



**Summary Report of the EURIM Personal Identity Group meeting,
1400 – 1600 hours, 23 February 2006, Room 'C' 1 Parliament Street, Westminster**

Chair: Stephen Darvill (LogicaCMG); Rapporteur: Dave Wright (EURIM)

SUMMARY OF MAIN POINTS

1. The Work and Pensions Longitudinal Study replaces a number of disparate systems that produced a knowledge gap because they did not talk to each other. It can be defined as a longitudinal, spells-based administrative database that allows a single client history view and inter-client group longitudinal analysis.
2. Data sharing provisions introduced in the Employment Act 2002 allowed DWP to receive employment data from HMRC for use in tackling fraud and error. DWP and HMRC have since cooperated to legislate for a data sharing initiative which enabled DWP to use HMRC employment data for research, statistical, performance, training, fraud prevention and verification purposes. However, WPLS data can currently only be used for DWP and related purposes; any developments leading to requests to share data externally would have to be discussed when and if they arise.
3. The highly sensitive data held within the WPLS on secure servers is subject to stringent safeguards and security measures, and can be used only for specific named purposes as defined in legal and permissive gateways. Access is restricted to individual business case applications, and while identifiable information is used for matching purposes, personal data is anonymized in almost all cases for end usage.
4. The WPLS is instrumental in shaping DWP thinking on the better use of customer information to improve service delivery, and the potential for its use in forward strategy thinking is a priority for consideration. Practical use and experience are necessary for realistic testing of systems: DWP uses customer experience and feedback from those people responsible for inputting data to the system to improve data quality both internally and through external suppliers.
5. The WPLS dataset is a significant resource, and much modelling through multivariate analysis has been applied to look at specific issues, both in terms of strategy and looking at combinations of characteristics or similarities in customer segmentation.
6. DWP is in discussion with other Government departments on a number of initiatives, in which legislative clarification is being sought. What could be achieved or enhanced by data sharing is very much a part of future strategy, and developments are expected once the Ministerial Committee on Data Sharing is functioning.
7. A EURIM planning meeting on March 10 envisages a 'showcase' event for MPs in support of the 'Transformational Government' strategy. The idea has received significant industry support, with the aim of showing successful working examples of IT-enabled transformation projects that can inform MPs and allow them to question those involved.
8. A meeting to discuss aspects of EU interoperability and shared services is planned for April.

1. Welcome and Introduction

1.1 Members and observers were welcomed to the meeting, with introductions around the table.

2. Work and Pensions Longitudinal Study – Neil McIvor

2.1 The WPLS can be defined as a longitudinal, spells-based administrative database. 'Longitudinal' refers to the benefit histories of individuals over time, while 'spells' refers to the specific periods of benefit receipt or employment of an individual that are held as a separate record in the database. This allows a single client history view and also inter-client group longitudinal analysis.

2.2 WPLS was introduced in January 2004, and provides, for every single DWP client, information on the initial and subsequent benefit or pension received. Interventions (e.g. assistance, training) and their consequences are also recorded, together with personal details (name, address etc.) and income, savings and investments. Housing and council tax benefit data are currently being integrated into the scheme.

2.3 The WPLS was preceded by a number of disparate systems that produced a knowledge gap because they did not talk to each other, and there was no record of a client's long-term employment or benefit profile. WPLS provides the opportunity to improve the analytical evidence base, operational effectiveness, understanding of the long-term effectiveness of the strategies employed, and the targeting of resources.

2.4 Details of the contents of the WPLS database are listed in Slides 6 and 7 of the presentation. Once details for a client are listed they are held for reference, so that any subsequent work experiences are recorded and can be tracked. The volume of data has led to the creation of a 'long but thin' database, with personal identifiers allocated that allow back-linking to source data to retrieve more information, e.g. for analytical purposes. Identifiable information is used for matching purposes, but is anonymized in almost all cases for end usage.

2.5 The WPLS holds around 143 million separate linked employment spells, with a further 146m HMRC employment spells (all except the HMRC's most secure and sensitive cases) unlinked as yet to DWP clients. Access to the unlinked database by officials is allowed only for very restricted purposes, tied to the longitudinal element of the database. For example, consider an individual losing their job this week and signing on to job seekers' allowance. Once the JSA spell had fed through to its end-product, linking back to the previous employment record will provide the full individual history (back to 1998).

2.6 Data held in any administrative database are unlikely to be 100% accurate, and data quality is an issue for the WPLS. Better data quality can be achieved by completing the loop with those people responsible for inputting data to the system. This might be enhanced by financial incentives to organisations supplying the data, or performance measures. DWP has control over its own data, but there are effective routines in place with HMRC.

2.7 Development and design of systems can achieve so much, but practical use and experience are necessary for realistic testing. DWP uses customer experience and feedback from its customers to improve data quality both internally and through external suppliers.

2.8 The data sharing provisions introduced in Employment Act 2002 opened the way for DWP to receive employment data from HMRC. Initially, this was 'snapshot' data for use in tackling fraud and error, but DWP and HMRC have worked together to legislate for a data sharing initiative. This was followed by amendments to Social Security (1992 and 1998) regulations, which enabled DWP to use HMRC employment data for research, statistical, performance, training, fraud prevention and verification purposes.

2.9 The gateways under the 1992 and 1998 Social Security Acts allow savings data to be used for fraud prevention and verification. The new Tax Credits Act (2002) permitted the use of tax credit information for research and statistical purposes, while the Pensions Act (2004) gave new powers and functions to facilitate retirement planning, enabling DWP to extend the use of savings data as an exclusion in pension credit take-up campaigns.

2.10 In addition to legislative powers, permissive gateways were also agreed with HMRC. All major DWP benefit programmes include consent forms for users to permit linking data as a condition of

receiving benefit. This allows DWP to link data with customers, and informed consent to share data is starting to be requested for customer surveys on an opt-in basis.

2.11 DWP must also observe data protection legislation and in particular Article 8 of the Human Rights Act, whenever they wish to extend the uses of sharing data, especially with respect to proportionality and foreseeability (ensuring the customer understands the occasions on which DWP will share their data). The uses to which specific data is put is also listed on the DWP website.

2.12 WPLS uses a combination of Oracle applications with SAS intelligence and analytics processes to develop the data within a UNIX environment, and has the flexibility to utilise this for initiatives within the Information Directorate, isolated from the wider departmental data. The development of fuzzy matching algorithms allows links between internal and external data, but it is important to understand the source data for capture, internal manipulation and extraction to WPLS. Internal manipulation and cleansing of data is undertaken for aggregated analytical purposes within the Information Directorate team. Uncleansed data is used for fraud investigations.

2.13 WPLS was developed primarily for research and statistics as well as management information; accordingly, there is the potential to support other services. One area is the provision of research and statistics on the success of Jobcentre Plus, whose analysts are using the WPLS to look at customer segmentation, e.g. whether a particular type of customer history can be linked to likelihood of gaining employment with or without interventions, thus allowing focus on those people who need more help.

2.14 WPLS data can help to evaluate individual policies in the short term, but more importantly also in the medium and long term. This applies equally to fraud, e.g. does an individual having an incorrectness in a claim make it likely that this will recur in the future? If so, DWP can better target resources and investigative efforts, as it is much easier to exclude people from interventions than to include them. WPLS has also been used in a targeted pension credit take-up campaign, eliminating those ineligible or already claiming benefit, and improving efficiency from the previous 'scattergun' approach and improving customer experience.

2.15 Slide 12 represents an individual's profile, created from data held on the WPLS database. This person has claimed income support since January 1999. A 'new deal' intervention in 1999 failed, but in early 2002, another new deal intervention, combined with basic skills screening, resulted in employment. This data, when aggregated with other cases, may yield useful patterns that result from various initiatives that can further improve the quality of interventions.

2.16 The highly sensitive data held within the WPLS is subject to stringent safeguards and security measures, and can be used only for specific named purposes as defined in the legal and permissive gateways. Access is restricted to individual business case applications, with data held on secure servers. The data owner has to clear the business case if the application is for anonymized information; more sensitive identifiable information has to be cleared by both the Head of Data Services in the Directorate, and the Ethics Committee.

2.17 All users accessing WPLS data must present evidence of attendance at data protection seminars, and make a business case for the need to use the data, which is subject to monitoring by a team which ensures observation of protocols. The Ethics Committee comprises statistics and research experts, which advises on and monitors the use of the data, and also monitors users through trace files and audit trails.

2.18 WPLS data (240 million records, 40 GB) sits in a secure server, and any subsets of data generated by analysts must be stored in that secure environment. Protocols govern this procedure, e.g. filenames must be prefixed by the log-in name, so that all access to data subsets can be monitored to ensure that data is used only by authorised people for legitimate purposes.

2.19 The Information Commissioner, who can perform an audit at any time, is informed of all developments of the WPLS. Enterprise Privacy Group (<http://www.privacygroup.org>) were proactively invited to pre-audit the WPLS system to ensure standards were being met, and their recommendations subsequently acted upon (e.g. the development team should sit in areas protected by swipe card access, to prevent unauthorised staff etc. from obtaining information by proximity).

2.20 Data retention policy allows employment, benefit and earnings information on customers to be held for 25 years, in anonymized format. For data relating to interest on savings for those aged 60 and

over, data can be held for 5 years. Data subsets generated for operational purposes are destroyed after 1 year.

2.21 Customer data is obviously a key part of the development and implementation of the Information Directorate strategy. The WPLS is instrumental in shaping DWP thinking on the better use of customer information to improve service delivery. However, the use of the WPLS beyond the management of information, statistics and research is yet to be determined, but the potential for use in e.g. forward strategy thinking is a priority for consideration.

Questions and Answers

2.22 What is meant by the term 'anonymized'.

Certain personal details (e.g. name) are removed; post code and sector are retained for analytical purposes. For ad hoc purposes, rounding policies are applied, with cross-tabulation across variables where identification of specific target groups is difficult. The NINO is encrypted, and a DWP unique identifier added that links to HMRC records, and can be used to back-link to source data.

2.23 To what extent is distinguishable, as opposed to identifiable data, able to serve?

Distinguishable data is almost always used in the WPLS, and protocols are applied to the conversion of identifiable to distinguishable data.

2.24 Was data cleansing undertaken for the WPLS database?

Data cleansing was applied to non-evidential data, e.g. instances of overlapping claims, but not to externally-sourced data.

2.25 Is WPLS data exported to HMRC?

Rules of non-disclosure are in place, and if any external data, e.g. from HMRC, is manipulated in any way through the Directorate, this cannot be sent outside, although DWP can act as an agent for HMRC to supply data to authorised requestors.

2.26 The WPLS dataset is a significant resource for multivariate analysis; what work has been done in this field?

Analysts look at specific issues, and a lot of modelling work is being undertaken, both in terms of strategy and looking at combinations of characteristics or similarities in customer segmentation e.g. the type of person in which incorrectness appears, or indicators of behaviour and success in obtaining employment.

2.27 Might WPLS be extended to include other types of data, e.g. from the private sector?

WPLS is very much the customer-driven, and the data held is normally fit for purpose. If a business case was made for additional data, and an appropriate legislative gateway existed, then it could be used. The Directorate would consider whether legislation was appropriate in specific cases, e.g. Fraud and Error work involving information from the private sector; there is also an initiative under way looking at data from credit reference agencies for Fraud and Error purposes.

2.28 Are there any areas where DWP might want to augment data, but is prevented from doing so by legislation?

DWP is in discussion with other Government departments on a number of initiatives, in which legislative clarification is being sought. What could be achieved or enhanced by data sharing is very much a part of future strategy, and developments are expected once a Cabinet Office Committee was able to make an announcement.

2.29 How open is DWP to another agency requesting information from the WPLS database?

Currently, WPLS data can only be used for DWP purposes, or a joint Government initiative where DWP has an interest. Local Government requests are not be entertained unless there is a direct DWP interest.

2.30 What encryption was used?

Encryption is applied to the DWP unique identifier, and to the NINO by DWP-generated algorithms that allow linkage between an encrypted NINO of one data source to another.

2.31 How open is DWP to external sources of information, e.g. from the public reporting suspected cases of benefit fraud by individuals they knew, and would this appear on the WPLS?

Hotlines are provided and any information supplied is logged. Risk-scoring might be used in assessment before referral to frontline fraud investigators. A programme protection database linked to the WPLS holds data on referrals and interventions, and subdivision of the variables in different source databases looking at e.g. benefit fraud, or hotline referrals can be linked individually or in combination to the WPLS.

2.32 Who manages the WPLS servers and data, are they visible from the office, and if not, how can you know how secure is the bulk of the data?

Servers are located and managed in Newcastle by Directorate IT teams. As part of the Directorate, the teams are involved in the entire security system and the protocols and measures surrounding this.

2.33 How does the WPLS fit in with the CIS (Customer Information System) development project?

These are currently separate areas; the Directorate is modernising its feeds with a view to using CIS to improve the quality of information held within the WPLS.

2.24 Is child benefit data stored on the WPLS?

Concerns around consent have led DWP to avoid recording child benefit data on the WPLS, even though it is held within DWP. This also applies to CSA data, but projects are afoot to see if legal gateways are available to better use CSA data with other relevant data in WPLS.

2.25 Does the Directorate monitor outcomes of business case applications?

Monitoring is not undertaken directly, but because it is used by a relatively small customer base, user fora have been convened to discuss the issues involved, and this provides unofficial feedback to the Directorate.

2.26 What policy developments had been triggered by statistical analyses of the WPLS data?

A great deal of input goes into the fraud and error fields, but no specific information is available about this. The data had really only been used 'in earnest' since early 2005.

2.27 Is the introduction of identity cards and the National Identification Register likely to lead to requests for data from other departments and agencies that might compromise the integrity of the WPLS?

WPLS data can currently only be used for DWP and related purposes. Any developments leading to requests to share data externally would have to be discussed when and if they arise.

2.28 Can you provide any more information about the Cabinet Committee on Data Sharing referred to earlier?

Dave Wright informed that this was the Ministerial Committee on Data Sharing (known as MISC31), which is to be chaired by the Chancellor of the Duchy of Lancaster, when and if this appointment is made.

3. Recent Developments – Dave Wright

3.1 There was little to report on ID cards, where there had been a hiatus in communication. The planned meeting between Margaret Moran MP and Baroness Ashton at DCA is set to follow the first meeting of MISC31, to be chaired by the Chancellor of the Duchy of Lancaster – but this appointment is still awaited.

3.2 The Data Sharing Subgroup is holding a planning meeting on March 10. This envisages a 'showcase' event for MPs in support of the 'Transformational Government' strategy, to be held e.g. in the House of Commons corridor. The idea has received significant industry support, with the aim of showing successful working examples of IT-enabled transformation projects that can inform MPs and allow them to question those involved. Details of the planning meeting had been circulated to the EURIM Personal Identity Group list, and members are encouraged to attend.

3.3 Responses to an email circulated to the EURIM Personal Identity Group on the issue of shared services and cross-border interoperability at EU level have suggested that an exercise may be worthwhile. Further action is likely to be dependent on the industry being able both to take a pan-European position and identify clear cases for interoperability solutions that will create future business. A major aim will be to develop realistic routes to achieve international interoperability: there are concerns that unworkable interoperability rules will result that are not supported by the industry. EURIM may be able to help with any formal response to the Commission document by eGU in relation to eGIF. Invitations will be sent to David Myers (Head of Shared Services, Cabinet Office) and Ilias Chantzos (Auxilliary Administrator for the European Commission), with a view to a meeting in April.

3.4 It was agreed to circulate the current version of data sharing grid to members, and invite further contributions.

4. Date of Next Meeting

4.1 It was agreed to set a date for the next meeting of the full group in response to developments.

4.2 The Data Sharing Subgroup is holding a planning meeting on March 10 (details circulated).

4.3 A meeting to discuss aspects of EU interoperability and shared services is planned for April.