**An Overview of Data Needed to Develop
a Front-line Decision Support System (FDSS):
A Summary of the Georgia Experience**

**Kenneth J. Kline
Christopher J. O'Leary
W.E. Upjohn Institute for Employment Research
300 South Westnedge Avenue
Kalamazoo, Michigan 49007
(269) 343-5541**

**kline@upjohninstitute.org
oleary@upjohninstitute.org
www.upjohninstitute.org**

February 10, 2003

**An Overview of Data Needed to Develop
a Front-line Decision Support System (FDSS):
A Summary of the Georgia Experience**

Table of Contents

1) Introduction

The Workforce Investment Act (WIA) of 1998 emphasized the integration and coordination of employment services. Central to achieving this aim is the federal requirement that local areas receiving WIA funding must establish one-stop centers, with providers of various employment services within a local labor market assembled in one location. A major challenge facing staff in one stop centers is the ability to quickly prepare individualized service strategies for customers.

The Frontline Decision Support System (FDSS) is a set of administrative tools developed by the W.E. Upjohn Institute for Employment Research with support and cooperation from the Georgia Department of Labor and the U.S. Department of Labor, to help frontline staff and customers make better decisions about job prospects and appropriate services.

The Upjohn Institute is in a unique position to design, develop, test, and implement FDSS. The Institute not only conducts employment-related research, but also administers the state and federal employment programs that are the responsibility of the local Workforce Investment Board.

FDSS is comprised of two independent but complementary modules that produce recommendations to help staff and customers make informed decisions. Recommendations provided by these tools are based on recent activity in the state of Georgia including labor market information, historical service data, and customer characteristics. The two modules of FDSS are:

Systematic Job Search Module (SJSM)

- Provides customized information about the job market

> Reemployment Probability and Estimated Earnings
> > Probability of Return to Work in Prior Industry
> > Expected Job Growth in Prior Occupation
> > Likely Reemployment Earnings
>
> Related Occupations

Service Referral Module (SRM)

- Provides customized information about the likely effectiveness of various reemployment and training services

> Service Referral
> Training Statistics

This brief paper documents the assembly of various data sets from Georgia Department of Labor administrative systems. At the end of this document, Table 1 lists the variables entering each of the three main FDSS statistical models: return to prior industry, likely reemployment earnings, and employability. The data source file for each variable is also listed. Each data source file is overviewed in this paper. The Georgia FDSS lists additional variables

in the customer background characteristics for informational purposes for frontline staff, these variables do not enter statistical models of FDSS. Additionally, the prior occupation code is a key to presenting information on the local expected job growth in the prior occupation and information on related occupations.

2) Data Overview

Since the FDSS project began in Georgia in late 1999, the Upjohn Institute has constructed a data set which follows the activity of persons in several employment related programs including Unemployment Insurance (UI), the Employment Service (ES), JTPA, WIA, and Welfare-to-Work (WTW). The merge of the Georgia Department of Labor data from these programs along with individual wage and ES202 employer information received from Georgia State University has resulted in a longitudinal data file of program participants covering the time period January 1, 1996 through June 30, 2002. Individual, quarterly wage and ES202 employer information date back to January 1, 1993.

Due to resource constraints, the Georgia Department of Labor was not in a position to construct the longitudinal data file for the Upjohn Institute to analyze. Instead, in most cases, the full, administrative data files were sent to the Upjohn Institute from which information relevant to FDSS was extracted and used to construct the longitudinal data file. Including the update to the FDSS models which is underway as of this writing, this has involved the transfer of roughly 30 gigabyte of raw data to construct a data file which documents the program participation over time of about 1.2 million persons.

The following will describe every file sent by Georgia DOL to the Upjohn Institute and discuss the concepts extracted from those files that were used to create the FDSS statistical models.

3) UI Data File

The only file sent by Georgia DOL to the Upjohn Institute that was an extract of a larger data set and not that full data set itself was the UI data file. This data contained UI claims information for all persons entering UI from January 1, 1996 through December 31, 2002.[1] The variables extracted from the Georgia UI data files include: Social Security Number, Benefit Year Begin date (BYB), Benefit Year End date (BYE), the Weekly Benefit Amount (WBA), the maximum amount of benefits the UI client could potentially receive in his/her benefit year, the maximum number of weeks the client could potentially receive benefits in his/her benefit year,

---

[1]Although the data extend through December 31, 2002, it is necessary that the UI clients used for analysis have completed their benefit year. Given that, and allowing for a one quarter lag for the UI data to become complete, the longitudinal data file only contains UI clients with Benefit Year Begin dates (BYB) on or before September 30, 2001.

the actual amount of entitled UI compensation received by the client in his/her benefit year, the number of weeks the client actually received a benefit in his/her benefit year, and lastly, his/her profiling score if one was generated.

Two of the above concepts require further explanation—entitled compensation received in the benefit year and the weeks of compensation in the benefit year. The actual amount of entitled UI compensation received by a client in his/her benefit year may be greater than the sum of the checks issued to the client. In any given week for which the client submits a claim for benefits, should the client be determined eligible for the benefit, his/her entitled UI compensation is the weekly benefit amount minus deductions due to earnings or partial unavailability for job search in the week, but before any deductions for taxes, child support, repayment of past overpayments or other deductions. It is this amount—the amount of entitled UI compensation—that is important for analysis.

The second concept requiring further explanation is the number of weeks the client actually received a benefit in his/her benefit year. Depending upon state law, this number may be greater than the maximum number of weeks the client could potentially collect benefits (entitlement). The maximum number of weeks the client could potentially collect benefits assumes the client is entitled to the full WBA in the weeks that he/she claims and is defined as the maximum amount the client could collect in his/her benefit year divided by the WBA. If the client reports earnings and is therefore only entitled to less than his/her WBA, that week counts fully toward the number of weeks the client actually received benefits. Depending upon state law, it is theoretically possible for a client who is partially employed to report some earnings in each week of his/her benefit year and file a claim for each of the 52 weeks of his/her benefit year. In each of those weeks, if still found eligible for a partial UI payment despite some earnings, the client would receive less than his/her full WBA; however, the week counts fully toward the calculation of the number of weeks for which the client received a benefit in his/her benefit year. Therefore, it is theoretically possible, depending upon state law, for a client to have a value of 52 for the number of weeks the client received a benefit in his/her benefit year.

4) Master/Combined Intake File

In Georgia, the combined/master intake file contains a record for every person who registers with the Employment Service regardless of whether the client is part of UI. All clients interacting with the ES are part of this file. Along with the client's SSN, this file contains important demographic information. The important concepts needed for construction of the longitudinal data file include: county of residence, office code where the client registered, the zip code of his/her residence, age, race, veteran status, and educational attainment. Other important variables include handicap/disability, welfare receipt status, and whether the client is economically disadvantaged, a dislocated worker, currently employed or currently in school.

3

5) Work History Files

Prior to March 2001, the work history file in Georgia was limited to the most recent three jobs for a client. That constraint was removed when the internet based Georgia Workforce System (GWS) for one-stop career centers became operational. Key variables in this file include the client's SSN, the start and end dates of employment, the reason the client left the employer, and new to the GWS, some indicators about the benefits the client received from the employer (sick leave, retirement, vacation, etc).

Most important to the FDSS models are the start and end dates of employment from which job tenure is calculated. The separation reason is also important. Indicators about benefits received from an employer are not part of current, FDSS models, though, over time, as sufficient sample size accumulates, their importance could be investigated. Though the client's occupation code (Dictionary of Occupational Titles (DOT) or Standard Occupation Code (SOC)) is taken from another file (ES Services file described later in this document), if that variable existed in the work history file, it would be of great value and would be used for analysis.

Another file in the Georgia system further summarizes the client's work history, only from the point of view of the different occupations the client has had over time. Along with the client's SSN, a DOT code is included for each occupation the client has had in the past and the months of experience the client has had in that occupation. Also included is the most recent salary amount associated with that occupation.

6) Worker Profile Data File

Another file, closely related to work history, provides a profile of the UI or ES client. It includes up to four DOT (occupation) codes, with the first one reflecting the client's most preferred occupational choice. Along with each DOT code is the number of months of experience in that occupation and the date the client last worked in that occupation. Other data in the file which provide a profile of the job seeker include the client's "required" salary, shift availability (1st, 2nd, 3rd, rotating), desired job duration (full-time, part-time), willingness to commute and whether the client has a driver's license or requires public transportation. The file further includes information concerning any driving licenses (commercial or other) the client may possess and whether the client has any physical limitations.

7) ES Services Data File

While all Georgia DOL data files provide at least some information useful to creating statistical models, the ES services file provides critical information for both the earnings and service/training referral portions of FDSS in Georgia. The file documents all "transactions" with the employment service for UI and non-UI clients. All records are uniquely identified by the client's social security number, the date of the transaction and a code which identifies the

transaction type. These records contribute to the time and characteristic profile of a client by identifying when the client registered with the ES (new applicant record), the services he/she received, when those services occurred, and what descriptive characteristics apply to that client.

A critical analysis variable is the client's occupation. Attached to new applicant records in the ES services file is the client's DOT code, known as the ES registration DOT code, which typically is the code associated with the client's occupation in his/her last job. An algorithm developed by the Upjohn Institute translated the DOT codes to SOC codes. The latter have become the standard for occupation classification (O*Net) and are used in the FDSS models.

As discussed in section 3, descriptive characteristic information is available in the combined/master intake file; however, based on recommendation from Georgia DOL, descriptive characteristic information used in the FDSS statistical models were taken from the ES services file. These include: age, gender, race, veteran status, handicap/disability status for veterans and non-veterans, education attainment, whether the client was recently separated from the military, welfare receipt status and whether the client is classified as economically disadvantaged. Other characteristics include migrant or seasonal farm worker status, food stamp receipt, and whether the client is a dislocated worker, currently employed or in school. Lastly, the ES services file enabled identification of the beginning and end of services received under UI profiling and the Claimant Assistance Program (CAP), which is a Georgia specific system for early referral to reemployment services of job separated UI beneficiaries.

8) Job Orders File

The Georgia data system also contains useful information about job orders which employers place with the ES. Important variables in the file include a unique job order number, the date the order was posted, the county where the job is located and variables which describe the job. The descriptive variables include months of experience and the education/training level required, minimum age requirements, hours per week and the salary being offered. Other important variables include the DOT code that applies to the job and the SIC (standard industry code) of the employer offering the job.

9) JTPA/WIA Data

The data contained within the JTPA and what later became the WIA data files form the foundation of the training referral portion of FDSS. Separate data files are maintained for JTPA/WIA which was especially needed before the "one-stop" approach where only a small percentage of JTPA/WIA clients would be found in ES related data files. Currently, WIA clients can be found in ES data files with the addition of WIA registration and exit records in the ES services data file. However, separate files are still needed to provide more detailed information about the services WIA clients receive as well as important descriptive characteristic information.

The primary key for the JTPA/WIA data is the client's social security number and the date of his/her eligibility for JTPA/WIA programs. Also included are records of all services received which are identified by a project ID code and a date the project/service began. Descriptive information about each client includes age, gender, race, county of residence, veteran status including disability status for veterans as well as non-veterans. Other characteristics include educational attainment, current school status, draft compliance and a vector of variables which describe any barriers to employment. These include skills deficiencies, learning disabilities, school drop-out, offender or felony status. Other barriers include homelessness, youth below age appropriate education level, youth pregnancy or parent status, displaced home maker, substance abuse, long-term TANF (welfare) status, or whether the client lacks significant work history. Data on other employment barriers in the data include lack of transportation, child care requirements, presence of pre-school children and a field for "other" barriers.

The JTPA/WIA data also contain information about past employment for all clients and include start and end dates of work, salary, usual hours worked and whether the company is the client's separating employer. Follow-up (post-termination) employment information is also collected.

The JTPA/WIA data also include information about public assistance receipt including TANF, food stamps, refugee assistance, general assistance and SSI. The data also include variables related to assistance for transportation, health or family care, housing, counseling, needs based assistance, and other assistance received.

Lastly, an important date in the JTPA/WIA data is the date the client terminated from the program and the reason for the termination.

Though not used in the current models for FDSS, the JTPA/WIA data files also contained information concerning Welfare-to-Work. That data were not used because of small sample sizes. FDSS models developed for economically disadvantaged and welfare clients used the ES data rather than welfare-to-work because the ES data provided much greater sample size and were considered more broadly applicable to the range of disadvantaged clients who were participating in services provided by the employment service.

10) Individual Wage and Employer ES202 Data

The expected labor market outcome and services/training referral information presented to each client in FDSS has wage data as its foundation. The wage data received by the Upjohn Institute from Georgia State University is state-wide data for all individuals in the state of Georgia and for all employers reporting wages. The data are quarterly and begin in the first quarter of 1993 and currently extend through the fourth quarter of 2001 (receipt of the first and second quarters of 2002 is pending as of this writing).

For each SSN, the individual wage data file contains the amount of wages reported in the quarter for all employers reporting wages for the given individual. Therefore, for persons with multiple employers, multiple records exist for a given SSN and year/quarter combination. The critical variables in the individual wage file include the social security number of the client, the year and quarter of the observation, an identifier which uniquely identifies the employer and the amount of wages reported by that employer for that SSN. From this file, the number of employers a client had in a quarter is known as well as the total wages reported for that client by all his/her employers.

The employer identifier attached to each SSN/record in the individual wage file is merged with the employer ES202 data. The important variable taken from this merge is the industry (SIC) code of the employer for the given year and quarter. Other data in the ES202 data file include the number of employees associated with the employer ID for the given year and quarter and the total wage bill reported by that employer.

11) Labor Market Information (LMI)

The FDSS models in Georgia use two variables related to local labor market information—the number employed and the number unemployed. The data are monthly for each county in the state of Georgia with the FIPS code serving as the unique identifier for each county. The LMI data should extend backward in time to a minimum of 18 months prior to the start date of client inflow used for model estimation. For example, the initial models developed for Georgia have used data for clients entering UI and other programs beginning January 1, 1996. This requires LMI data going back to at least July, 1994.

The LMI data in Georgia typically have a three month lag in availability with September/October 2002 being the most recent month available as of this writing (January 2003).

12) Data Transfer Issues

The transfer to the Upjohn Institute of the large administrative files maintained by Georgia DOL was not a simple task. Georgia DOL uses IBM 3480/3490-E cartridges for its backups of the IBM, EBCDIC, VSAM administrative files. The Upjohn Institute has an ASCII based, Compaq Open VMS Alpha system which uses DLT IV tapes as its backup media.

Fortunately, Western Michigan University in Kalamazoo, Michigan has both types of computer systems with the capability to transfer data between the two using FTP. Therefore, the Upjohn Institute contracted WMU to read the IBM 3480/3490-E cartridges, transfer that data to the ASCII-based, Open VMS system and copy that data to DLT tapes.

One very important aspect of this data transfer problem was that some of the files contained packed/compressed data fields. The compressed/packed data fields do not transfer

properly when using FTP to go from the IBM world to the ASCII based Open VMS world. Therefore, additional time and expense was required for WMU to decompress the data fields on their IBM system using SAS, re-write the entire file, and then transfer the data to the Open VMS system for writing in ASCII to DLT tapes.

	Table 1 Variable Summaries			
Variable	**Return to Prior Industry**	**Earnings**	**Employablity**	**Source**
Intercept	x	x	x	na
Natural log of maximum prior quarterly earnings	x	x		Individual wage file
Prior Wages, 5 Qtrs Before Ref Date			x	Individual wage file
Number of Employers, Qtr T-5			x	Individual wage file
UI client	x	x	UI & ES	UI data file
Exhausted Prior UI Claim		x	x	UI data file
Age as of Reference Date	x	x	x	CI100 (CI Master file)
Age Squared	x	x	x	CI100 (CI Master file)
Months Tenure on Prior Job		x	x	CI110 (CI Work History file)
Months Tenure Squared		x	x	CI110 (CI Work History file)
Education, Less than High School	x	x	x	CI100 (CI Master file)
Education, GED	x	x	x	CI100 (CI Master file)
Education, Some College	x	x	x	CI100 (CI Master file)
Education, Bachelor Degree	x	x	x	CI100 (CI Master file)
Education, Advanced	x	x	x	CI100 (CI Master file)
Education, currently enrolled in school	youth & disadv	x		JS200 (ES Services file)
Youth, Ages 14 through 21	diff. model	diff. model	x	CI100 (CI Master file)
Veteran	x	x	x	CI100 (CI Master file)
Dislocated Worker	x	x	x	JS200 (ES Services file)
Welfare Recipient	diff. model	diff. model	x	JS200 (ES Services file)
Economically Disadvantaged	diff. model	diff. model	x	JS200 (ES Services file)
Employed currently	x	x		JS200 (ES Services file)
Has No Driver's License			x	CI120 (CI Profile file)
Available for Rotating Shifts			x	CI120 (CI Profile file)
County monthly unemployment rate (t-3)	add	x	add	LMI Data file
County annual empl. growth (month data) (t-3)-(t-15)		x	add	LMI Data file
Prior Industry: Ag, Forestry, Fish	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: Mining and Construction	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: Trans, Comm, Utilities	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: Wholesale Trade	youth & disadv	youth & disadv	x	ES202 Employer file

9

	Table 1 Variable Summaries			
Variable	**Return to Prior Industry**	**Earnings**	**Employablity**	**Source**
Prior Industry: Retail Trade	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: FIRE	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: Services	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Industry: Public Admin	youth & disadv	youth & disadv	x	ES202 Employer file
Prior Occupation: Management, Business, Financial	x	x	x	JS200 (ES Services file)
Prior Occupation: Professional and Related	x	x	x	JS200 (ES Services file)
Prior Occupation: Services	x	x	x	JS200 (ES Services file)
Prior Occupation: Sales and Related Occupations	x	x	x	JS200 (ES Services file)
Prior Occupation: Office and Admin. Support	x	x	x	JS200 (ES Services file)
Prior Occupation: Farming, Fishing and Forestry	x	x	x	JS200 (ES Services file)
Prior Occupation: Construction and Extraction	x	x	x	JS200 (ES Services file)
Prior Occupation: Install, Maintenance and Repair	x	x	x	JS200 (ES Services file)
Prior Occupation: Transport and Material Moving	x	x	x	JS200 (ES Services file)
Field Service Office: DeKalb			x	JS200 (ES Services file)
Field Service Office: Gwinnett			x	JS200 (ES Services file)
Field Service Office: North Metro			x	JS200 (ES Services file)
Field Service Office: South Metro			x	JS200 (ES Services file)
Field Service Office: Cobb/Cherokee			x	JS200 (ES Services file)
Reference Date in 2nd Quarter	x	x	x	See reference date definition (*1)
Reference Date in 3rd Quarter	x	x	x	See reference date definition (*1)
Reference Date in 4th Quarter	x	x	x	See reference date definition (*1)
Ref Date 3 Qtrs After Max Wage		x		See reference date definition (*1)
Ref Date 4 Qtrs After Max Wage		x		See reference date definition (*1)
Ref Date 5 Qrtrs After Max Wage		x		See reference date definition (*1)
Days Left in Current Quarter		x		See reference date definition (*1)

(*1) The reference date for UI clients is the benefit year begin date of the client's most recent valid UI claim. For ES clients, the reference date is the transaction date associated with the most recent of: 1) New applicant record in JS200 ES services file, 2) Renewal record in the JS200 ES services file, or 3) Applicant active at beginning of current reporting year (July 1, yyyy) record in the JS200 ES services file.