

Issues on the Evaluation of Revised Employment Action Plan/NEES

Professor Colm Harmon (UCD School of Economics)

Paul Morrin (DSP)

Dr Sean Murphy (UCD School of Computer Science)

HIGHLIGHTS	3
1. INTRODUCTION.....	4
2. DESCRIPTION OF CURRENT INTERVENTION PROCESS MODEL.....	4
3. IDENTIFICATION OF EXISTING DATA STREAMS	5
3.1 CURRENT DATA COLLECTION CAPACITY	5
3.2 REPORTING CAPACITY	8
3.3 EVALUATIVE CAPACITY.....	10
4. RECOMMENDATIONS ON INFORMATION AND DATA SYSTEM REQUIREMENTS.....	11
4.1 STRUCTURE.....	11
4.2 EASE OF ACCESS.....	12
4.3 LINKAGES TO EXTERNAL DATABASES.....	13
4.4 DESIGN FOR EVALUATION	13
5. ENHANCED REPORTING FOR DSP.....	14
5.1 NUMBERS PROCESSED WEEKLY/MONTHLY/YTD WITH COMPARATORS	14
5.2 NUMBERS IN ALMP	15
5.3 LENGTH OF TIME SINCE LAST EMPLOYMENT.....	15
5.4 TYPE OF PREVIOUS EMPLOYMENT.....	16
5.5 LENGTH OF TIME SINCE LAST INTERVENTION	17
5.6 QUEUE LENGTH FOR TYPES OF INTERVENTION	17
5.7 NUMBERS OF SANCTIONS	18
6. ADDITIONAL ISSUES.....	19
6.1 ENHANCED REPORTING MECHANISMS.....	19
6.2 ENHANCED CAPACITY FOR DATA ANALYSIS.....	19
6.3 DATA COLLECTION.....	19
REFERENCES.....	21
APPENDIX 1 – TERMS OF REFERENCE	22
APPENDIX 2 – NOTE ON EVALUATION DESIGN	23
APPENDIX 3 – SOURCES OF DATA INFORMATION	25
A3.1 LIST OF INFORMATION SOURCES.....	25
APPENDIX 4 - A SHORT NOTE ON PERFORMANCE OF DATABASE MANIPULATIONS FOR ANALYSIS.....	27
A4.1 EXPERIMENTAL PLATFORM.....	27
A4.2 EXPERIMENTS.....	28
A4.2.1 <i>Experiment set 1 – adding records to a database.....</i>	<i>28</i>
A4.2.2 <i>Experiment set 2.....</i>	<i>29</i>
A4.3 CONCLUSION FROM THESE SIMPLE EXPERIMENTS	29
APPENDIX 5 – SUMMARY CHART FOR PROSPECTIVE EVALUATION	31

Highlights

- Overall the impression is that the DSP collect a very sizable volume of data, and has made significant progress in developing linkages between their data holdings and other sources within the public sector, including Revenue and Education. The data from FAS will fall under the structure of the DSP adding an additional data stream.
- However, while the PEX profiling process is a major step forward in terms of the information gathered, some critical information points (education levels, reasons for signing off, destination of employment if you have found a job etc) are either missing, not gathered systematically by any source, or may only be available from secondary sources after a sizable delay.
- DSP's IT systems have evolved over time, responding to different requirements and different capabilities of the platforms they use, which inevitably results in information being located in different locations and in different formats; further, in some cases, there is limited amount of managed duplication of information in different systems. Moreover the infrastructure is largely aimed at the provision of a claimant system rather than for the production of management information.
- Reporting is largely 'point in time' instead of an episodic, longitudinal, process. There are specific difficulties in linking current claim data to previous claim or work histories due to the lack of a 'single client' approach to the data. We would recommend restructuring of the claimant analysis database to address this.
- Due largely to the fact that the IT systems role in the management of the claimant system is paramount, there has been little attention given to the access points for extracting data from across all available data sources. This is a relatively easy issue to address ideally against some agreed 'pro forma' set of data requests for standard management reporting and a joint business/IT project will progress this on foot of the issues raised. Section 5 of the memo has some specific recommendations regarding the potential of the current data infrastructure in reporting a number of key metrics.
- Over and above reporting requirements and critical data gaps, the DSP reporting infrastructure is not currently optimized to produce the sort of data outputs required for large-scale programme evaluation. These evaluations typically require access to large volumes of data at a claimant level and the production of these files would appear to place a considerable delay in evaluation exercises. This needs to be addressed in the context of the overall review of reporting infrastructure and recent evaluators (such as the ESRI) may be able to share their experiences vis a vis their other evaluations.
- There appears to be some blockages in the pipeline between the IT/Data divisions of the DSP and the research/analysis divisions. The root of this would seem to be the need to protect the integrity of the system for claimant purposes but a more systematic working agreement between the internal divisions would seem an appropriate target for DSP Senior Management. Moreover the analytical capacity is under-resourced and investment in both people and training would be recommended – for example the scope for internal multivariate analysis of data is very limited.

1. Introduction

During December 2011, UCD was asked to provide a view on how the Department of Social Protection (DSP) manages the Live Register (LR). The main objectives of the short study were to understand how DSP collects data relating to the labour market and how it uses same to understand the dynamics within the market and to influence policy mechanisms. The study offers some recommendations on how to improve data collection and ways to use the data more effectively, both in regular reporting and in the context of longer-term evaluation of particular activation schemes.

The Terms of Reference (ToR) of the work are included in Appendix 1.

This structure of this document largely follows the issues outlined in the ToR. In section 2 there is a discussion of the current intervention process model. This is followed in section 3 by a discussion of the existing data infrastructure, including some commentary on the current reporting capabilities and capacity to evaluate activation programmes. Next, in section 4 there are some remarks on where the data infrastructure can be improved to perform more useful/effective reporting and analysis. Finally in section 5 some additional observations on workflow and gaps are made.

2. Description of Current Intervention Process Model

From 'Pathways to Work'; through regular and on-going engagement with people on the Live Register, access will be provided to opportunities for work, education and training. 'Pathways to Work' is based on five strands:

- more regular and on-going engagement with the unemployed
- greater targeting of activation places and opportunities
- incentivising the take-up of opportunities
- incentivising employers to provide more jobs for people who are unemployed, and
- reforming institutions to deliver better services to the unemployed.

Through implementation of these five strands, people who are unemployed will be given the opportunity to acquire suitable skills and/or qualifications. The following targets for achievement by 2015 are set out in the 'Pathways' document;

- (i) For each person currently unemployed

The Department of Social Protection will engage with and provide supports (such as referral to education, training, work placement, or job search assistance) to each person currently unemployed to increase their prospects of securing a job. There will be a particular focus on the long-term unemployed with a view to ensuring that at least 75,000 of those currently long-term unemployed will move into employment by 2015. This is more than double the current rate.

- (ii) For each person who becomes unemployed in the future

The Department of Social Protection will engage with each person who becomes unemployed in the future with a view to reducing the average time spent on the Live Register from 21 months today to less than 12 months by the end of 2015.

(iii) For Employers

As part of this 'Pathways to Work' approach there will be more meaningful engagement between the Department of Social Protection and employers so as to ensure that employers have access to and are offered, on request, candidates to fill full-time vacancies.

The Department of Social Protection will increase the proportion of all jobs that are filled with candidates from the Live Register via its employment services to a level which meets international best practice. Appropriate research and analysis will be undertaken during the second quarter of 2012 to establish robust targets and measurement metrics in line with this commitment.

In advance of this analysis, an interim target will be set of increasing the proportion of vacancies filled by the Department's employment services from the Live Register to 40% by 2015. This target will be reviewed once the relevant analysis has been completed.

(iv) For the Tax-Payer

To ensure that each person in receipt of job-seeker payment fulfils their personal responsibility to engage fully with the employment and training supports provided by the State, as a pre-condition for receipt of their welfare payments.

This will contribute to the Department of Social Protection meeting its target of achieving control savings on expenditure on welfare payments of which a significant proportion is expected to be in respect of working age payments.

3. Identification of Existing Data Streams

There are two core, related but in part distinct, demands from the analysis and interrogation of the data holdings – firstly, to provide timely, informative and more forensic analysis of data on an ongoing, routine basis to obtain a better understanding of the operation of the current systems; and, secondly, to provide an infrastructure for evaluation of activation programmes.

The Department has a sizable volume of the data required to carry out systematic reporting and evaluation. Much of this is currently held across multiple systems and data extracts; ISTS extracts of live and closed claims, Revenue data (COE, P35, self employed incomes), Central Records, PEX data, DSP and FAS EAP/APCM data and BOMi data for mainly long term schemes. Having the data dispersed over this multitude of systems – for which there are good historical reasons – does give rise to some issues when new analyses are required, which will be discussed further below.

3.1 Current Data Collection Capacity

We have examined detailed information on the data sources within the DSP and related data flows into and out of DSP and cognate Departments and agencies such as the Revenue Commissioners. The following table gives a simple overview of core data availability.

Data item	Data available	Data Location/Source	Comment/Observations
Demographic; age, sex, marital status, nationality	Yes	CRS	Most of this data is held in a 'point in time' format.
Previous work history	Yes	Revenue data	This data is not held for analysis in an episodic or longitudinal (multiple observations on the same individual over time) format, which is more desirable.
Sector of employment	Yes	Revenue data	Ditto
Previous welfare history	Yes	CRS	Ditto
Duration on benefits	Yes	CRS	Ditto
Previous earnings	Yes	Revenue data/PEX question (CRS)	Revenue as above; PEX data has limited availability
Previous welfare incomes	No		Not available in a readily accessible source.
Claim closures/reasons	Yes	ISTS	Accessible.
Persistence of closure	Yes	ISTS	Not available in a readily accessible source.
Level of education	Yes	PEX question (CRS)	Limited availability in a systematic format except for Welfare Offices participating in the 'PEX' profiling pilots.
Level of SW entitlement	Yes		Accessible

The Department recognizes the need to hold this data in a suitable reporting environment for analysis and a joint business/IT project will be set up to progress this requirement on foot of the issues raised in this report.

The primary data source available to the department in this context is the large set of ISTS data arising from the ISTS system. This is mostly a transaction-based record of payments to individuals covering a number of different types of payments including LR payments (of different types). It is worth noting that the design of this system focused on the processing associated with managing payments to large amounts of individuals rather than collecting data for the purposes of analysis of the labour market; that said, the data contains much useful information from which more insight relating to operations of DSP and its 'customers' can be gleaned.

Apart from the ISTS data, it is worth noting the following sources of information:

- ⤴ Information from Revenue: There is good information flow between DSP and Revenue relating to employment. The primary sources of information are a Commencement of Employment notification (when Revenue has been notified by an employer that an individual has taken up a new employment) and information from (employer) P35s which indicates how much an individual has worked within a given year. This information is stored on DSP's IT systems and is available for analysis.
- ⤴ Information from FAS: Communications with FAS is largely in the DSP→FAS direction, but DSP does receive an important feedback signal that indicates if an individual did not present at FAS for interview.
- ⤴ PEX profiling information: DSP has recently started collecting information that is more detailed on individuals in order to better predict their suitability for activation schemes. It makes sense to use this data for the purposes of evaluation of such schemes

DSP's IT systems have evolved over time, responding to different requirements and different capabilities of the platforms they use. As is natural in such situations, this has resulted in information being located in different locations and in different formats; further, in some cases, there is limited amount of managed duplication of information in different systems. For the purposes of both reporting and evaluation, having information distributed over multiple systems can pose some problems. It means that the information may not be in the right format in the right place for some specific type of analysis – this is especially true in a context of evolving requirements where new types of analysis can be requested at any time.

Given the above, it is worth highlighting some of the relevant DSP IT systems and indicating which data resides in which systems. The IT systems relevant within the context of this work then are:

- ⤴ The ISTS system, which contains all of the individual payment/transaction information, has complete information relating to each payment made to individuals on one of a number of employment related schemes (UA, UB, Jobbridge, etc). Each record contains information on the individual, the scheme they are currently on, the office they are associated with, the amount of the payment, etc

- Querying the ISTS system generally returns the above transaction records; as such the information is very much point in time and very granular.
- ISTS is largely a single (albeit big) application within the DSP's IT infrastructure which focuses on transactions in relation to DSPs short term schemes.
- △ The Central Records System (CRS) contains a large amount of information relating to individuals and their interaction with DSP more generally – it contains information for most individuals in the state. It contains records relating to the individuals, what DSP knows about their employment history (i.e., information in respect of any period of employment for which PAYE/PRSI has been paid), a summary of their interaction with the schemes supported by ISTS and DSP's other payment systems (the duration of the interaction and the scheme they were on), etc.
 - CRS is a primary information store for DSP – it has tools to afford flexible access to information stored in this repository and the DSP policy department makes extensive use of Business Objects to provide information relating to the live register status to interested parties.
- △ BOMi is the next generation of DSP's IT systems and the department is working hard to integrate its existing systems with this new platform. This process delivers services through modern, efficient channels and also maximizes the volume of data held on customers through increased usage of scanning and online submission of claim forms. When all customer data is held on BOMi, reporting environments can be built across multiple 'domains' (customer details, means, payments, schemes, etc) for activation and other purposes. The main DSP schemes and interventions being considered in this report are only partially recorded on BOMi at present (mainly PEX details and engagement with DSP jobs facilitators so far) and thus the reporting requirements identified will need to span all of the systems set out above.

A reference to the documents provided by DSP describing the above is included in Appendix 3.

3.2 Reporting Capacity

At present, reporting relating to the LR is done primarily on point in time data obtained from ISTS. The regular reporting is somewhat rudimentary – partially imposed by some limitations in the way data is made available for analysis – but more advanced reporting is done from time to time, typically driven by CSO needs.

The regular reporting involves determination of the following key metrics:

- △ Number of individuals currently on the LR – this is calculated by determining the total number of individuals on welfare schemes (JA, JB, JB credit signing etc) at a given time based on the ISTS records, excluding a small number of categories who are not expected to be genuinely engaged in job search such as short term workers;
- △ Inflow to the LR during a given period – this is calculated by determining the number of claims records which have a registration date within the specified range;

- each ISTS record contains a CLM_REG_DATE field which indicates the date the claim was registered;
- △ Outflow from the LR during a given period – this is calculated by determining the number of claims that are closed during the given period;
 - each ISTS record contains a CLM_CURR_PRIM_STAT_CODE and CLM_PRIM_STAT_DTL_FROM_DATE – if the former is 'Closed' then the latter field indicates the date of closure of the claim and 'closure reason', and these records are moved to a separate file for analysis

The Department has the ability to determine the above metrics based on the ISTS records, but it should be noted that it can be time-consuming to generate the above due to some resource limitations. Such resource limitations (mainly hardware issues) should be remedied such that it should be quick and straightforward to extract the above information.

As well as the above, the department is in a position to determine the mean duration of a claim, the distribution of the duration current claims for any given claim type at present etc. As with the more rudimentary reporting, obtaining these statistics can be time-consuming as it involves somewhat inefficient data processing.

The Department reports separately on 'open' and 'closed' ISTS extracts /point in time data at present, and this gives a limited view of a client's progression through the system over time. For example, a person may be fully unemployed at the start of a period, progress to an activation scheme (e.g., Back to Education), become casually employed afterwards, and subsequently (finally) close their claim. None of these transitions can be picked up based on extracts, other than the final claim closure which is archived separately. The customer's progress can be reported at points in time based on extracts, which involves matching up to 600,000 claim records at various dates, and can be (separately) matched to closures files covering a period, where the Department also holds useful information in relation to the reason a customer closes their claim. This is obviously time consuming and inefficient for reporting and evaluation purposes. The Department reports extensively on these extracts in relation to e.g., average length of time on benefits for current customers, their average weekly welfare entitlements, etc. and much of these point in time data are reproduced in the Minister's Live Register monthly memorandum. In addition, some further analysis of progression can be done by matching against e.g., claim extracts 6 months or 1 year on and closures for a customer cohort.

As set out above, the Department has some difficulty at present generating information which links current claim information to previous claim history or previous work history. While the Department currently has much of this information, there are some issues where the information is spread across disparate systems (CRS and ISTS specifically), and there are no specific linkages for reporting purposes between the concrete ISTS records comprising of a single encounter with the DSP, which contain much useful information – and are archived onto ISTS claim extracts at PPSN level - and an aggregated view which is retained in CRS and also held at PPSN level.

3.3 Evaluative Capacity

For evaluation activity we are assuming, either in a basic fashion or based on a more sophisticated econometric model, taking some formal comparison of a group involved in activation (a treatment group) with a group who are similar but not involved in any activation programme (a control group). Appendix 1 outlines a more sophisticated style of evaluation model but regardless of the degree of complexity, a number of key requirements follow. Put simply, there are two dominant needs – firstly to proactively select people for the treatment and understanding what happens with them through and after the scheme; and secondly to select people with similar characteristics to those who have gone through the system and understanding their path. However from a perspective of data requirements any evaluation will need the same information on both the treated group and the ‘control’ group. There are three basic approaches to compiling the necessary data.

- ⤴ Rely on administrative data (from DSP data, from Revenue etc).
- ⤴ Rely on specialized baseline, within-activation, and follow-up surveys linked to the administrative record.
- ⤴ A combination of these.

Monitoring systems are also required, particularly where randomization is used or where the delivery process may influence the outcome (such as a suspicion that the client base from a particular welfare office differs systematically from the client base in another location). The main steps in the analysis of the data include:

1. Decide on the appropriate ‘filter’ to define the population of any control group, based on the characteristics of the treatment group (e.g. 18-24 year olds, unemployed for defined period etc). More sophisticated econometric tools require as much information as possible about pre-programme outcomes to deal with the potential for complications such as selection effects (see Appendix 2 note).
2. Comparison of pre-treatment outcomes for participant group and comparison group, using whenever possible, the same outcome variables that are the focus of the evaluation e.g. exit from unemployment. This gives a baseline perspective.
3. Comparison of outcomes for treatment and control groupings and a check of response rates to follow-up survey (or within treatment surveys if implemented).

The summary chart in Appendix 5 outlines the process in a simple fashion. From our sense of the data we can:

- ⤴ Choose some set of PPSNs (or equivalent identifier) perhaps with some filter (when individual participated in scheme, gender, age, amount of time unemployed prior to participation, etc.), and similarly from a control set. We can presumably select these from the larger CRS and filter out those who have participated on the scheme. There may be a dependency on some historical work related information or information relating to their engagement with DSP in the selection process. This would mean that we would need some access to previous work history information as part of the selection process. However it is current practically impossible to select/filter based on education level (with the exception of clients entering via the PEX trials).

- ⤴ Upon selection of both the sample and control groups, it is necessary to compare the outcomes as defined from the outset. Some outcomes are easily achievable – exit from DSP claim for example. However many others are less readily available – for example, it is often not easy to determine why a claim has been terminated due to lack of data in the ISTS data set¹.
- ⤴ Some data gaps are quite significant therefore – prior education levels for non-PEX participants would seem a gap that should be filled both prospectively and retrospectively, though there may be significant associated costs

4. Recommendations on Information and Data System Requirements

4.1 Structure

The current data infrastructure is largely adequate in terms of the availability of data. The ISTS records contain comprehensive data on each LR transaction - who receives the payment, how much they receive and under what scheme they receive payment. The DSP also has quite comprehensive knowledge of each individual's employment history as obtained through Revenue information. Also, with the enhanced PEX profiling information, it is possible to perform more selective analysis of sample and control groups.

However, while there is a large amount of data available to DSP, there are some limitations relating to how the data is structured. The available data needs to be (i) longitudinal at customer level to measure progress over time; and (ii) episodic to separate the interventions, Social Welfare claims, and employment spells in the customer history (while allowing for the fact that some interventions will close/interrupt a claim, while others will not). Most data analysis within DSP is based on extracts of live and closed ISTS claims, which do not meet either of these criteria.

More specifically, the formation of sample and control sets for evaluation of a specific programme based on combinations of personal characteristics and employment information is non trivial as it involves combining data sets in ways that have not been combined heretofore. Also, determining outcomes for sets of people such as subsequent employment is non-trivial.

However, while the current infrastructure is suboptimal for the reporting and analysis requirements in the context of this work, it appears to contain sufficient data to perform useful analysis.

As mentioned earlier in 3.2 under 'Reporting Capacity': at present, there is poor linkage between the ISTS records and any kind of longitudinal view of an individual's interaction with

¹ It may be possible to make more inferences regarding what happens to an individual when an ISTS case is closed by performing deeper analysis of the data in the system, eg trying to find employment records at some future date, perhaps trying to find subsequent engagements with DSP on other schemes etc

the department. This can be done by creating a new record in which ISTS records from a given claims_continuous interaction are linked to the claims_continuous record. The availability of PPS Numbers across all records should facilitate this linkage.

4.2 Ease of Access

It will be necessary to invest in reporting environments for accessing the information, which as of now is in disparate databases with equally disparate structures. Such an environment could be relatively simple if focused on, say, a simple EAP transaction but would be significantly more complicated for interventions that are more elaborate. For example, live register information is available via ISTS, but it is probably too fine grained (i.e. a query for a particular PPSN will probably return a large set of ISTS events). This could be addressed with more episodic analysis of the ISTS information. The FÁS system could act as a base for this eliminating sizable investment in time and effort, as it is already developed on a longitudinal and episodic basis for analysis of FAS customer data.

As noted elsewhere ideally we need to see an access process that will retrieve and tabulate data for reporting purposes, and allow for the creation of evaluation samples for programme assessment. The requirements of any access 'tool' can be seen as:

- ⤴ Facilitating easy of access to key indicators for reporting or programme evaluation.
- ⤴ Obtaining a set of individuals who form the sample and control set
 - For the purposes of evaluation, it may be necessary/desirable to specify particular characteristics of the individuals in the sample set, e.g. age, gender, marital status, etc
- ⤴ Obtaining outcome data for both the sample set and the control group - some of the key questions revolve around having information on the employment history of the individuals combined with the ISTS records of same. Quite a number of different outcomes could be of interest here e.g. how many obtained gainful employment over some timeframe? How many moved completely off the LR? How many moved off the LR for some period but returned later? How many pursued self-employment? How many did neither reappear on the LR nor Revenue employment information and what happened to them?

Given the above requirements, a key question is whether the data infrastructure exists to construct such a tool; it is also interesting to understand how it can be built on the existing capabilities with a minimum of effort. An access 'tool' may operate as follows, for example, in retrieving analysis samples for programme evaluation:

- ⤴ Specify which programme is under evaluation
 - This may include a time period
- ⤴ Specify characteristics of sample set of individuals for purpose of evaluation
 - The characteristics could include age, gender, duration on the LR, etc
- ⤴ Determine sample set based on programme and above characteristics

- This will involve choosing from the set of individuals on the programme (during the interval) – possibly at random - and determining if they fit the profile stipulated. Choosing from the set of individuals on the programme can be done, for example, by querying the claims_continuous record
- △ Determine control group based on individual's characteristics
 - This may involve selection of individuals at random from CRS and ensuring that they did not participate in the programme (during the interval). A more sophisticated evaluation design will specify other issues to control for econometric problems.

4.3 Linkages to External Databases

The Department has excellent information exchange with Revenue in particular. DSP receives information from Revenue and stores information in its own IT systems: this is done primarily for two specific purposes – identification of fraud and determining how much social insurance contributions any individual has made.

With regards to the former, Revenue provides DSP with a 'Commencement of Employment' (CoE) notification when they have been notified that an individual is commencing employment. The employer needing to know the individual's tax status to make payroll calculations usually triggers this. This is a timely information source that can be used to identify individuals who have terminated their linkage with DSP to commence employment.

Throughout the course of this work, the DSP kindly performed an analysis of the quality of the CoE data to help understand how reliable it is as a source of information. The resulting analysis found that most of the notices are received (over 80%) and they are received in quite a timely fashion (mostly within 3 months of the event).

As well as the CoE notification, DSP receives annual information related to each individual's employments for a calendar year. This information is as provided by the P35 form, in which each employer must provide information on each employee including how many weeks have been worked in a calendar year and total income received in their employment. DSP's IT systems store this information and it is possible to determine, for example, which sector a given individual has worked in, how many employers a given individual has had within some timeframe, etc. It is worth noting that from the point of view of episodic analysis of an individual's labour market activity, the employment information is suboptimal in the sense that the annual employment information is not currently aggregated for analysis purposes into a single employment record over a longer time horizon.

4.4 Design for Evaluation

A universal point is that the design process for programmes needs to ensure evaluation is possible. Ex-Post assessments of current programmes are fraught with pitfalls. It is not in the scope of this document to develop this point in detail but the key issue to consider is the selection process (particularly with voluntary schemes) and critical issues surround content, the potential timeline (short/medium/long run impact) etc. When interventions are rolled out in multiple sites, the design must include adequate resources for monitoring the

implementation at all sites. In many evaluations, the choice of the comparison group is dictated by data availability.

The planned focus of output should also be decided from the outset to ensure a lack of 'drift' – did the programme have the impact intended in the time expected? Did participants find work within the timeframe expected? Did participants in subsidized schemes move to the non-subsidized sector at the same employer? What was the cost per participant?

The fundamental issue in design is the choice of the comparison group. Typically, we will have identified the treatment group. We will then begin with a population for a control or comparison group, and then impose additional restrictions (data or statistical assumptions) to select a best match to the characteristics of the treatment participants. Comparison groups maybe be stratified by geographical location; by time period (so we do not necessarily need the treatment and control group to be contemporaneous); or by eligibility.

5. Enhanced Reporting for DSP

Here we take a number of key reporting domains and examine how the reporting requirements could be achieved using current infrastructures. These reports are:

- ⤴ Numbers processed weekly, monthly and year-to-date with year-on-year comparators by age and gender;
- ⤴ Numbers in active labour market programmes with totals starting/completing the programmes and referred through the NEAP or otherwise
- ⤴ Length of time since last employment
- ⤴ Type of previous employment – maybe NACE sectors
- ⤴ Length of time since last intervention (including those with no intervention since the current LR sign on)
- ⤴ Queue length for types of intervention by local area office and region
- ⤴ Numbers of sanctions applied by region and at each intervention stage.

5.1 Numbers Processed Weekly/Monthly/YTD with Comparators

Numbers processed weekly/monthly means anyone who has either has an awarded, refused or withdrawn application in a given week/month. To obtain these numbers, it is necessary to determine the number who have had a new ISTS claim awarded/refused/withdrawn with the specified date range. The Department has well developed process for determining this information at the so-called scheme level, where scheme could be Jobseeker's Allowance or Jobseeker's Benefit; however, many ALMP interventions are implemented as subcategories of a particular scheme or 'low level schemes' on ISTS – as this information is only available in the ISTS records, it is more difficult to work with as numbers refused, withdrawn or pending decisions (i.e., 'queueing') are not available.

To obtain the numbers process on a given ALMP scheme, for example, it is necessary to query to the set of ISTS data to determine the number of individuals who fit a given selection criteria, eg have claims of a given type and have claim activity within some specified date range. This can involve dealing with a large amount of ISTS records – many multiple records

per individual – unless the reporting environment supports queries in which unique fields can be identified.

Note that it is possible to have some approximation of this data using the claims_continuous records within CRS; however, these records do not contain sufficient information to differentiate between all of the different schemes offered by the Department. For example, using this approach, it is not possible to differentiate between an individual on a Jobseeker's Allowance Penalty Rate or someone who is on an internship or Springboard programme. It could be possible to modify the claims_continuous data structure to include more detail on the specifics of the scheme the individual is on: in this way, it could be possible to provide for more useful analysis using this summary record, rather than requiring more complex analysis of the detailed ISTS data.

FAS has a good case tracking system, so it is relatively straightforward to obtain information relating to numbers processed within their system for different time periods.

5.2 Numbers in ALMP

It is not entirely clear which ALMP schemes are considered here: indeed, it is not entirely clear which schemes qualify as ALMP schemes per se. Here, the focus is on FAS schemes.

Clearly, FAS schemes fall into the category of ALMP schemes and the aggregate numbers who are at different stages within FAS schemes at any point in time can be obtained from their case tracking system. FAS can provide details on the numbers of people referred through NEAP. For their programmes, it is possible to determine the number of people starting and completing the programmes.

5.3 Length of Time Since Last Employment

Determining the time since the last employment looks difficult to determine exactly from the current information gathered by DSP. The Department has the following information relating to employment:

- ⤴ No of weeks worked in any given calendar year for each individual; it should be noted that this information is received in P35 annual returns and hence there is a significant timelag between the cessation of employment event and receipt of this information. Further, this information is insufficient to determine the specific dates during which the individual was working
- ⤴ Cessation of employment notifications received from Revenue – these notifications are received and inserted into an employment record within the CRS. Note that this employment record may not have complete information relating to the employment (eg when it commenced, for example); also while Revenue provide this in a systematic way, and any data they receive is likely to be received by DSP, the data can have some deficiencies as employers may not report cessation of employment, there may be errors in there reports, etc.
- ⤴ PEX information for some individuals – this provides approximate information on when people who are commencing a claim last worked (eg in the last year, in the last month, etc)

None of the above information gives a simple approach to determining time since last employment, which is guaranteed to be correct. However, when the above information

sources are combined, it is likely that it will be possible to obtain a good estimate of the length of time since last employment for many individuals. Combining the above information does require some work and some analysis is necessary to understand in more detail the quality of the information (Cessation of employment in particular) and also if date of cessation of employment can be inferred from P35 returns in many cases.

Without further analysis, it is not possible to state definitively how specifically to combine the above information, nor how accurate the result will be. However, it is likely that some combination of the above will give quite a reasonable estimate of the length of time since last employment. (It has been noted that there may be an effort within CSO to combine information from Revenue, DSP, education and FAS to understand an individual's interaction with all these systems and to construct an easily navigable path through their education/training/employment activities. DSP should investigate if this is so as this may be useful in the context of knowing concrete information relating to an individual's previous employment).

5.4 Type of Previous Employment

Type of previous employment is somewhat ambiguous and disparate and potentially conflicting sources of information could be used here. The following sources of information could be used to determine this:

- ⤴ information pertaining to previous employer(s): information on the employer(s) as determined above could be used to determine which sector an individual worked in; there could be issues that a given individual may have had multiple previous employments, perhaps in different sectors;
- ⤴ information relating to occupation:
 - information relating to an individual's occupation is stored in the ISTS records;
 - information relating to an individual's occupation from ISTS are also stored in CRS – interestingly, these records can have a start date and end date associated with them; also, it is possible that an individual could have multiple occupation records
- ⤴ 'desired occupation' is also stored for those individuals who have had interaction with FAS.

The question then arises what the purpose of this analysis is: is it to determine which sectors are seeing the most stress in the labour market, or is it to determine which skills are having difficulty finding employment? Of course, these are not unrelated questions, but they are distinct and result in different approaches when working with the Department's data.

If the focus is on sectors, then the approach should be based on analysis of previous employer's industrial sectors. If the focus is on skills, then the approach should be based on using the individual's occupation record(s) and/or the occupation information in the ISTS data. Some care must be taken to deal with the case in which an individual has more than one occupation record.

Information relating to previous employers should be both available and accurate in the main as it comes from the P35 records. As such, the information is available to perform analysis of the types of previous employment for a given set of individuals, such as those who have just

joined the LR. Issues could arise, however, when an individual has multiple previous employers – differentiating between them and identifying which was the last employer could be a difficult in the context of a single database query pending further analysis of the quality of employment cessation date information as mentioned in 5.3.

5.5 Length of Time Since Last Intervention

There are a number of different possible interventions and some of them are dealt with in different ways within the Department's IT systems. This gives rise to some difficulties answering this question in the most general sense. A key issue in this regard is the fact that some interventions interrupt a single ISTS claim, while others don't; hence, the notion of an intervention is not reflected clearly in the Department's IT systems.

As the Department's IT systems have different implementations of interventions, it is not straightforward to establish the length of time since the last intervention. It does necessitate analysis of the ISTS records, as the date ranges relating to interventions that do not interrupt a claim are only visible at this level (and even this requires some work to determine).

To determine the length of time since the last intervention for an individual, then, it is necessary to extract ISTS files pertaining to that individual, process the ISTS files looking for the claim codes pertaining to interventions and determining which was the most recent. To perform this over an aggregate set (such as those who signed on to the LR this month) is more complex and seems to be beyond the capabilities of the database query based operations – within the confines of the current data structures – that are mostly used within the Department. (Of course, it is possible to write some scripts to extract this information and process it accordingly without much difficulty in the context of a reporting environment).

As noted above, it may be possible to modify the claims_continuous record to reflect some of this information; in this way it would mean that it was not necessary to process a set of ISTS files to determine the last intervention for any individual and thus obtaining this information would be more efficient.

5.6 Queue Length for Types of Intervention²

The number and types of interventions possible is substantial and a queue length indicator (or the number of 'pending applications') does not make sense in the context of some of the interventions. The following interventions are considered here:

- ⤴ Jobbridge: a Jobbridge is essentially a matching service, there is no specific queue associated with it and it is ignored here
- ⤴ FAS Community Employment Schemes effect recruitment through a Job Bank – this does not have a notion of a queue or waiting list per se
- ⤴ FAS training schemes: FAS has a comprehensive case tracking programme which can be used to determine queue lengths for access to their programmes at different points in time
- ⤴ Tus schemes: – ISTS claims are closed and paid under separate arrangements for

² It may be more useful to have information relating to the amount of time it takes an individual to get called rather than the amount of individuals in the queue.

the TUS scheme – no queue data available at present

- △ Back to education schemes – from a DSP point of view, these are simply a claim category on ISTS; no information is available within this system relating to processing times or queue lengths.

5.7 Numbers of Sanctions

ISTS records support the notion of penalty rates for JA and JB claim categories. These can be used to determine the amount of individuals on penalties, broken down by location if desired, at any point in time.

Another data point in the penalty discussion arises in the information flow from FAS to DSP: in this information flow, FAS highlight those individuals who have not presented for interview. This information could also be used to determine the numbers to whom penalty rates should apply. However, DSP does not as a matter of course store this information with the individual's record – this information is only reflected in the application of penalty rates in the ISTS system and is at the discretion of the case officer dealing with the individual. This information should be used to have more information regarding the statistics of those who are interacting properly with the system.

Obtaining information on how many individuals have been sanctioned during a given time period is a little more problematic. At present this must be done by obtaining a (potentially large) set of ISTS records which have subcategory codes corresponding to penalty rates and determining the number of unique PPSNs that these apply to.

As this is a particularly important piece of information in relation how individuals are interacting with intervention schemes, it would be useful to make this available at a more aggregate level within CRS/BOMi – this could be done by modifying claim_continuous indicating that a particular individual was on a penalty rate for some period. This would make it easier to analyse how many individuals were on penalty rates, how long they were on penalty rates for, etc.

6. Additional Issues

6.1 Enhanced reporting mechanisms

As mentioned in 3.2 and 4.1, the current reporting infrastructure is deficient in the sense that the full resources of the Department are not currently being leveraged. The Department has a wealth of data at its disposal but has not set up appropriate reporting environment to process this data to perform effective reporting. This needs change. The following two recommendations can help alleviate this.

- ⤴ In the near term, some ad hoc processes need to be defined to enable the information to flow more freely within the Department, enabling the appropriate data to be used as the basis of the reporting. These processes include reporting arrangements for 'Pathways to Work' targets and enhanced analysis of claim closure rates for various interventions and customer groups. Representatives of the Department have demonstrated an ability to realize these as necessary so it would seem that this needs structure rather than invention.
- ⤴ In the medium term, a more systematic approach is required which renders straightforward the extraction of appropriate information from Departmental databases. This will probably require the development of a distinct IT system that mirrors aspects of the Department's data holdings, but is sufficiently separate from the operational systems that it cannot affect them. A project should be established with business and IT involvement to develop this reporting environment, with the aim of developing an integrated customer history across all interventions and work outcomes for reporting purposes.

6.2 Enhanced capacity for data analysis

The Department appears resource constrained in terms of expertise relating to statistical modelling and analysis. This is an important function when it comes to understanding the operation of the labour market and making appropriate, data driven, policy recommendations, particularly given that the set of data the department has is large and has gaps in some areas (eg understanding an individual's work history, making inferences on total labour market activity from limited PEX information, etc).

It would be beneficial for the Department to expand the modelling and analysis skillbase to effect deeper analysis of activities within the labour market. This would incur costs in the shorter term, but would reduce the amount of outsourcing necessary within the department over the longer term.

6.3 Data collection

While the department does have a large set of data at its disposal, there are significant deficiencies in the data when it comes to performing credible analyses of activation programmes. Some elements that are missing from core claimant files (education level, exit/destination information) seriously hinder capacity to develop appropriate policy

responses. It is difficult to examine, for example, the duration history by education level of claimant. Greater effort is needed to ensure this information is gathered.

- ⌘ A more aggressive rollout of the PEX claimant forms should be considered (though with associated costs) to aid analysis of the performance of activation schemes.
- ⌘ Mechanism to more accurately understand the reasons individuals have left the LR should be investigated.
- ⌘ Conceptually it will need to become a 'norm' to be able to analyse episodically and longitudinally the claimant records for reporting purposes. Similarly for more sophisticated evaluation the ability to deliver sizable files for treatment/control group assignment will need a responsive IT/data system.

References

Angrist, Joshua and J. Pischke. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton NJ: Princeton University Press.

Card, David; J. Kluve, and A. Weber. (2009). *Active Labor Market Policy Evaluations: A Meta- Analysis*, NRN Working Papers 2009-02.

DiNardo, John and D. Lee. (2010). "Program Evaluation and Research Designs," NBER Working Papers 16016, National Bureau of Economic Research.

Heinrich, Carolyn, A. Maffioli, and G. Vázquez. (2010) *A Primer for Applying Propensity-Score Matching*. SPD Impact-Evaluation Guidelines. Technical Notes. IDB-TN-161, Inter-American Development Bank

Lechner, Michael. (2002). *Microeconometric Evaluation of Active Labour Market Policies*. University of St. Gallen, Department of Economics, Discussion Paper No. 2002

Appendix 1 – Terms of Reference

1. To set out a detailed description of the current intervention process model and its individual elements including targets for service delivery and throughput at defined milestones (the group interview pilot, the subsequent NEAP interventions and monitoring and the associated control measures following people's failure to engage with activation).
2. To identify the existing data streams within and available to the Department of Social Protection on that Department's intervention with unemployed people on the Live Register.
3. To make recommendations on appropriate information and any attendant data and systems requirements to measure the impacts of the intervention regime upon client off-flow rates.

Data requirements to be considered, but not limited to, the following:

- Original claim to jobseekers' payment (fulltime/part-time and credits) (data collection point for activation purposes via profiling (PEX))
- Invitation to client to participate in group engagement (at 3-6 months) where client remains on the Live Register
- Follow-up intervention following group engagement whether:
 - Referral to appropriate interventions ("the 65%") and measuring the impact of those interventions over time
 - Referral to monitoring and control ("the 35%") and the assessing impact of those measures over time.

To also include data emanating from existing data-streams at local office level. For example, weekly reports from local offices used for compilation of weekly Live Register reports and the quarterly report to Government on the Live Register.

4. To make recommendations on the appropriate management reporting systems to produce weekly and monthly reports on
 - Numbers processed weekly, monthly and year-to-date with year-on-year comparators by age and gender;
 - Numbers in active labour market programmes with totals starting/completing the programmes and referred through the NEAP or otherwise
 - Length of time since last employment
 - Type of previous employment – maybe NACE sectors
 - Length of time since last intervention (including those with no intervention since the current LR sign on)
 - Queue length for types of intervention by local area office and region
 - Numbers of sanctions applied by region and at each intervention stage.

Appendix 2 – Note on Evaluation Design

Active labor market programmes (ALMPs) include three broad classes of interventions – training programmes, subsidized employment programmes (such as Community Employment (CE) or JobBridge), and job search assistance programmes.

The demands of a full evaluation impose more on both the data infrastructures and the basic design of the ALMP – in other words the programme needs to be designed for evaluation from the outset. Moreover, a robust evaluation needs to ensure careful assessment of the outcomes of focus – evaluations rarely drift into subject areas such as how satisfied individuals are with the ALMP experience, rather are focused on quite specific outcomes such as whether the ALMP raised the employment and/or earnings of participants at some point after completion of the intervention.

The critical issue is to design and implement the ALMP and its evaluation so to address the key question of how the intervention group would have fared if they had not participated in the ALMP. This is fundamentally different to simply examining a group before and after an ALMP. This ‘counterfactual’ approach instead attributes an impact of an ALMP as the marginal impact of the ALMP over what would happen without the intervention. This is a more challenging approach requiring two basic research designs for an evaluation – experimental or quasi-experimental.

- ⤴ In an experimental or randomized design, a group of individuals who satisfy the eligibility requirements for the programme are randomly divided into two groups: the treatment group, who are assigned to receive the programme, and the control group, who are assigned to not. Because assignment is random, the treatment and control groups would be expected to have similar experiences in the post-program period in the absence of the programme intervention, so the outcome can be viewed as a causal impact of the intervention. To take a simple example, the PEX profiling system could be evaluated by random assignment of individuals to the PEX system, with comparisons of the outcomes to a cohort progressing through the normal EAP process (adjusting for the sample characteristics).
- ⤴ A quasi-experimental (or Difference-in-Differences) design uses a non-randomly selected group of ALMP non-participants to proxy as the control group in the experimental design. To take an example from the current Irish environment, a design to evaluate something like JobBridge would match the participants in the internship scheme to a sample of non-participants using a sophisticated econometric technique. For any individual intern, this effectively finds a paired match who is a close proxy for them but is NOT participating in JobBridge. We can then follow their outcomes, and the change in the differences between these two individuals over the period of the internship (i.e. the difference in the differences) is attributable to the internship. Aggregated over all participants this gives the programme effect.

The challenge is twofold – effort must be invested in ‘matching’ the control group sample to the intervention sample, which is an analytics capacity problem. The second more

fundamental problem is that we must be confident that there is no systematic 'selection bias' element underlying the two cohort.

Some econometric assumptions are generally needed. To illustrate the point we can look to the JobBridge example - we can accept that there may be differences in that the participants in JobBridge (as a voluntary scheme) may be more motivated than non-participants. As long as this pre-internship difference would have remained fixed in the absence of the programme (or that the selection bias is constant over time), then we can use the difference-in-differences approach to estimate a programme impact – we would simply adjust the post-ALMP outcomes to account for the fact that differences existed pre-ALMP. Once a potential comparison group is selected, refinement will almost certainly require access to pre-treatment condition. This is less true for ALMPs aimed at certain sectors like youths who may be relatively homogenous in terms of education and experiences. For adult unemployment experiences, life history to the point of job loss is a major factor.

Appendix 3 – Sources of Data Information

DSP provided much detailed information on the operation of their IT systems covering many different aspects of their systems. Their systems are large and complex with a long history and it is difficult to succinctly capture them here. Here, we just note the sources of information provided by DSP which were used to understand how their systems work.

A3.1 List of information sources

The following information was provided by DSP throughout this work:

- ⤴ Memorandum of Understanding on Co-operation and Mutual Assistance between the Department of Social Protection and the Central Statistics Office
 - sent by Paul Morrin 7 Dec 2011
 - main source of information on ISTS records
- ⤴ Revenue/DSP Data Exchanges
 - sent by Paul Morrin 8 Dec 2011
 - rich source of information on two-way information flow between DSP and Revenue
- ⤴ JCP B form
 - sent by Paul Morrin 8 Dec 2011
 - form for collection of richer information relating to Social Welfare recipients – so-called PEX related information
- ⤴ DSFA ICT Systems
 - sent by Paul Morrin 9 Dec 2011
 - presentation given by Sean Fay to delegation from Ukraine in January 2011
 - good summary of evolution of DSP IT systems
- ⤴ EAP related information
 - sent by Paul Morrin 12 Dec 2011
 - comprises of description of how EAP information is dealt with as well as specifics of mechanisms and information formats shared with FAS
- ⤴ BOMi object model formats
 - sent by Paul Morrin 13 Dec 2011
 - detailed breakdown of structure of information stored in BOMi system
- ⤴ claim_continuous data structure

- sent by Paul Morrin 14 Dec 2011
- specific information record from CRS which is useful for performing some longitudinal analysis
- ⤴ CRS Data models
 - sent by Sean Fay 6 Jan 2012
 - detailed description of data models comprising CRS
- ⤴ Analysis of quality of Revenue Commencement of Employment Notifications
 - sent by Sean Fay 11 Jan 2012, 17 Jan 2012
 - this analysis gave some insight into the quality of the CoE data provided by Revenue – what percentage of notifications are received, how timely are they?

Appendix 4 - A short note on performance of database manipulations for analysis

[DSP has very extensive experience dealing with large scale databases supporting complex mission critical operations. These basic experiments were performed to obtain some concrete results for performance/scalability considerations database manipulations similar to some of those necessary to perform the types of analysis specified in the ToR.]

Throughout the course of this work, DSP raised the issue that they have observed cases in which there can be significant delays in processing the data necessary to understand the labour market. In light of this, they raised some questions about the import of the way data is structured and the impact this may have on performance.

In general, the workload of the policy division is very modest in the context of the today's computing power and any reasonable database system should be able to accommodate the needs of the department without difficulty. More specifically, the policy division deals with sets of ISTS records reflecting activity typically on weekly/monthly basis – such sets would comprise of approximately 600k records. A year's such records would account for 6m records.

To demonstrate typical scaling properties of modern database system, a simple experiment was performed in UCD in which records were added to a database that is similar to those the DSP deals with and certain queries were made. The experiment shows that current hardware can easily deal with the types of workloads the department has, although it is worth noting that can be relatively easy to use tools in a way that delivers significantly suboptimal performance.

A4.1 Experimental platform

The experiment was performed on the following system:

Hardware specifications:

- ⤴ Intel i5-680 dual core processor (4M Cache, 3.60 GHz)
- ⤴ 8GB (2X4GB) Non-ECC,1333MHz DDR3 RAM
- ⤴ 320GB hard disk

Software:

- ⤴ Debian (version) v6.0.3 Operating system
- ⤴ MySQL v5.1.49 database
- ⤴ Python 2.6 with SQLAlchemy 0.7.4

The experiment was constructed in Python using the SQLAlchemy library – this affords great flexibility over the use of the system. Of course, other systems could have been used, but the team has most experience working with these technologies.

A4.2 Experiments

The following experiments were performed to gain an understanding of the difficulty of dealing with the types of records DSP has to deal with; in particular, it was interesting to note if the workloads arising from the DSP activities would generate substantial load on systems, giving rise to delays and performance problems.

The experiments that were performed:

- ⤴ Add 500k records to the database from remote client – this was meant to emulate the addition of ISTS records to a database
 - Add 500k records which are independent and perform db query
 - add 500k records which are linked to other records (PPSN) and perform query based on join operation
- ⤴ Add 500k records to the database which are linked to PPSN records, a previous employment record and an employer record – deduce which records are new and determine which employment sector these records relate to

The findings from these experiments are noted below.

A4.2.1 Experiment set 1 – adding records to a database

The first set of experiments was conducted to understand how long it takes to add a significant set of records to the database. All of the records were of a similar size to ISTS records (about 60 fields, most of which were strings of 50 characters in length).

The experiment involved populating the records with some padding and adding them to a database. In this case, the database was empty at the start of the experiment. More specifically, the experiment consisted of a python script which initialized the database, created the records and wrote them to the database using SQL. (In this case, both the database and the client operated on the same machine, but they were separate and distinct processes – this result would hold if the operation was performed over a high performance network).

It was found that it took just over 8 minutes to add all of the records to the database. This seemed like quite a long time for quite a rudimentary operation.

On a little further investigation, it was found that the delays related to the insertion of each record into the database individually and the associated recalculation of indexing tables each time.

An alternative approach to add significant numbers of records to a database was considered. MySQL supports an operation in which a set of records can be loaded from a file and inserted into a database. To do this, the file must be resident on the server – somewhere that the database process can find it – and the database must be instructed to read the file into the database.

Inserting records into the database in this manner is very significantly FASTER: using this approach, 500k records could be inserted into the database in under 3 seconds.

The primary reason for this enormous performance gain is that all of the records can be added to the database in one operation and then indexes can be calculated based on the addition of all of the records.

These experiments were then broadened out to include a second set of records which were linked to the first set. In this case, a separate record was created which was linked to each of the first 'ISTS-like' records – this second record was intended to emulate a PPSN record. The experiment then involved adding the 'ISTS-like' records and the 'PPSN-like' records and making a simple query. More specifically, each of the 'ISTS-like' records contained a date and the query was to determine which set of PPSNs had a (linked) ISTS record which fell within a specified date range.

In this case, the focus was on the query operation as it has been shown above that it is possible to add records to the database in a very short time. It was found that the query operation – a query of 500k joined records resulting in about 40k responses – took under 3 seconds.

The conclusion of these simple experiments is that it is entirely possible to add records to a database in a manner similar to the ISTS records would be added to a system in a short period of time; also, query such a database should only take a few seconds.

A4.2.2 Experiment set 2

The second set of experiments built on the first set of experiments. The objective was to emulate the process of determining which sectors new registrants to the LR come from. The work involved emulating the record structure of the DSP and querying the database to obtain the results.

More specifically, records were created to emulate ISTS records, PPSN records, employment records and employer records. ISTS records were linked to PPSN records; employment records were also linked to PPSN records and employment records were linked to employer records. The experiment involved the following steps:

- ⤴ create all the records
- ⤴ query the database to determine the set of PPSNs which are new to the register (a data based join query linking ISTS records and PPSN records) – this resulted in approx 40k records
- ⤴ iterate over all these PPSN records, determining last employment information for each PPSN, extract sector of employer
 - when iterating, collect sector statistics

Once the records were created, it took approximately 30 seconds to generate a table of employment sectors and amount of ISTS records associated with same.

A4.3 Conclusion from these simple experiments

While database technology is mature, it has a significant amount of complexity. Working with large databases is an everyday occurrence and the types of work required by DSP are quite

modest. Hence, modern database systems should easily be able to accommodate their needs.

That said, it is possible to use databases inefficiently resulting in poor system performance. One example is the rudimentary approach to adding records to the database highlighted above. In most cases, such problems arise when the correct use of the system is not being applied. Hence, some care must be taken when working with databases to achieve good performance.

A final related remark: today's databases contain much intelligence in which they continuously monitor their usage and dynamically adapt to workloads. In particular, they often adapt their indexes to the workload they perceive so as to improve performance. Given this, it is not often necessary to put a lot of effort into database optimization save for the very largest data sets.

Appendix 5 – Summary Chart for Prospective Evaluation

AVAILABLE:

Employment cessation date
Duration of unemployment spell
Prior intervention/training

Available:

Exited unemployment

Control Group

T_0

$\Delta I MP$

T

Treatment Group

REQUIRES INVESTMENT:

Education level attained
Previous earnings
Prior work experience (spells)
Most recent work experience
Prior employment sector (NACE)

Unknown:

Exit to self-employment.
Employer type/NACE code
Employment status (FT/PT)
Subsequent earnings
Entrance to FT/PT education
Emigration

Additional surveys conducted on the T and C cohorts between T_0 and T_1 .