a project within each of their organisations. This also helped with early buy in to the project and what it is trying to achieve, as well as identify what processes needed to be followed.

Although each provider/scheme has different procedures/committees involved, the following highlights the different parties and layers that anyone considering undertaking a similar project might want to consider²:



² This is not an exhaustive list and individual providers/schemes may have other committees/departments also involved.

Understanding and agreeing where liability sits among and between participants is essential

As The Pensions Data Project requires the transfer of data from one party to another and merging of data from multiple parties, extensive thought was given to identifying all of the different scenarios where liability could sit in the case of a potential breach or potential breaches. These scenarios are clearly documented and relevant protections and financial liability established and accommodated.

Determine who will have access to the merged data

In order to avoid unauthorised access to data and to protect the data in line with data protection law, clear processes and documented access need to be established.

In the case of The Pensions Data Project, only named personnel within the data processor will have access to the hashed data. Any access will be:

- Authorised and accessed via passwords.
- Data will be handled with utmost confidentiality and at a security level similar to use with government engagements.

Similarly, in order to avoid any findings which could advantage one competitor over another, it was agreed that no provider would have access to the amalgamated raw data and would only have access to summary data through a visualisation tool, based on pre agreed analysis (see point on data scope for more information).

What aspects need to be taken into account when dealing with data?

Areas that need to be considered include:

- Completing a Data Protection Impact Assessment to identify possible risks and issues for the project.
- Engage a trusted third party to act as data aggregator.
- Develop and agree a set of specific research questions to be addressed and use this to develop a detailed and clear data scope.
- Specify how data should be laid out and formatted.
- Test, test and test again the data definitions using both dummy and real data.
- Establish the impact of a provider's privacy policy on format of data including:
 - o Whether data needs to be pseudonymised or anonymised.
 - o The process for pseudonymisation / anonymisation.
 - o The process and protocols for merging and storing data.
 - o How data will be matched.
- Determine a secure and compliant way to transfer data.
- How long and where raw data is held.

Complete a Data Protection Impact Assessment to identify possible risks and issues for the project.

A DPIA (a Data Protection Impact Assessment) is a process designed to help systematically analyse, identify, and minimise the data protection risks of a project or plan. It is a key part of an organisation's accountability obligations under the UK GDPR, and when done properly helps assess and demonstrate

how data protection obligations have been complied with. A DPIA needs to be carried out when a type of processing is likely to result in a high risk to the rights and freedoms of individuals.

As outlined by the Information Commissioner's Office (ICO) a DPIA should begin early in the life of a project before data processing is started and should run alongside the planning and development process. As part of the DPIA the following needs to be documented:

- a) The purpose of the processing, including whether there is a lawful basis for processing, intended outcome for individuals
- b) The context of the processing including the source of the data, likelihood of individuals expecting their data to be processed, any issues of public concern
- c) The nature of the processing including how data is collected, stored, used, accessed, shared, retention period, security measures etc.
- d) The scope of the processing including the nature of the personal data, the volume of personal data, the duration, the number of data subjects involved etc
- e) The risks associated with the processing, and mitigations being put in place to minimise these risks.

The Pensions Data Project processes personal data on a large scale, and will match and merge it. A DPIA was needed to be carried out by each of the individual providers, and to support this, an overall DPIA was done at the project level. The DPIA helps understand and identify the potential risks and importantly, makes sure the right mitigations are implemented.

Engage a trusted third party to act as data aggregator

Once the governance for the project has been agreed and set up, the objectives clearly defined, and the outcomes agreed, a process needs to be agreed about how and who will merge and analyse the data. Where multiple competitors are involved, it is recommended that the analysis and merging of the data needs is carried out by a third-party supplier. This might include someone who has no inherent commercial interest in the outcome of the research because they are not a provider and who may have more bandwidth and capabilities than the other participants within the project.

As no one organisation involved in The Pensions Data Project was in a suitable position to provide such a role, the Board decided to procure a third-party supplier who had the required skills, knowledge of pensions and previous experience of carrying out analysis on large data sets. PwC offered to carry this out on behalf of The Pensions Data Project.

Develop and agree a clear, explicit data scope so all parties are working to the same parameters

Having defined the research questions to be answered, the first challenge is to identify what data is needed. This might also include what data it would be useful to capture but are unable to due to the data not being available, (for The Pensions Data Project this included for example what industries people work in, what their earnings are).

The second challenge is to identify the scope of the members included in the data extract (i.e. the sample size) and specifically are there any members who should not be included to ensure compliance with the principle of data minimisation. In the case of The Pensions Data Project, it made no sense to include those who have £0 in their pot.

In addition, to complete a DPIA the scope of the data processing needs to be clearly documented, this includes documenting what the processing will cover for example:

- the nature of the personal data being shared;
- the volume and variety of the personal data;
- the sensitivity of the personal data;

- the extent and frequency of the processing;
- the duration of the processing;
- the number of data subjects involved; and
- the geographical area covered.

Any data scope should include explicit instructions and examples of how the data should be laid out and formatted

The data scope should include explicit instructions as to how to lay out the data. Having data in a consistent manner facilitates the smooth merging of it when multiple providers are involved.

It is imperative that the instructions are as explicit as possible, removing any possible mis-interpretations. The Pensions Data Project employed multiple test runs, including examples of how the data should be laid out, which improved the success of merging significantly.

To ensure the best possible matching of the data from multiple providers, clear instructions are essential to ensure consistency in how each data field is formatted. A key lesson from the testing is that providers accept and store data in varying formats. For example:

The infographic is a dark blue rectangular area containing several white and yellow boxes with text and icons. It is organized into several sections:

- Using all lower case whilst others use upper and lower case:** A box with a yellow circle containing the letters 'g', 'a', and 'k' in different cases.
- The use of diacritics (i.e., accents or cedillas):** A box with a yellow triangle containing the characters 'É', 'é', and 'đ'.
- Use of numbers and other special characters:** A box with a yellow hexagon containing the characters '!', '1', '4', and '?'.
- Recording dates of birth in the American format (MM/DD/YYYY) or English Format (DD/MM/YYYY):** A box with two flags, the American and British flags.
- Where no data is recorded, what happens?:** Two boxes: one says 'For some the approach is to leave the field blank' and the other says 'whereas others put 0 or NA or another variation'.
- Varied use of Punctuation:** A section with four yellow circles containing different date formats: '01011980', '01/01/1980', '1/1/1980', and '1-1-1980'.

It is crucial, therefore, that all permissible and not permissible data is clearly outlined, and explanations given as to what to do if data is not saved in that way in their databases.

Test, test and test again the data definitions and formatting with an initial trial project that is more limited in scope than the ultimate goal, including using both dummy and real data.

The data scope needs to be robustly tested. The Pensions Data Project tested its data scope three times using dummy data before moving to a sample of real data. The testing of dummy data, where known issues had been introduced, highlighted that nothing should be assumed and that every eventuality needed to be articulated and anticipated.

Establish the impact of a provider's privacy policy on format of data to be shared with the data processor

One of the key considerations, that any project needs to consider is what implications each provider's privacy policies and notices may have and in particular whether:

- 1) The data can be processed for research purposes in the first place,
- 2) Personal data can be shared with third parties for this purpose.

The format of the data to be shared with the data processor will be impacted by what has been communicated to members, in particular:

- Whether the personal identifiable data needs to be anonymised or not
- If required, what method will be used to make the data anonymous

Confirm whether data needs to be pseudonymised or anonymised

Definitions:

Anonymisation is the complete and irreversible removal of any information that could lead to an individual being identified, either from the removed information itself or this information combined with other data.

Pseudonymisation is the process of replacing the personal identifiable data with artificial identifiers or pseudonyms so that discovering the identity of an individual is not possible without additional data).

Depending on the specific data projection considerations that apply to each of the data owners, a decision needs to be made about whether the data shared with the data processor will be anonymised or at least pseudonymised.

As the data from the different providers in The Pensions Data Project needed to be merged with data from other providers, the processor needed to be able to tell which data from different providers related to the same data subjects, but in doing this, they did not need to be able to identify these subjects. As there was no requirement for them to be able to identify these subjects, it was agreed that data would need to be pseudonymised rather than anonymised to enable the data to be merged.

Initially, the project also wanted to understand the level of accuracy of each of the different provider's data, and how this was affecting the match rate. However, as one of the provider's privacy statements did not allow for analysis of their data by a third-party provider it was agreed that once merged, the hashed personal identifiable data would be removed and deleted thereby making the data, prior to any form of analysis, anonymous. This meant that only high-level analysis will be carried out on data quality.

If the data needs to be pseudonymised or anonymised, identify the optimal secure option

Having established whether the data needs to be anonymised or pseudonymised, the next step is to determine what method to use.

In the case of The Pensions Data Project, it was decided to pseudonymise the data by hashing the data.

Hashing is a method of one-way data anonymisation. This means a predetermined key cannot be used to unscramble the information. Unlike encryption, hashing typically serves as a checksum to ensure that a particular piece of data or a file hasn't been altered.

A hash is a mathematical computer process that takes information and turns it into letters and numbers of a certain length, making storing and finding information quicker because hashes are usually shorter and easier to find. It also makes information unreadable, and the original data can become confidential.

The Pensions Data Project will use a hashing technique called Secure Hash Algorithm 256 (SHA 256) and is a maths process that generates a 256 bit (64 character long) random sequence of letters and numbers (hash) out of any input). SHA 256 is considered as one of the most secure ways to protect digital information.

The implications of a capital instead of a lowercase letter

"I like bitcoin" when hashed equals:

ad3e58f21b94f32dcadca6b71df4c31a18179f38011551a17a80d0ff065d22c5

"I like Bitcoin" when hashed equals:

d988ca30eaa88c0410ad6e48a5297c0d505dcee572f9884f1a6fa2cbc8dedc86

The number of possible combinations of letters and numbers produced by SHA 256 exceeds the number grains of sand on Earth, which makes "guessing" the data hidden within the hash virtually impossible. Hashes cannot be reversed, so the process is typically used for computer security.

Resolve the process and protocols merging and storing data

In order to comply with each provider's data compliance regulations and security procedures and privacy statements, the merging and storing of the data also needs to be established, including understanding and agreeing the following:

- Where will data be stored;
- How data will be matched;
- Who will have access to the merged data; and
- How long data will be held for before being destroyed.

Due to the confidential nature of the data, it is important to understand and agree where data will be stored and what security procedures will be in place to prevent unauthorised access.

It was agreed that the data received by PwC via Mft2Go would be dropped into DASH (Dash is a user interface which allows large files to be securely copied across the internet into folders which can then be used for creating analytical web applications) where the data will then be amalgamated.

Ensure suitable consideration is given on how data will be matched

There are likely to be a number of items upon which data can be matched, including date of birth, national insurance number (NINO), surname and address, and the level of accuracy required.

As data for The Pensions Data Project is being shared in hashed form, merging needs to generate the most accurate matches and therefore it was decided to primarily match on NINO and date of birth. Both of these fields have been found to be the most accurate as evidence suggests surnames and addresses are most likely to not have been updated.

Additionally, many providers find that women generally do not update them when they change their names (for example due to marriage) and few members update providers when they change addresses, especially if they are no longer working for the company who set up that specific pension.

It was decided that including full postal codes in the matching would run the risk of potentially identifying people and that the data could no longer be seen as anonymous. Therefore, only the first part of the post code will be provided by providers and will only be used in the research analysis to identify any regional differences in the pension information.

If possible, it is recommended that data is shared unhashed in order to maximise the quality of the matching. Where data is provided unhashed there is an opportunity to use fuzzy matching techniques to improve the quality of the matching (fuzzy matching identifies the likelihood that two records are a true match based on whether they agree or disagree on agreed various identifiers).

To provide some evaluation on the level of data accuracy within The Pensions Data Project, once the matching process on date of birth and NINO has taken place, further analysis will be carried out to see what level of match would have occurred if surname had also been included in the process.

Determine a secure and compliant way to transfer data

There are multiple ways to transfer data in a secure manner, such as using external media, e.g., managed file transfer, email, USB drive etc. In the case of The Pensions Data Project, the most secure method was felt to be by using PwC's managed file transfer process (Mft2GO). Other methods were felt to be less secure.

A Managed file Transfer (MFT) is a technology platform that allows organisations to reliably exchange electronic data between systems and people in a secure way to meet compliance needs, including having algorithms and encryption to secure data, alongside IDs, passwords and SSH keys to verify authenticity.

In the case of The Pensions Data Project nominated individuals from each of the providers will be emailed an invitation giving them access to this platform. Only users emailed with an invitation will have access to this platform and similarly only the individual sending the request to share data will be able to view the shared files.

Agree and document for how long and where the raw and merged data will be held before being destroyed

In line with applicable Data Protection Legislation, formalised written agreement should indicate how long data should be held for and any mitigating circumstances where data can be held for longer. This will also prevent unauthorised or unnecessary access to both the raw, pseudonymised data and also any amalgamated data.

In the case of The Pensions Data Project, the raw pseudonymised data will be kept until the data had been successfully merged, with data being deleted from DASH and Mft2Go within 1 week. The merged data set, on which analysis will be undertaken by PwC, will be retained for a maximum of 12 months after publication of the analysis, or as long as is required for replication of the results and / or for regulatory purposes.

How the people involved in the project can make all the difference to its success

Areas that need to be considered include:

- Agree project goals, objectives and outcomes in line with competition law safeguards
- Build trust and confidence
- Patience
- Ensure resources are available
- Learn and share
- Communication

Agree clear, common project goals, objectives, behaviours and outcomes, which are defined and documented

Discussing and agreeing the purpose of the project, what the desired outcomes are and how the participants will work together sets the necessary framework to encourage the appropriate non-competitive behaviours.

Allow time and space to build trust and confidence and to establish ways of working between the project participants

There are usually significant cultural barriers to transparency around data and information. Competitors do not usually share information and thoughts with others. Investing time to build trust and confidence in working together reaps returns at points in the project where the going can get challenging.

Be patient. Be prepared to go through multiple iterations to achieve the desired outcome

The novelty of The Pensions Data Project means that no precedent has been set and we have worked without any best practice guidance available. The project team is learning as it goes, with many aspects being unforeseen or unknown. Having spent time building a team means that such challenges can be faced head-on and worked through in a collaborative and positive way.

Ensure that the necessary resources are in place in all participants, ahead of being needed, specifically from the data, legal and policy team

The Pensions Data Project team have competing work priorities. Getting the necessary time and resource therefore requires good-will and commitment to work which are for the overall good for future policy making.

Throughout The Pensions Data Project different disciplines have been required, specifically:

- **Policy;** who understand the importance of the project and how it will enable evidence based policy making in the future.
- **Data;** who understand how each participating organisation functions, how it manages its data and how its operations and systems work. Their role has been to enable discussion on the data and process practicalities.
- **Legal;** Creating the necessary governance between the participants in the project and between them and PwC has required significant resources to prepare, review and sign off the necessary agreements and documents.

Learn from the process, learn from others and share lessons

The Pensions Data Project has endeavoured to learn from the various activities and steps in the project process. Depending on resources, participants have been able to carry out some elements prior to others. Sharing these experiences and lessons enables other to avoid making the same mistakes and speeds up the process.

There is no such thing as over-communication

Communication within the project and among external stakeholders is key to keeping the momentum going and ensuring everyone is on the same page.

Final thoughts

- You never know everything you'll need when you're dealing with a ground-breaking project or even one which is seeking to do novel things with existing data.
- Getting the right people in the room is essential as a little good-will goes a long way. Having experts helps you not go round in circles.
- Tolerance and a sense of humour are essential to see you through the tougher times.

³ Wealth and Assets Survey QMI - Office for National Statistics (ons.gov.uk)

⁴ Annual Survey of Hours and Earnings (ASHE) - Office for National Statistics (ons.gov.uk)

Appendix: The Pensions Data Project

Background

The development of public pensions policy in Britain and overseas is increasingly being informed by evidence-based research. However, currently in the UK there is no central longitudinal research base of people's total retirement savings. This exists in other countries (such as the USA) and is a powerful tool for evidencing how individual citizens' retirement savings, aggregated across their different schemes and providers, are evolving over time.

Discussions about establishing such a research database for the UK have taken place for many years across Government and the pensions industry. The Pensions Commission identified this data deficiency for making evidence-based policy in their First Report in November 2004. More recently, discussions about addressing this need for evidence have been led by the Pensions Policy Institute (PPI).

What is the purpose of the project?

The project will amalgamate data across multiple providers in order to get insight into people's savings patterns. It will provide the industry with a central longitudinal research database of people's total retirement savings.

Initially, it will aggregate across the Defined Contribution (DC) pots (trust) which individuals have with different providers, thus generating insights for both pension providers and Government.

In the longer term, there will be an opportunity to provide other industry-wide insights on the adequacy of individuals' total retirement savings and their resulting incomes by linking to:

- Defined Contribution (DC) contract based
- Defined Benefit (DB) scheme data
- Decumulation data
- Other data sets, such as the Annual Survey for Hours and Earnings (ASHE)

Why is the project needed?

Although there is the Wealth and Assets Survey³ and the Annual Survey of Hours and Earnings (ASHE)⁴, there is currently no single combined data set that shows us how people are saving for their retirement.

These surveys, along with the ONS produce aggregate statistics and individual modelling, but do not answer any of the following types of questions:

- o How many pension pots do people have?
- o How often do they move provider?
- o How much is their pension wealth?
- o How does this change over time?

The Pensions Data Project seeks to answer all these questions. It will also help to provide additional evidence to inform policy making.

Proof of concept phase

Over the 2019 Summer, a Proof of Concept (PoC), using realistic, but entirely fictitious, DC pensions data from several unidentifiable providers explored the different technical options for a) pseudonymising/encrypting personal information (so that individuals aren't identifiable), and b) combining the encrypted data. The exercise proved that the aims of the project could be achieved technically.

The Private Beta phase

Since 2019 a Private Beta phase has been undertaken. This has continued to build on the proof of concept, with the aim of:

- Assessing the feasibility of such a cross-industry data sharing exercise and;
- Assessing the prevalence of smaller pension pots among certain groups of UK workers.

Who is involved in the Private Beta phase

The Private Beta Phase is being controlled by a small group of master trusts and the PPI who are all committed to evidence-based policy making and further research into long-term savings and are happy to take part in this initial phase. Participants include B&CE provider of The People's Pension, Legal & General, Nest Insight the public benefit research and innovation centre set up by Nest, Now: Pensions, Pensions Policy Institute and Smart Pension.

The Private Beta Phase is being funded by the Association of British Insurers (ABI), the Department for Work and Pensions (DWP), the Pensions and Lifetime Savings Associate (PLSA) and The Pensions Regulator (TPR).

How will it work

The planned approach is for each provider to transmit pseudonymised data using SHA 256 from their DC book to a trusted central repository (for the Private Beta phase this will be PwC). The individual records will be merged and aggregated. Poor / bad data, where possible, will be identified and highlighted to the relevant provider. Analysis of the data will be undertaken subject to adherence to a data access policy.

Analysis of this combined dataset (by policymakers, academics & other research partners), will provide vital inputs to inform the long-term development of retirement saving (and wider) policy.

The Pensions Data Project has different aims and serves a very different purpose to the Pensions Dashboard initiative: dashboards is a digital interface that will show one individual their lifetime pensions savings in one place (if they ask), whereas The Pensions Data Project will aggregate individuals' pension pots and provide a greater understanding of the overall pensions landscape at an aggregate level which will inform policymakers about pension savings at a societal level.

Acknowledgements and Contact Details

This paper has been written by Nicky Day, Project Lead, The Pensions Data Project.

About The Pensions Data Project

The Pensions Data Project is an exciting new pensions research initiative, managed independently on behalf of the entire UK pensions industry by a small group of master trusts and the Pensions Policy Institute (PPI) who shares a common goal of wanting to contribute to a wider societal benefit where everyone has better provision and can achieve a positive outcome in retirement. The crucial new facet, which does not currently exist anywhere else, is the ability to link across the various pension pots which individuals have with different providers, thus generating unprecedented levels of insight for both pension providers and Government

Participants in The Pensions Data Project include: B&CE provider of The People's Pension, Legal & General, Nest Insight the public benefit research and innovation centre set up by Nest, Now:Pensions, Pensions Policy Institute and Smart Pension, as well as PwC who will be processing the data on behalf of The Pensions Data Project.



The Pensions Data Project Private Beta Phase has been kindly sponsored by a consortium of funders, The Association of British Insurers (ABI), the Department for Work and Pensions (DWP), the Pensions and Lifetime Savings Association (PLSA) and The Pensions Regulator (TPR).

Sponsorship has been given to help fund the research, and does not necessarily imply agreement with, or support for, the analysis or findings from the Project.

How and why to be involved

There are many benefits to being involved in this Project. We are looking for providers, funders and other interested parties to participate in developing further the first cross-industry pensions dataset. For further information please contact:

Nicky Day, Project Lead

The Pensions Data Project

Email: info@pensionsdatapoint.org.uk

Website: www.pensionsdatapoint.org.uk

