

A stylized map of New York State is positioned behind the main title. The map is filled with a gradient of green, matching the overall color scheme of the document. The background of the top section features a pattern of thin, light green lines.

New York **Makes Work Pay**

Developing a path to employment for New Yorkers with disabilities

www.NYMakesWorkPay.org

Report

Data Integration for the Purpose of Tracking Employment Outcomes of People with Disabilities in New York State

Preliminary Assessment

February, 2010

Lars Vilhuber
Sarah von Schrader
Employment and Disability Institute

This document is being published as part of the New York Makes Work Pay Project, a Comprehensive Employment Services Medicaid Infrastructure Grant funded by the U.S. Department of Health and Human Services, Centers for Medicare and Medicaid Services (CMS) to the New York State Office of Mental Health (OMH) and its management partners the Blatt Institute at Syracuse University and the Employment and Disability Institute (EDI) at Cornell University. The New York Makes Work Pay Initiative is currently funded for calendar years 2009 and 2010 and will provide an array of services to individuals with disabilities and the agencies and advocates that serve them, helping to remove obstacles to work and pave the way to self-supporting employment.

www.NYMakesWorkPay.org

Data integration for the purpose of tracking employment outcomes of people with disabilities in New York State

Preliminary Assessment

Lars Vilhuber, Sarah von Schrader, (EDI/ILR, Cornell University)

February 11, 2010

1. Purpose of this Project and Report

Nine New York state agencies track employment, earning and health-related data for New Yorkers with disabilities. Each agency has individual case management systems; however, there is currently no single method for integrating all of these data systems. Such an integrated system would provide a much more comprehensive and accurate picture of the progress of the comprehensive employment system toward successful employment of people with disabilities. It would allow the measurement of key outcomes and indicators across state agencies and inform decision makers as they implement policies related to employment, rehabilitation services and health care supports.

In this report, we present a brief analysis of our cross-agency inventory of data management systems, as well as an inventory of other known databases that integrate agency case management data. In addition we present:

- A listing of specific employment outcome variables that minimally should be considered for collection going forward by the individual agencies,
- The potential for one of the existing systems to serve as a potential shared data exchange for this information,
- The potential for using shared data legislation that covers DOL data as a vehicle for providing a more robust snapshot of employment outcomes of New Yorkers with disabilities served in the various systems.

Since the first draft of this report, New York state agencies have come to an agreement on the use of New York Department of Labor's One Stop Operating System (NYOSOS) to be used as the platform for an integrated data system. We welcome this agreement, and strongly agree with that decision.

2. Scope

The focus of the Medicaid Infrastructure Grant is on enhancing employment outcomes for people with disabilities in New York State. The nine state agencies in-scope for this report are:

- New York State Office of Mental Health (OMH)
- New York State Office of Mental Retardation and Developmental Disabilities (OMRDD)

- New York State Office of Vocational and Educational Services for Individuals with Disabilities (VESID)
- New York State Commission for the Blind and Visually Handicapped (CBVH)
- New York Department of Labor (NYSDOL)
- New York State Office of Alcoholism and Substance Abuse Services (OASAS)
- New York State Office of Temporary Disability Assistance (OTDA)
- New York State Division of Veterans' Affairs (VA)
- New York State Department of Civil Service (CS)

All agencies already track by some method the employment outcomes of their clients.

However, most of these methods are based on interviews at points-of-contact, and follow-up is difficult, impossible, or too expensive. These methods do not provide a full picture of employment outcomes, particularly after the client's case is closed. One agency, OMH, in addition, uses a biennial survey of all their clients to assess current employment outcomes.

3. Participating agencies and databases

The following agencies provided the authors with information for this report:

- New York State Office of Mental Health (OMH) – NYISER (New York Interagency Supported Employment Reporting), covering (supported employment) case management by OMH, CBVH, OMRDD, and VESID (various emails and phone interviews) – database schema
- New York Department of Labor (NYSDOL) – NY One Stop Operating System (NYOSOS) case management system – description of select database elements in spreadsheet form
- New York State Office of Alcoholism and Substance Abuse Services (OASAS) - copy of client discharge report form (email)
- New York State Commission for the Blind and Visually Handicapped (CBVH) – report on data collected for clients (email)

Additional information was independently obtained on

- RSA-911 (based on a research version of the file, with additions based on PD-07-01, dated Oct 5, 2006, VESID and CBVH report data in RSA-911)
- New York State Office of Vocational and Educational Services for Individuals with Disabilities (VESID) – CaMS case management system
- NYSDOL wage records (WR) - (based on first author's extensive personal experience at the U.S. Census Bureau)

We do not have information at this time about the databases used by VA, OTDA, or CS, nor do we have database schema for OASAS' Information Processing Management and Evaluation System (IPMES) or CBVH's internal database structure.

In general, we have not built a proper database description, in particular with respect to agency-specific unique own and foreign keys. Rather, we have focused on data elements that can be used to combine records across agencies.

4. General data description

Table 1 provides a tabular overview of common and unique database features for each database. Further description of what the authors have surmised about common and unique data elements is presented below. It is important to highlight that this is a partial overview only – for reasons of practicality, the authors did not have access to all data elements of all databases. In particular, NYOSOS is a very large and complex database system, and the elements described in the overview here are only a small subset of the actual database schema.

4.1 *Common data elements and identifiers*

In general, all databases have personal identifiers that can be used for linking. NYISER, NYSDOL-WR, NYOSOS, CBVH, and CaMS all have the Social Security Number (SSN) as a primary or secondary unique ID. OASAS discharge reports have partial SSN, plus date of birth and partial names. We expect that the use of standard off-the-shelf statistical matching software would be sufficient to match persons based on those details to the other databases with reasonable reliability.

All the above databases also already contain some element of employment tracking; we have some level of information on this tracking for:

- NYISER: JOBS table, at quarterly level (?)
- NYSDOL-WR: core wage record, at quarterly level
- NYOSOS: presumably contains employment tracking, but was not part of the information received, which focused on disability-related elements
- VESID CaMS: STATUS_BY_MONTH (?) and OCCUPATIONS table (frequency?)
- OASAS and CBVH: client discharge report and/or internal database (presumably required for CBVH RSA-911 reporting)

By far the most complete coverage is expected to be present in the NYSDOL-WR database, since it is based on mandatory employer reports for the unemployment insurance system.

4.2 *Unique elements and missing elements*

- NYSDOL-WR does not have occupation or education. It provides a very long and complete within-state employment history. Employment status is current at each quarter, with some lag. NYSDOL-WR does not tabulate hours of work per week, only earnings.

- CaMS has indicator of “competitive employment,” occupation (DOT coded), industry (SIC, although we do not know if NAICS coding has been introduced or back-coded for consistency).
- OASAS, CaMS have variables on living arrangements.
- NYOSOS, CBVH, OASAS have variables on the medical status and history relevant to each agency
- OASAS does not have occupation, or employment history beyond completed length of current employment at time of discharge.
- NYISER has indicator of “integrated employment” (we do not have a definition) and employment model, job termination reasons.
- NYOSOS has indication of the presence of employment barriers, but details were not available to the authors at this stage
- Most non-NYSDOL-WR databases have information on (presumably self-reported) language and ethnicity
- RSA-911 and databases reporting to it (CaMS, CBVH) have information on jobs only at the time of closure (and first contact). RSA-911 has no information available for open cases, although agency databases may have updated status information as of the time of last contact.

4.3 *Coverage issues*

In general, the individual case management databases provide many rich details on the case or the individual. Issues may arise with incomplete reporting, since few fields are required. Of particular concern is incomplete coverage, both in terms of non-response or erroneous response by individuals, but also in the brevity of the follow-up period (if any). Any improvement on the status quo does not come from the availability of a common database structure, but rather from coverage improvement. The broadest coverage on jobs is obtained by directly accessing the NYSDOL-WR database. Improved timing information (on job entry and exit) may be gained for particular jobs by referring to the agency-specific databases, which in general ask for a specific job start and end date. On the other hand, improved coverage on other person characteristics may be gained by joining all databases.

In the next phase of this project, we will re-interview each agency with respect to the data content, rather than the data structure, to assess the actual coverage provided by each agency’s database.

5. Known challenges

5.1 *Reconciling different data sources*

Challenges arise from several factors. From a data content perspective, consistency between administratively reported elements (NYSDOL-WR) and client-reported elements (others) needs to be assessed. Research into the correspondence between survey-reported and administratively reported jobs¹ has shown the correspondence to be imperfect at best. When combining disparate data sources, algorithms will need to be developed in order to disambiguate and reconcile them. This is very much an empirical exercise, and will require access to the data in order to explore and resolve.

5.2 *Data sharing: legal aspects*

The second challenge arises from the need to share data. Restrictions (e.g., NYS law, HIPAA) may impose constraints. In general, program-supporting data sharing is permissible, and has been done in the past. A later version of this report will focus on the legal possibilities, and possible justifications that can be made to allow for sharing of data in support of suggested new statistics. This requires specific statistics to be elaborated, so that possible justifications can be taken into account. We expect this to be an iterative process.

It should be noted that other data sharing initiatives are under way within the state of New York. We note in particular a pilot “Data Sharing Assessment Project” involving NYSDOL and the NYS Insurance Fund (NYSIF), which could potentially be of use to the other agencies in this report at a later stage.²

6. Previous successful uses of merged data

OMH regularly receives data from the Veteran Administration, for administrative purposes. It is unknown to the authors whether employment information is exchanged at that time.

In 2005, VESID provided NYSDOL research staff with an extract of their respondents under 34 CFR 361.38(d). NYSDOL then tracked those respondents/consumers in the NYSDOL-WR database, and tabulated the employment outcomes. In particular, the available employment criterion was “one quarter of work in the next year,” with employment rates much higher than

1 Martha Stinson (2002), “Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data,” U.S. Census Bureau LEHD Technical Paper TP-2002-24. Jesse Bricker and Gary V. Engelhardt (2008) “Measurement error in earnings data in the Health and Retirement Study.” *Journal of Economic and Social Measurement*, 33(1). Kristin McCue, Sule Celik, Chinhui Juhn, and Jesse Thompson (2010), “Understanding Earnings Instability: How Important Are Employment Fluctuations and Job Changes?” presentation at 2010 Meetings of the Allied Social Science Associations.

2 <http://www.labor.state.ny.us/cioshares/activeprojectlist.shtm> as of 2010-01-18, 12:01PM.

captured by VESID's tracking system (email correspondence with Frank Coco, 11/2/2009). The stability of employment was also explored. The example highlights the benefits of cross-agency data access for improved tracking of employment outcomes.

7. Minimal database structure for subsequent analysis

We used the information provided by the agencies to construct a simplified virtual view of the cross-agency data elements as a single database (see *Table 1: Inventory of common data elements across agency databases, and suggested common elements*). This database obviously does not presently exist, but the goal here is to identify a small (and thus hopefully feasible) subset of the data available across all agencies that can be combined to create a dataset that is more useful to current administrators than the sum of its parts.

Clearly, all agencies benefit from common information on the demographics of their clients, and the existence and structure of both NYISER and NYOSOS attest to that. However, many additional case management structures collect this information in a myriad of different, slightly incompatible or inconsistent ways. A standardized table or database of person characteristics would greatly enhance both access to clients, and later standardization of contact with clients. This table would contain gender, race, ethnicity, language, education (formal and observed, such as reading and math skills), family status, veteran status, and other common demographic characteristics, captured at different points in time. Note that the concepts of “at closure” and “at application” no longer apply in a consolidated database view; rather, discrete data capture dates would be recorded, which can be linked to actions on a client’s case in order to identify a particular agency’s case history. Within this set of variables, the residential address could be either queried, looked up within, or validated against other state records (tax rolls), if not already being done. Many of the demographic characteristics could also be derived from DMV records for a significant fraction of clients, if legally allowed.³ We would expect that this set of variables would be the most complete.

A second group of variables is also commonly recorded: Participation in various support schemes not provided by the agencies sharing data. These programs such as SSI/SSDI/Workers Comp/etc., but also the different forms of medical insurance (Medicaid/Medicare, private, other coverage, no coverage). This set of variables may not be complete in any or all of the underlying agency databases based on data provided by the client, depending on the legal requirement to provide the agencies with such information. However, it can presumably be filled in or looked up when access to tax records or Medicaid/Medicare data is permitted for program administration.

3 Numerous states use access to their DMV database and to their tax records for precisely this purpose.

The third group of variables is direct employment related variables. Three categories of such variables are commonly found. The first category relates directly to the basic characteristics of employment held at various times: earnings and hours at one or more jobs. These can be partially obtained from NYSDOL-WR data; however, the NYSDOL-WR data do not contain hours, nor do they contain other more detailed information such as occupation or part-time status. On the other hand, they do contain or can be linked to employer location and sector of activity. While some of the agency databases do contain client-reported work location, the employer's sector of activity is not reported in any of the databases.⁴ Depending on the focus of the agencies, the basic earnings information may be sufficient.

The second category of employment-related variables relates to the methods of obtaining and losing employment, and are typically not (fully) captured by other administrative data. It is in this category that we also find the largest variation across agencies, and some notable absences. At a minimum, the reasons for losing jobs, and the methods of finding jobs (whether provided by the agencies or not) should be captured. Reasons for finding jobs are currently limited to activities and services provided by the agencies.

The third category is barriers to work or work impediments. For the agencies involved in this project, that typically means impediments or issues related to disabilities, although the concept can be (in particular for NYSDOL) broader. NYOSOS already captures a variety of family, health, and education-related variables, but does not capture some of the more detailed job characteristics as they relate to integration or support. Whereas NYISER captures job termination variables, and the type of job (integrated/sheltered/etc.) provided through one of its contributing providers, it does not seem to capture other methods of finding jobs.

The data collection done by the agencies is longitudinal, that is, data is typically collected at several points in time. It is clear that improvements can be immediately achieved by reducing the number of repeated queries for common variables. Thus, a consolidated database would require as much burden on the client at first contact, but much less upon subsequent contact, providing the frontline service provider with more time to elicit more detailed information, or reducing the burden on the client. Updating information of repeat clients of the system, rather than individual agencies, could also provide an excellent mechanism to capture information otherwise hard to obtain, such as job finding tools outside of the agency program, or occupation in jobs held since previous contact.

However, it is very critical that a consolidated database also provide a mechanism for individual agencies to fill in the specialized information that is not of high value to other agencies, but vital for service provision within that agency (e.g., detailed diagnostics on

4 Note that work location in the NYSDOL-WR data may pose challenges for large employers, since the worker wage records are not directly associated with the establishment location.

blindness or substance abuse). This is likely to be less an issue of database design (the internal structure might simply be adjusted to contain additional linked tables), and more a feature of the user interface to the database (for instance, by identifying a service provider or agent connected to a particular agency, additional screens are offered, or the workflow is modified).

In sum, our review suggests that the following benefits are likely to be obtained by using a consolidated database design:

- Better (more complete and accurate) data on common clients. In particular, standardized information on race and ethnicity, as well as improved (standardized and geo-coded residential address information) will allow for a common view on the distribution of client characteristics.
- Better measurement of multiple and cross-agency client contacts.
- Through access to NYSDOL-WR, better and more complete information for clients both before and after contact with the participating agencies. In particular, information on post-agency labor market information (employment and earnings) will be markedly improved. Information on the employer geographic location and sector may allow for improved planning of agency activities.

The overview also suggests a small number of variables for which collection could be implemented in order to improve service provision:

- Collection of information on non-agency means of obtaining employment.
- Consistent methods across all agencies of collecting information on reasons for job loss.
- Finally, the consolidated database requires a small number of critical features to be adequately addressed. The structure must accommodate agency-specific elements that go beyond the common elements. Based on the current usage by agencies of their individual case management systems, these elements might be related to job support methods, diagnoses, or particular histories.
- The structure must accommodate future enhancements in a straightforward fashion. We outline several employment-related variables that could conceivably be added to such a database in the near future; other elements not currently captured include self-employment and entrepreneurial outcomes. While the initial focus should remain tightly centered on the integration aspects, expandability needs to be a feature of the structure, not an afterthought.
- The support structure for the database needs robust data interchange features with existing databases (upload/download), ideally in a largely automated fashion. This is important for two reasons. First, in a transitional phase, both a new integrated database and legacy case management systems need to co-exist, without loss of data or functionality. A robust data interchange mechanism reduces the burden of maintaining the two systems in parallel, and reduces the risk of data loss or inaccuracies. Second, for

analytical and reporting purposes, the migration of historical data into the new system is desirable.

8. Expanding set of variables

Additional indicators may be internally computed at NYSDOL, or captured as noted above as updates to a cross-agency database.

The longitudinal nature of the NYSDOL-WR database is ideal to compute information on the duration of jobs or the stability of post-contact employment. The described database structure does not contain the combining rules, only the basic structure. We have in mind the following menu of combining rules. The usefulness of the statistics derived from such combining rules needs to be tested and assessed by the participating agencies in terms of their usefulness and their (computational) cost.

- The consolidated database contains timing information on last contact by each agency. Information at next contact about a complete job history may be incomplete (typically, only information on the last job is collected). Filling in holes through the WR data allows the agency to have a more complete between-contact employment history.
- For each job, identify the associated industry (based on QCEW linked to WR data). This will allow the assessment of the industry distribution, as opposed to the occupational distribution of jobs (which can be performed on the consolidated database without access to the WR data, conditional on the completeness of the data).
- For each job, assess distance between client's home (available as of time of agency contact) to the employer's workplace (or possible workplaces). This will allow the agencies to assess whether they are successfully reaching out to local employers when mobility is an issue.
- For each worker, assess post-contact employment stability in terms of jobs/industries/location.

9. Next steps

The usefulness of a consolidated database, and how close the current system is to such a consolidated database, can only be assessed in conjunction with actual coverage and completeness statistics in the current set of disjunct or partially joined databases. While the structure of the database may contain all the critical elements, if the actual elements are missing because not entered at client contact, the value of consolidation becomes higher. We will thus propose to the participating agencies a number of database queries, to be executed on the internal data. These queries will assess how complete the coverage of existing variables in the databases is, and provide an indication to both benefits and challenges of consolidation.

Furthermore, we propose to construct a second set of database queries that will explore the usefulness of enhanced statistics, as briefly outlined above. Involving a small set of participating pilot agencies, we suggest a number of currently not available statistics that can be computed with the NYSDOL-WR database for (a subset of) the pilot agencies' clients. A short report that can be circulated within the group of participating agencies will be the preliminary output of such an exercise, and a more informed suggestion as to the direction of database consolidation at the end of the activity period.

In both cases, access (possibly limited) to the internal data is required. The queries can be executed by agency staff, if resources are available. Alternatively, Cornell provides a secure data facility, the Cornell Restricted Access Data Center (CRADC), and can provide guarantees compliant with 34 CFR 361.38(d), allowing for the temporary transfer of the confidential data for analytic purposes to Cornell facilities. Our staff could then perform consistent queries across all databases, and provide the pilot reports outlined above without further draw on agency resources.

The agencies have agreed upon a timeline for the implementation of a consolidated database, which includes (a) finalization of a database structure (b) addressing any outstanding legal issues and (c) implementation within the next year. We will work with the relevant working groups to assist in all of these tasks, as needed.

Table 1. Inventory of common data elements across agency databases, and suggested common elements

	CBVH	NYOSOS	NYSDOL- WR	VESID/RSA911	OASAS	NYISER	<i>Suggested COMMON</i>	<i>Notes</i>
Potential linking information								
<i>Alternate identifier</i>								
SSN	x	x	x	x	last 4 digits	x	x	Replaces SSN
Last name	x	x	x	x	o	x	x	
First name	x	x	x	x	o	x	x	
DOB	x	x	o	x	x	x	x	
Geographic information (with history)								
may be derived								
City	x	x		x	o		x	
County	x	x		x	x	x	x	
Zip	x	x		x	x	x	x	
Street address							?	
Demographic information								
Race	x			x	x		x	Standardized
Ethnicity	x			x	x	x	x	Standardized
Hispanic					x	x	x	Standardized
Primary language	x	ESL flag		o	x	o	x	Standardized
Gender	x	x		x	x	x	x	
Other non-health information (time-varying)								
Veteran status	x	x		x	x		x	
Education	x	x		x	x		x	Standardized
Language proficiency	x	x			x		x	
Family status		x		x	x		x	
SSI/SSDI	x			x			x	
Medicaid/Medicare coverage	x			x			x	

Table 1. Inventory of common data elements across agency databases, and suggested common elements

	CBVH	NYOSOS	NYSDOL- WR	VESID/RSA911	OASAS	NYISER	<i>Suggested COMMON</i>	<i>Notes</i>
Employment related variables								
Earnings	x	x	x	x	x	x	x	<i>Standardized</i>
Hours				x		x	x	
Part-time status				x	x		x	
Complete in-state employment status			x				x	
Occupation	x			x			x	
Industry			x				x	
Location	x		x				x	
Reasons for leaving	x			x			partial	
Methods of finding job				partial			new	
Special employment (integration)				x		x	x	
Benefits	x						partial	

Note: this table is based on a partial description of the databases, and is not meant to provide a full description.

Note: x denotes that a data element is present. o denotes that it is absent. A blank cell means that the authors had no information.

www.NYMakesWorkPay.org

Contact Information

Employment and Disability Institute
ILR School / Cornell University
201 Dolgen Hall
Ithaca, New York 14853-3901
607.255.7727 (voice)
607.255.2891 (tty)
607.255.2763 (fax)
ilr_edi@cornell.edu
www.edi.cornell.edu

Partnering Organizations

New York State Office of Mental Health
Employment and Disability Institute (Cornell University)
Burton Blatt Institute (Syracuse University)

This publication is available in alternate formats. To request an alternate format, please contact us using the information provided above. It is also available online in both text and pdf format. They are located at www.NYMakesWorkPay.org.