

4-2006

Data Integration and Storage: Managing and Using Home and Community-Based Services Data for Quality Improvement

Carolyn E. Gray MPH

University of Southern Maine, Muskie School of Public Service

Maureen Booth MRP, MA

University of Southern Maine, Muskie School of Public Service

Follow this and additional works at: <https://digitalcommons.usm.maine.edu/aging>

 Part of the [Health Policy Commons](#), [Health Services Research Commons](#), and the [Public Policy Commons](#)

Recommended Citation

Gray, Carolyn E. MPH and Booth, Maureen MRP, MA, "Data Integration and Storage: Managing and Using Home and Community-Based Services Data for Quality Improvement" (2006). *Disability & Aging*. 118. <https://digitalcommons.usm.maine.edu/aging/118>

This Briefing Paper is brought to you for free and open access by the Cutler Institute for Health & Social Policy at USM Digital Commons. It has been accepted for inclusion in Disability & Aging by an authorized administrator of USM Digital Commons. For more information, please contact jessica.c.hovey@maine.edu.

Data Integration and Storage
***Managing and Using Home and
Community-Based Services Data for
Quality Improvement***

April 2006



UNIVERSITY OF
SOUTHERN MAINE



Rutgers Center for
State Health Policy

April 2006

Discussion Paper

Community Living Exchange

Funded by Centers for Medicare & Medicaid Services (CMS)

Data Integration and Storage: Managing and Using Home and Community-Based Services Data for Quality Improvement

Carolyn Gray
Maureen Booth

UNIVERSITY OF SOUTHERN MAINE
Muskie School of Public Service

This document was prepared by Carolyn Gray and
Maureen Booth of the Muskie School of Public Service.

Prepared for:



Rutgers Center for
State Health Policy

Susan C. Reinhard & Marlene A. Walsh



Robert Mollica

The Community Living Exchange at Rutgers/NASHP provides technical assistance to the Real Choice Systems Change grantees funded by the Centers for Medicare & Medicaid Services.

We collaborate with multiple technical assistance partners, including ILRU, the Muskie School of Public Service, National Disability Institute, Auerbach Consulting Inc., and many others around the nation.

This document was developed under Grant No. 11-P-92015/2-01 from the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services. However, these contents do not necessarily represent the policy of the U.S. Department of Health and Human Services, and you should not assume endorsement by the Federal government. Please include this disclaimer whenever copying or using all or any part of this document in dissemination activities.

Acknowledgments

The authors would like to thank Kim Murray for contributing to this paper and providing insight into the technicalities of data integration. We would also like to thank our Muskie colleagues Julie Fralich, Eileen Griffin, Taryn Bowe, Stuart Bratesman, Bob Keith and Lisa Marie Lindenschmidt for reviewing drafts of this paper and offering helpful assistance.

Early in the design of this paper, a small group of 2003 Systems Change QA/QI grantees volunteered to help conceptualize the design of this paper and comment on drafts. Our thanks to the members of this subgroup for guiding us in the development of the paper and insisting that it translate arcane technical terms into lay language. Members included Suzzanne Freeze (OH), Jolene Kohn (MN), Terri Richards (TX), Bruce Rosen (NY), Angela Berger (GA), Darlynn Thomas (SC), Ruth Burgess (WV) and Kraig Nagel (CA). Special gratitude to Florida, Minnesota, and Ohio for allowing us to share the designs of their data integration plans in Appendix B.

Table of Contents

Managing and Using Data for Quality Improvement	1
Focus and Purpose of Data Use and Management Series	1
Data Integration	2
Overview of Data Integration	2
Data Identification	3
Methods for Linking Data	5
Storing Linked Data	6
Assessing Readiness to Integrate Data	8
Summary	10

Tables

Table 1. Quality Indicators that Require Linked Data.....	3
Table 2. Types of Primary and Secondary Data Source for HCBS QA/QI.....	4
Table 3. Issues to Consider when Assessing Readiness for Data Integration	8

Figures

Figure 1. Linking Long Term Care Data	6
--	---

Appendices

Appendix A. Data Source Assessment Worksheet.....	A-1
Appendix B. State Examples.....	B-1
Appendix C. Resources for More Information.....	C-1
Appendix D. Glossary of Terms.....	D-1

Managing and Using Data for Quality Improvement

The Data Management and Use Series represents the final installment in a group of papers synthesizing the ideas and practices of states as they improve the quality of home and community based services (HCBS) and supports for older persons and persons with disabilities.

In 2003, the Centers for Medicare & Medicaid Services (CMS) awarded grants to 19 states to enhance their quality management (QM) programs for HCBS programs.¹ CMS contracted with the Community Living Exchange Collaborative² to assist states in their grant activities by promoting information exchange and facilitating discussions on topics of common interest. As part of its work with the Community Living Exchange Collaborative, the Muskie School of Public Service, together with grantee states, identified three priority topics for working papers:

1. Quality Management (QM) Roles and Responsibilities
2. Discovery Methods for Remediation and Quality Improvement
3. **Managing and Using Data for Quality Improvement**

The Data Management and Use Series builds upon the concepts and techniques discussed in the two previous papers and provides additional resources for states as they seek to organize, analyze and report data in a way that informs decision making and supports quality management and improvement.

Focus and Purpose of Data Use and Management Series

As 2003 Quality Grantees move into the third year of their projects, their methods for collecting and automating HCBS waiver data are continuously improving, and program and outcome data are becoming more readily available. One challenge that is frequently articulated by grantees is how to organize, analyze and report this data in a way that is timely, accurate and cost-effective. States are challenged to integrate information from a variety of separate systems and present data in a format that is meaningful, purpose-driven and often dependent on the audience or stakeholder. CMS's requirement that states report data in a way that directly addresses HCBS waiver assurances gives each of these challenges additional weight.

A number of specific issues and questions were identified through monthly conference calls and one-on-one discussions with grantees. These include the following:

- **Performance Measurement:** How do states construct and use performance measures to evaluate HCBS programs?
- **Data Analysis:** How do states validate, clean and analyze waiver data in a way that supports project management and informs decision-making?
- **Data Presentation:** What types of tables, charts and graphics are used to present data, and how does the effectiveness of these formats vary depending on the type of information and/or pattern being conveyed?
- **Reporting:** What types of reports are generated from HCBS waiver data and how do these reports vary depending on the audience and purpose?

¹ QA/QI grantee states include: California, Colorado, Connecticut, Delaware, Georgia, Indiana, Maine, Minnesota, Missouri, North Carolina, New York, Ohio, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Wisconsin, and West Virginia.

²The Community Living Exchange Collaborative is a partnership of the Rutgers Center for Health Policy, the National Academy for State Health Policy and Independent Living Research Utilization. Under contract with the Technical Exchange Collaborative, the Muskie School of Public Service is the lead for providing technical assistance in the area of quality assurance/quality improvement.

- **Data Integration:** How are data from different sources linked to create a better understanding of HCBS performance?

This paper reports on data integration from a program manager's perspective. The paper is not meant to be an exhaustive research document, nor does it single out any one correct approach. The paper is meant to facilitate communication between program units and analytic staff and serve as one reference for states as they continue to improve upon data collection techniques and use this information for ongoing quality management and improvement.

Data Integration

Any quality assurance or improvement activity must have data to support its efforts. Data is often provided through different sources, such as survey results, claims data or provider files. Using data from various sources requires information to be linked. This paper outlines some of the factors program managers may need to consider in identifying strategies for integrating information to support their quality improvement activities. More detailed technical information, such as examples and resources for more information, can be found in the Appendix.

Overview of Data Integration

The term data integration is often used interchangeably to describe any one of the following activities:

1. **Synthesizing information from different data sources.** For example, to compare the number of hours authorized to those actually delivered, data from a service plan must be analyzed against information included in Medicaid claims files.
2. **Integrating two or more files.** Entire files may be linked on an ad-hoc or permanent basis. This type of integration can be illustrated in the case where a state combines multiple databases maintained at a county level into a central data base for aggregate state-wide reporting.
3. **Integrating information systems.** Interfaces may be constructed which allow one information system "to talk" to another. A state, for example, may develop an automatic "feed" from its financial eligibility system into its medical eligibility file to identify participants that fully qualify for a waiver program on an ongoing basis.

The focus of this paper primarily is on the first two types of data integration. For the more technically minded, the Appendix provides examples of states system integration efforts.

Understanding why and how data integration may fit into a quality management strategy is an important consideration. Data integration is not a prerequisite to quality management. It can, however, advance understanding of what is happening in a program and help target and focus quality improvement interventions. For example, as states prepare CMS evidence reports or management reports for internal use, there are many illustrations of where ad-hoc linking or integration could be valuable. Table 1 identifies several quality indicators which require linked data for their calculation.

Table 1. Quality Indicators that Require Linked Data

Name of Indicator	Data Source (Examples)	Data Element or Data Set
Percent of cases where the total hours of services authorized in a care plan for a month is equal to total hours provided in the month by the type of provider specified.	Plan of Care	Authorized hours by type of service
	Medicaid Claims	Paid claims by service
Percent of participants with a particular diagnosis (such as Chronic Obstructive Pulmonary Disease) who were hospitalized during specified time frame (e.g., previous 3 months).	Assessment data	Diagnosis
	Medicaid claims	Paid hospitalization claims

The following sections address issues that program managers may consider as they contemplate whether and what data to integrate. This is not a “how to” guide but can hopefully provide program managers with a working knowledge of the concepts involved in data integration and the framework for discussing these issues with information system staff or external technical experts within their states.

Data Identification

An initial step to integration is the identification of individual data sets, or data elements within a data set, that are of interest. There are several important aspects about your data that will determine whether it is possible to integrate.

Electronic versus paper format: Data that are in paper format must be converted into electronic mode as a prerequisite to integration. For example, paper data may include survey responses mailed back by members, plans of care or incidents/complaints. These will need to be put into a database management system for manipulation later in the integration process. Other data may already be in electronic format such as Medicaid claims, financial and eligibility records.

Primary versus secondary data: Understanding whether you, as a HCBS program manager, have access to the data of interest is a critical step in the integration process. Primary data are those that are collected and controlled by the HCBS program versus secondary data that are maintained by other units, agencies or departments. If the data you want to link are all primary, the issue is principally a technical one on how to link the data. On the other hand, if data are controlled by another entity, it is frequently necessary to develop data sharing agreements specifying the purpose, use and restrictions that may apply when using the data. Table 2 lists possible data sources in each group.

Table 2: Types of Primary and Secondary Data Source for HCBS QA/QI³

Primary Data Source	Secondary Data Source
<i>Person Level</i>	<i>Person Level</i>
Assessment	Medicaid Eligibility file
Level of Care Determination	
Service Plan	
Consumer Satisfaction Survey	
<i>Person and Provider Level</i>	<i>Person and Provider Level</i>
Chart Review	Medicaid Claims
Complaints/Incidents	Adult Protective Services
Caseload	
<i>Provider Level</i>	
Provider Audit	
<i>Program Level (aggregate)</i>	
372 Report	

Level of aggregation: Data must be in a comparable unit of analysis in order to be linked. For example, is the data on an individual person level, or does it provide only summary information? Knowing this upfront will determine whether and what integration may be possible.

- *Person level:* Data are captured about an individual member. For example, assessment data capture information on the “person level.”
- *Provider level:* Data captured about individual providers and/or provider agencies. Record reviews conducted during a provider agency audit are frequently conducted at the provider level.
- *Program level:* Some data may be available only at the program level, meaning that the data set itself cannot provide information on individual members, providers or agencies. The required 372 Report to CMS is an example of program data.

Common unique identifier: When working with person or provider level data, there must be a way to assure that the information from one file belongs to the same person or provider in another file. Using the example from Table 1, there must be a way to link approved hours from the service plan of one individual to the paid claims of that same individual. Most likely in that case, the identifier common to both data sets would be the person’s Medicaid identification number. The Medicaid identification number is the link between these two data sets. Beware of using identifiers that are not necessarily unique, such as a person’s name.

Common definitions: When combining or linking data, it is important to know whether data elements with the same name have the same definition. For example, does “restraint” mean the same thing in both data sets or does it include chemical restraint under one data set and exclude it in another?

Data accuracy: Before integrating data, the quality of data within each data set should be determined and addressed. A full description of the components for assuring data accuracy is addressed in the *Data Quality and Analysis* module of this series.⁴

³ Data sources listed under each category will be unique to each state.

⁴ Fralich, J., Booth, M., and Keith, R. (2006) *Data Quality and Analysis: Managing and Using HCBS Data for Quality Improvement*. Community Living Exchange: Portland, ME.

Appendix A provides a work sheet for use in conducting an inventory of data sources that are of interest to you in your quality management activities.

Methods for Linking Data

There are three basic methods for linking data each of which is described below.

Match merging or using a common identifier/variable that matches between two or more data sources. For example, a Social Security number is used to link eligibility files with claims. Whenever there is an **exact** match across data sets, the data can be linked. The danger of match merging, however, is that oftentimes one digit may be wrongly entered for a Social Security number. You generally end up with fewer “matched sets” than actually exist.

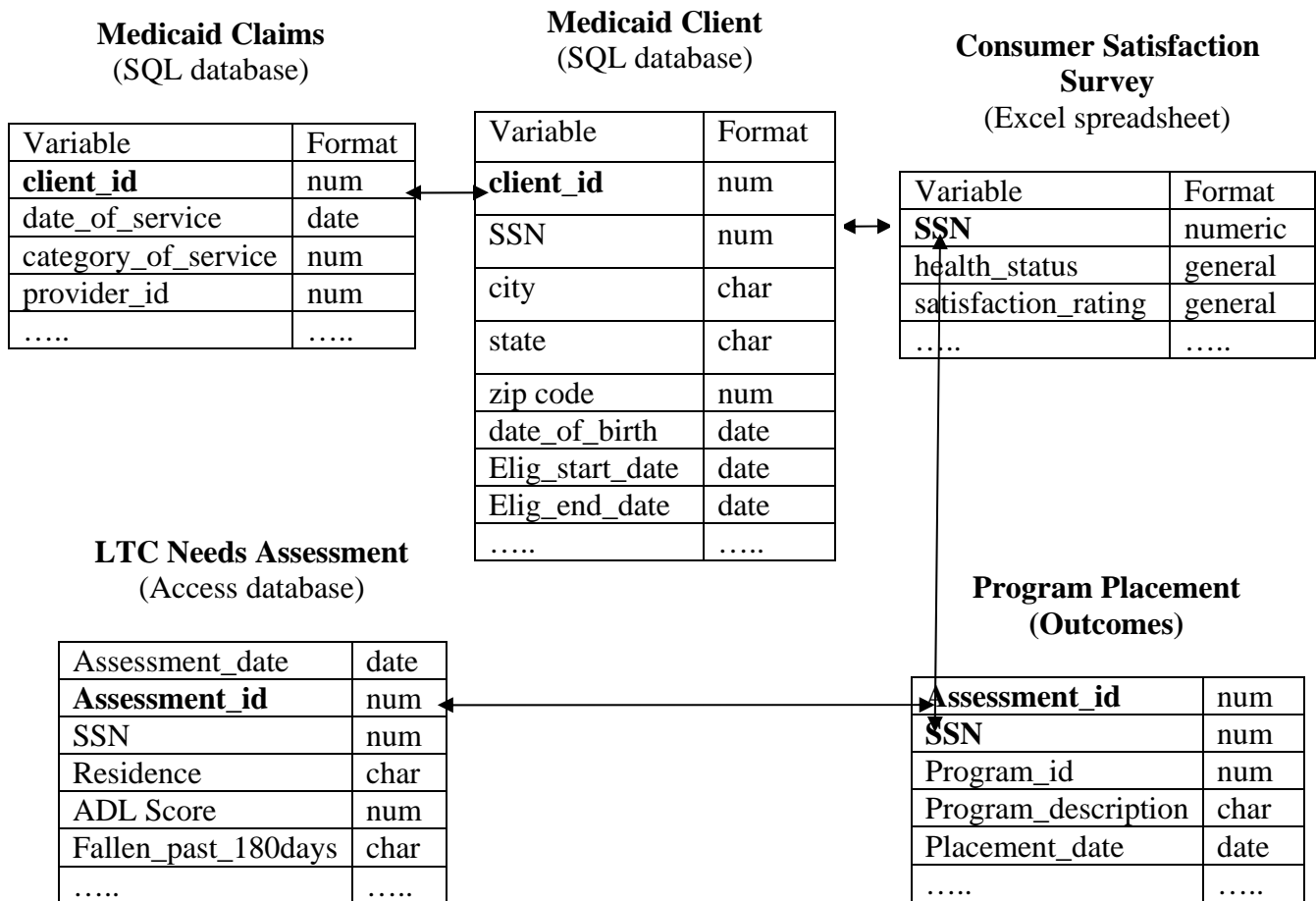
Sometimes you can use one data source as the link between two sources that do not have a common identifier. Figure 1 shows an example in which a Social Security number and client identification number may be on one but not all data sources. A third data source that has both these variables can be used as a bridge to make a link. For example, linking Medicaid claims with a consumer satisfaction survey may be possible by using the Medicaid client database as a link between these two sources. The Medicaid client database has both Social Security number and client ID, whereas Medicaid Claims and the Survey have only one of these elements.

Deterministic linking uses several variables to determine if the files will link. For each match with a particular variable, a certain number of points are given. Only after a threshold number of points are achieved is the link considered to be accurate. For example, agreement of a Social Security number might be given 20 points and the same last name may give you 15 points. If gender agrees, you might get 5 points. This method gives more opportunities for finding links, but can become difficult when assigning or determining which point values to give to different criteria.

Probabilistic linking. This method is similar to deterministic linking in that it uses several variables to find a link. The difference, however, comes in the points, or weights, that are given for each criterion. With deterministic linking, points are assigned before the data are analyzed. The points assigned are arbitrary. Probabilistic linking determines the weights (points) based on looking at the data first and then determining what the values should be in relation to each other. For example, a match on the last name “Smith” may be assigned fewer points for agreement than if there is a match on the last name “Wagner”, since “Smith” is a more common last name. In addition, if there is not a link with particular variable, points are taken away from the total amount. This method goes beyond the arbitrary point assignment and provides a more dynamic method for linking. This method, however, is complex in setting up a system of weights, thresholds and then linking the data.

In anticipation of linking the data elements or files, you may choose to leave the data sets stored on separate software programs (Oracle, SAS, Microsoft SQL server, Access, Excel, or other program) or you may choose a single relational database management system (DBMS) and export your data into a newly created single system (such as Oracle or Microsoft SQL server). Information Technology experts can help you make these decisions.

Figure 1. Linking Long Term Care Data



Storing Linked Data

There are two options for storing the data files that you wish to link:

- Leave data with the original source; access and link as needed
- Copy data into a central repository

Original Source of Data

Data can be stored in different databases and formats, such as a SQL database or spreadsheets, such as Microsoft Excel. You can link data from these different sources if you have a common identifier or other method for linking.

Different software packages can link data in the methods described above. Some packages are stand-alone, meaning they do not rely on other software packages to run. The cost of these packages can start at a few hundred dollars. Other software packages do not cost anything, but require you have software such as a SAS license.⁵ Your current system and needs will help determine which package or system development is needed.

⁵ See Appendix C.

Data Warehousing and Data Marts

A data warehouse is a repository where data from multiple databases is brought together for more complex analysis. Key components of a data warehouse include:

- Information from the different sources are copied and sent to the warehouse. This is often done on a routine schedule, such as once a month in a batch.
- Sometimes this warehouse can be located on a network server or web server.
- The previous historical information in the warehouse is preserved to allow longitudinal analyses.
- A quality assurance mechanism is developed to assure information copied and transferred from its source complies with standards before it enters the warehouse. This can include protocols for resolving conflicting information from two or more data sources.
- Once the information is stored in the warehouse, approved users have direct access to authorized data.

Appendix B includes a graphic illustration of a data warehouse under consideration in Ohio.

Another term that is often used with data warehousing is *data mart*. A data mart can either be dependent or independent. A *dependent* data mart is a subset of a data warehouse reserved for one particular users or task. For example, the Department of Human Services and the Department of Mental Retardation and Developmental Disabilities may all contribute to the same data warehouse. A “quality” data mart may be established with data elements relevant to quality assurance and quality improvement activities that would be accessed only by the QI team members from each department.

Data marts may also be *independent*. An independent data mart is similar to a data warehouse except that it houses information from only one department, agency, or section. For example, a quality unit could establish a data mart for its own use. This is an attractive alternative for organizing multiple data sets within the control of a single unit. Depending on the level of expertise available, internal staff may be able to develop this approach, or a consultant may be needed for initial set up and design.

Assessing Readiness to Integrate Data

This section reviews practical advice for program managers when making preliminary judgments about whether data linking or integration is feasible. Table 3 includes a check list of issues to consider when assessing data and system readiness for integration.

Table 3: Issues to Consider when Assessing Readiness for Data Integration

Data Readiness		
✓	Common Identifier	The presence of one or more common identifiers will dictate what can and cannot be linked.
✓	Paper/Electronic Format	Electronic format is needed for integration. Paper format can be converted to electronic, but the method for doing this must be determined.
✓	Common Time Frame	Data that are to be linked should represent the same, or similar, time frame of data.
✓	Primary/Secondary Data	Identifying what data you have control over and how to share data between departments or organizations.
✓	Level of Aggregation	It is ideal for individual level data to be linked, but aggregate data can potentially be integrated.
System Readiness		
✓	Funding	The current system level of integration desired can impact the level of funding need.
✓	Political Will and Agreement	A supportive leader who can bring parties together and work through data sharing arrangements and system requirements.
✓	Security and Confidentiality	Security is needed to ensure that information is protected. Technical and program staff need to understand required protections for accessing and using protected information.
✓	Staff Resources	Program managers will need to identify staff or consultants needed to develop a system for integration, linking and analyzing the data.

Common Identifier: A common identifier is needed to link two or more data sets. Examples can include Social Security number, ID number, or a combination of matches using different methods, such as using combined matches of last name, date of birth, Social Security number, etc. Where a common identifier does not exist, a third data set may be used to “bridge” the information (see Figure 1).

Paper/Electronic Format: Linking data requires the information to be in an electronic format. Prior to converting paper data into electronic format, the usefulness and accuracy of the data should be assessed. In addition to conversion of historical data, electronic methods for ongoing collection of the data should be established.

Common time frame for data reporting: It is ideal to link data that represent the same time frame. For example, assessment data for the first quarter should be linked with Medicaid claims data for the same time period. There will be times when exact time matches is not possible, or the data may not be complete for a particular time period (i.e. claims data may have a 3-6 month lag before considered complete). Significant differences in time periods among data sets will affect the value and usefulness of integration.

Primary/Secondary Data: It is easiest to link primary data over which you have direct control. A Memorandum of Understanding (MOU) will be necessary to govern the confidentiality and security of secondary data that you wish to link. Working out the details of such an agreement can be a time consuming, although an enlightening process.

Level of aggregation: Data can be at an individual or summary (aggregate) level. Linking data at the individual level provides the greatest option for grouping data into many different clusters, depending on the nature of your analysis. Linking aggregate data can be challenging. For example, a provider report in aggregate format may be linked with assessment data maintained on an individual level. To do this, however, requires that assessment data be manipulated and grouped so as to include only those people who are served by the provider agency during the time period represented in the provider report.

Funding: The sophistication of your current system will in part determine how costly it will be to integrate data, either on an ad hoc basis or through an established system such as a data warehouse. Data integration may be as simple as using existing applications and developing a process to share the information. It can also be far more extensive, involving multiple departments and data systems. The issue of whether to integrate is based in part on the cost benefit. Will integrating data be more cost effective than current manual or ad hoc practices? How long will it take for the agency/department to realize these savings?

Political will and agreement: Major changes in information systems require the engagement of a leader who supports the action. Data integration requires agreement by all parties who control the applicable data sources. The format, process, and outcomes of this integration are also important elements that must be collaboratively defined and agreed upon.

Confidentiality and Security: Integrating data invariably brings up issues of confidentiality, especially when more than one unit, agency, or department is involved. There may be internal restrictions on who has access to the data or protections placed on data by federal or state regulations. Ensuring security of information should be built into integration practices. A first step is to work with the affected entities in developing a Memorandum of Understanding (MOU) that specifies how, by whom and under what conditions data will be accessed.

Staff resources: Whether you are linking data or conducting preliminary work to prepare for data integration in the future, staff to support these efforts is instrumental. Linking data on an ad hoc basis requires staff trained in computer programming and data analysis. The development of data warehouses or data marts requires input from staff who to enter, analyze and maintain data on an ongoing basis. For in-depth discussion on the different roles to be filled when developing a data warehouse, see Alan

Simon's book, *Data Warehouses for Dummies*⁶. Decisions must be made on whether there is internal capacity to conduct integration activities or if an external consultant is needed.

The above issues will vary depending on the proposed scope of the integration effort. For example, issues of political will and confidentiality are less likely to be important if your proposal is to build a mini data warehouse for your own program data that you collect and control. In contrast, these issues may be very important if you are proposing to integrate your program data with the Medicaid eligibility file.

Summary

This paper was intended as an educational effort and “think piece” for HCBS program managers. Data linking and integration are not pre-requisites to good quality management but these tools can enhance your understanding of the data and where to target areas for improvement. As was mentioned, data integration can be done on an as needed basis, or on an ongoing basis. Developing a system for routine integration does not have to be large in scale. A data mart can be developed, such as one that looks at quality within the HCBS system (primary data). This may require a consultant, but it may provide an important tool in the long run for monitoring the quality of your HCBS program.

⁶ Simon, A.R. (1997). *Data Warehousing for Dummies*. Wiley Publishing, Inc. Hoboken, NJ.

Appendices

Appendix A. Data Source Assessment Worksheet

A worksheet to assess the different characteristics of various data sources.

Appendix B. State Examples

Examples of state's data integration systems are described or visually shown. For more information on a state, contact information is provided.

Appendix C. Resources for More Information

A resource list is provided to give additional information on the material presented in this paper. Resources for more technical guidance are also provided.

Appendix D. Glossary of Terms

A non-exhaustive list of terms and basic definitions that is often associated with data integration.

Appendix A. Sample Data Source Assessment WorkSheet

	Data Sources								
	Claims	Licensing	Incident Reports	Care Plan	Eligibility Files	Provider Files	Consumer Survey	Level of Care Determination	Assessment
Primary/ Secondary Data Source	2 nd	1 st	1 st	1 st	2 nd	2 nd	1 st	1 st	1 st
Paper Based			X			X	X		
Electronic	X	X		X	X			X	X
Time Period: When is it reviewed / conducted	Ongoing	Yearly	Ongoing	Yearly/ As Needed	Yearly/ As Needed	Yearly	Every 3 Years	Yearly	Every 6 mo.
Aggregate/ Individual Data	Individual	Aggregate		Individual	Individual				
Common Identifier	Client ID#			Client ID#	SS # DOB	Provider ID	SS # Client ID #	Client ID#	Client ID #

Appendix B. State Examples

Ohio

Contact Information:

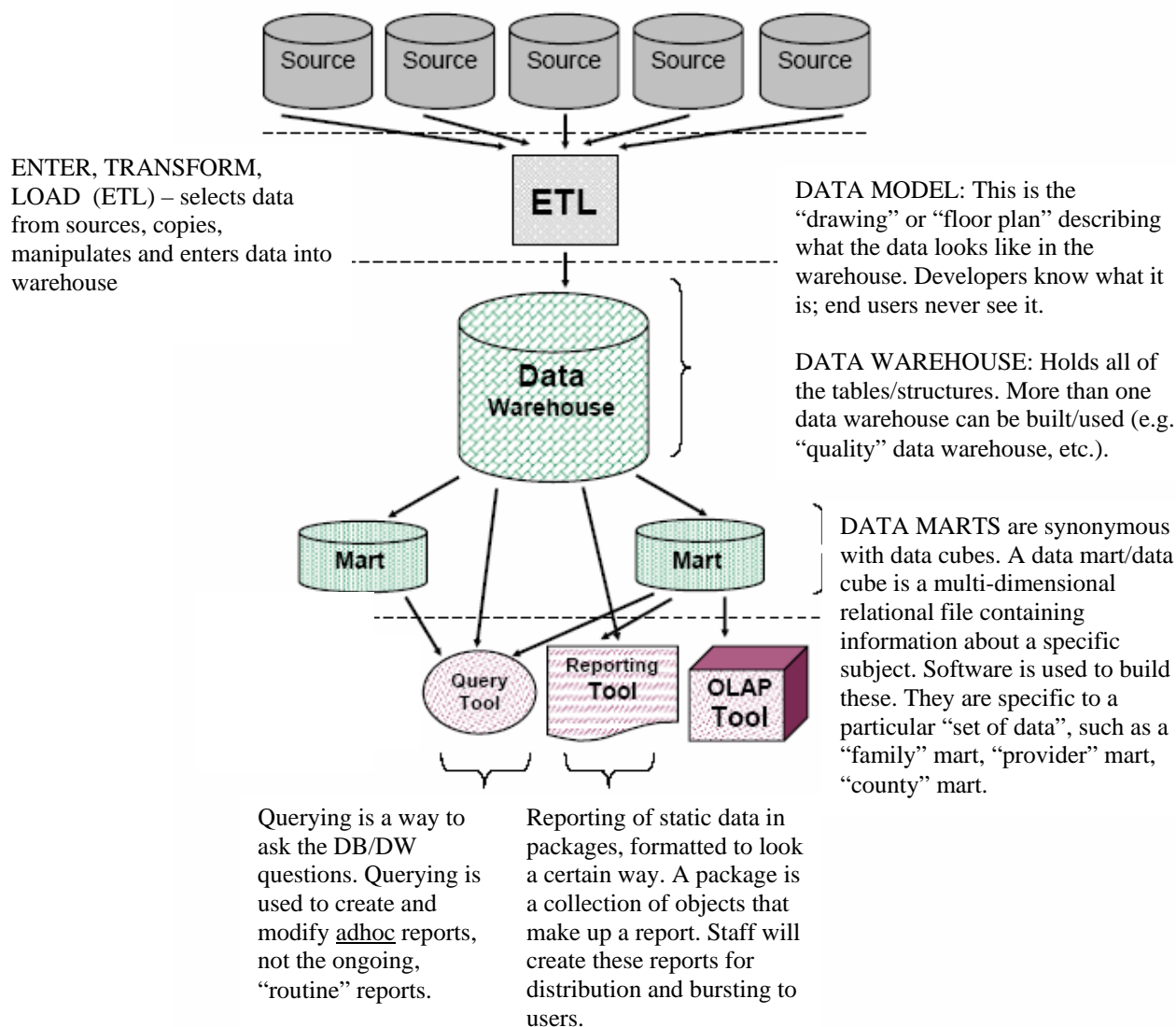
Suzzanne Freeze
Project Manager, ODMR/DD
suzzanne.freeze@dmr.state.oh.us
614.728.2518

Ohio's goal has been to develop an information management system, that is accessible to stakeholders from all facets of the service system, and can be used for decision-making on a daily basis. To do this, ODMR/DD has been developing a data warehouse to serve as the foundation for the information management system. The image on the next page represents the layout for the flow of data from individual data sources to the data warehouse to use in a variety of report formats.

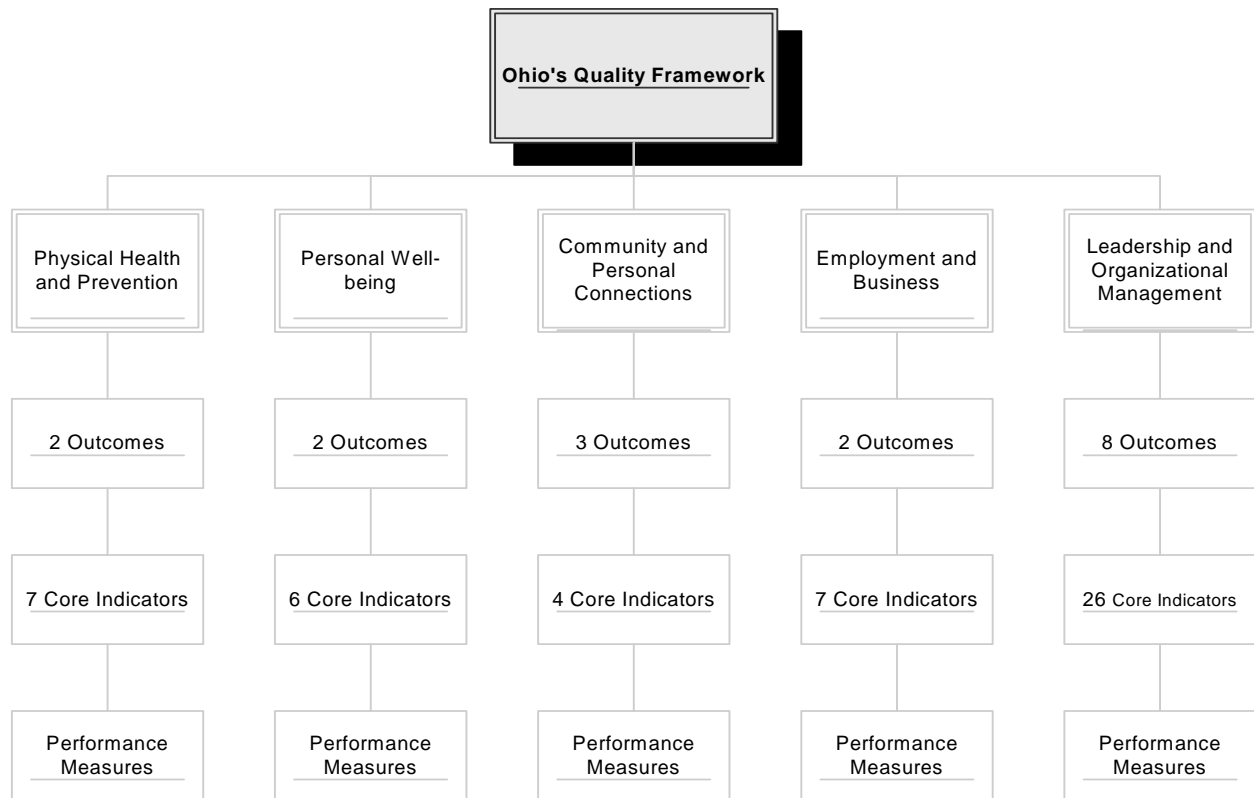
Ohio Dept MR/DD QA/QI Grant

The “Data Diagram” below is a visual used during presentations to evoke the sense of taking many disparate sources of data in to a warehouse and then purposefully separating the sources in to ‘marts’ for a variety of uses, such as generating reports specific to elements of the Quality Framework, or for answering queries related to CMS requirements. ODMR/DD has elected to use Cognos software for such querying and reporting.

Diagram –OHIO Data Diagram with Comments



Ohio's Quality Framework is the guide for questions of what data to collect. This diagram reflects the Framework's organization:



Minnesota

Contact Information:

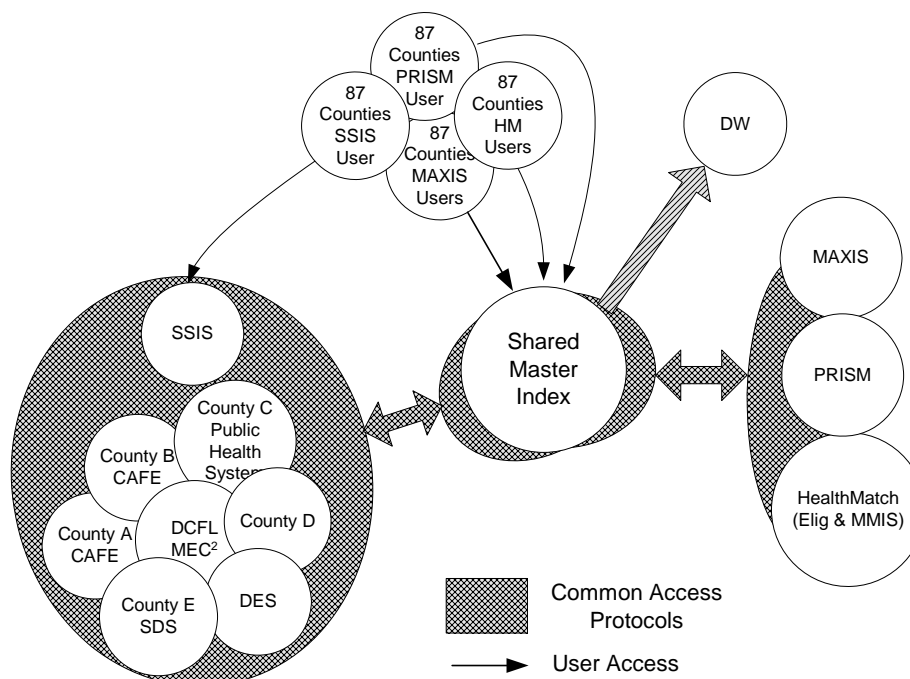
Shirley Manchester

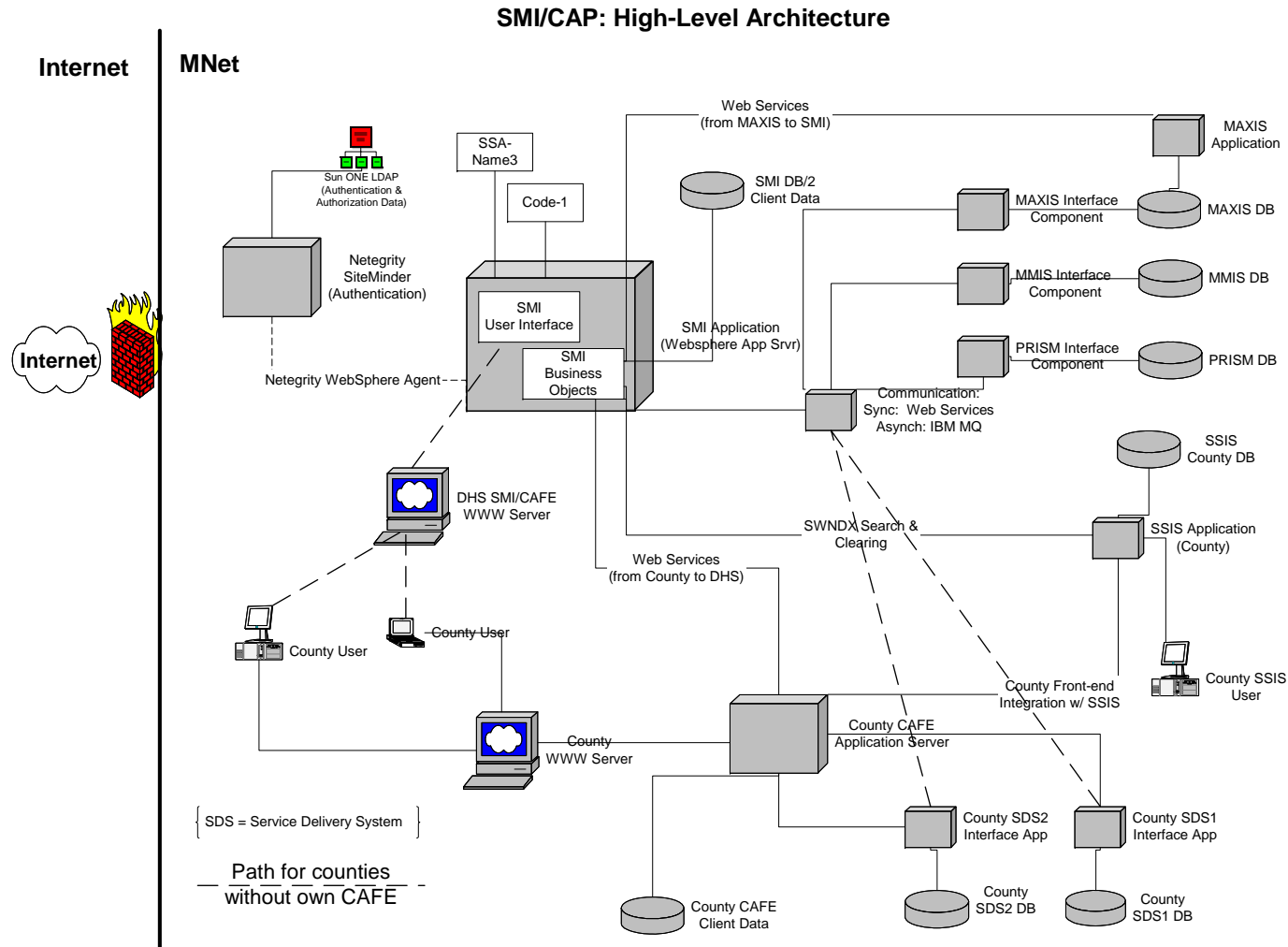
Shirley.Manchester@state.mn.us

Minnesota has developed a Shared Master Index / Common Access Protocol (SMI/CAP). By assigning each human services client an individual identification number that is used across multiple service delivery systems, the Shared Master Index (SMI) allows caseworkers to view and assess the range of services that individual clients and families are receiving and allows program managers and policy-makers to improve coordination of services and client outcomes. To learn more about this system, go to http://www.dhs.state.mn.us/main/groups/publications/documents/pub/dhs_id_050119.hcsp.

Identity Management. A common identifier is assigned so it becomes possible to track a person across multiple systems/programs.

Data Sharing and Synchronization between systems is performed following the protocol standards put in place by the project. Within the privacy and security requirements of MN Data Practices and HIPAA, any application that builds an interface using the CAP standard can access the features of the SMI and other systems that use that standard. Many county users will have direct access to the SMI while others will access it through their county-based systems. The common identifier created by the SMI makes it possible to automatically keep demographic and address data synchronized between the systems.





Florida

Contact Information:

Mindy Sollisch
Statewide Community-Based Services
Department of Elder Affairs
Sollischm@elderaffairs.org

Florida's Department of Elder Affairs has a data warehouse that links data from the Agency for Health Care Administration and the Department of Elder Affairs. The data warehouse holds information such as Medicaid Waiver claims, assessment of functional status, mental impairment, chronic health conditions, nutrition, and social support for elders who are candidates for nursing facility long term care. From this warehouse, information can be provided to do quality assurance through monitoring.

Florida's Department of Elder Affairs developed a Holistic Monitoring Tool to be used by Area Agencies on Aging for monitoring Aged/Disabled Adult Services Medicaid Waiver clients, claims and providers. The tool uses information from the data warehouse to provide information to support the review. By selecting the case management agency to be reviewed, a random sample of clients is generated. The screen will then show the names of the randomly sampled clients along with a checklist the reviewer must fill out (see below). For each client selected the tool displays claims information for the past six months along with a list of the programs that clients are/were enrolled in.

The tool also generates a client report for each participant in the sample (see below). This report includes the services received by the client and the average care plan cost. The procedure code, procedure name, number of units received, total amount paid and unit cost are given for each service that is listed. The names of the providers who serviced the sampled clients are also provided. The report also includes a list of all programs the clients are/were enrolled in and their enrollment status and start and end dates.

The tool also requires the monitoring of at least five claims for each sampled client. A screen is provided that contains claim-level questions. The tool also requires the monitoring of all providers that serviced the sampled clients. A screen listing the providers is included in the tool as well.

The Holistic Monitoring Tool also generates provider letters the Area Agencies on Aging use to request claim-level documentation for the review (including service authorization). Information from this review process is aggregated into Provider Report Cards and then analyzed. For more information, go to <http://www.nasua.org/waiverconference/hcbs2005/47%20Mindy%20Sollisch.ppt>.

ADA Waiver Holistic Monitoring Tool - Case Management

Action Edit Query Block Record Field Help

Clients

Last Name	First Name	SSN	Med ID	DOB	Prog	CM										CP				Care	Fair	HA	LO
						App AAA	Meet Need	Choice Prov	CIRTS Acc	CP Assess	CP Svcs	Other Fund	Notes CP	Home Visit	Assess Doc	Cost Doc	Plan Doc	Hrg Doc					
BEGIN	RON	999999999	999999999	01/04/1909	ADA																		
DOE	JANE	111111111	111111111	01/05/1942	ADA																		
JAMES	FRED	666666666	666666666	04/15/1926	ADA																		
JONE	MARY	333333333	333333333	12/06/1908	ADA																		
LAME	FLINT	888888888	888888888	01/08/1940	ADA																		
MATT	SALLY	111111112	111111112	03/25/1929	ADA																		
SMITH	GREGG	222222222	222222222	04/02/1910	ADA																		
STONE	JILL	555555555	555555555	03/19/1919	ADA																		
TAYLOR	ELIZABETH	777777777	777777777	10/12/1920	ADA																		
WAGNER	JANE	444444444	444444444	01/02/1929	ADA																		



PROVIDER REPORT - 10 CLIENT SAMPLE

Provider Name: ADT SECURITY SERVICES

Provider ID: 683529500

Client Name - SSN - Medicaid ID SMITH, MIKE - 222222222 - 222222222

Service Date	Procedure Code	Description	Service Units	Amount Paid
06/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	30	\$21.00
07/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
08/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
09/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	30	\$21.00
10/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
11/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	30	\$21.00

Client Name - SSN - Medicaid ID WEST, JANE - 333333333 - 333333333

Service Date	Procedure Code	Description	Service Units	Amount Paid
06/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	30	\$21.00
07/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
08/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
09/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	17	\$11.90
10/31/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	31	\$21.70
11/30/2004	S5161-U2	EMERGENCY ALERT RESP. - MAINT.	30	\$21.00

Appendix C. Resources for More Information

Data Management

Prepared by T. Gray, T. Hubley, R. Keith, M. Lahti, S. Loux, S. Mahimkar, G. Shaler. *A Guide to Effective Data Management in Applied Research Projects* (December 2005). USM: Muskie School of Public Service.

This report provides information on data management elements such as standards for management, data collection, validating, storage and analyzing data.

Data Integration Resources and Examples

Data Warehousing Website (<http://www.datawarehousing.com>)

This website has information on data warehousing, terminology and links to other resources.

CMS MITA (<http://www.cms.hhs.gov/medicaid/mmis/mita.asp>)

The Medicaid Information Technology Architecture (MITA) is an initiative of the Center for Medicaid & State Operations (CMSO) and is aligned with the National Health Infrastructure Initiative (NHII). MITA is intended to foster integrated business and IT transformation across the Medicaid enterprise to improve the administration of the Medicaid program. Its common business and technology vision for state Medicaid organizations will emphasize:

- A patient-centric view not constrained by organizational barriers
- Common standards with, but not limited to, Medicare
- Interoperability between state Medicaid organizations within and across states, as well as with other agencies involved in healthcare
- Web-based access and integration
- Software reusability
- Use of commercial off the shelf (COTS) software
- Integration of public health data

Simon, A.R. (1997). *Data Warehousing for Dummies*. Wiley Publishing, Inc. Hoboken, NJ.

This book provides an overview to data warehousing and the various elements that need to be considered before, during, and after a warehouse is created.

Research and Practical Experiences in the Use of Multiple Data Sources for Enterprise-Level Planning and Decision Making: A Literature Review – Using Information in Government Program (1999). Center for Technology in Government, University at Albany/SUNY.

http://www.ctg.albany.edu/publications/reports/multiple_data_sources/multiple_data_sources.pdf

This report provides the benefits, issues, methods and results of different organizations and states that integrated data either within their organization or with multiple sources. Case studies are presented for examples.

Linking Data

Walen, D., Pepitone, A., Graver, L., Busch, J.D. (2001). *Linking Client Records from Substance Abuse, Mental Health and Medicaid State Agencies*. US Department of Health and Human Services: Substance Abuse and Mental Health Services Administration, Rockville Maryland.

<http://csat.samhsa.gov/adbse/linking.pdf>

This report gives some examples of different methods for linking records from substance abuse, mental health and Medicaid state agencies.

Linking and Integration Software Packages

There are a number of different software packages that link data. Below is a non-exhaustive list of some packages. This paper does not endorse any of the following packages.

Cognos	http://www.cognos.com/
Link King (free)	http://the-link-king.com/
Microsoft SQL Server	http://www.microsoft.com/sql/default.msp
Oracle	http://www.oracle.com/
SAS	http://www.sas.com/

Sujansky and Associates (August 2004) *Patient Data-Matching Software: A Buyer's Guide for the Budget Conscious* <http://www.chcf.org/topics/view.cfm?itemid=104595>

This report gives a comparison of four commercially available data-matching software packages.

Appendix D. Glossary of Terms

Note: This is not intended to be an exhaustive list of definitions. Terms used here are often used in association with data integration and storage issues. These terms are technical in nature and are intended to provide basic information about the nature of its meaning.

Ad Hoc Report

A report generated on a one-time basis to answer a specific question or questions.

Data Integration

The process of combining data from different operational sources or data systems to facilitate access by users.

Data Mart

A subset of tables within a data warehouse containing data that is used by specific individuals or a specific department or office. Can also refer to a small independent data warehouse.

Data Set

A collection of data stored in an electronic file or files, usually as part of a database.

Data Sharing

Methods for allowing multiple persons to access and edit data safely.

Data Source

The origin of where data is collected, such as claims database or survey responses.

Data Warehouse

A large collection of data from many sources that is updated regularly and accessed by many people within an enterprise or agency on a read-only basis and used for decision support.

Database

An organized collection of data stored electronically, usually in multiple tables or files, which can be accessed using database management software for updating purposes or to answer questions and generate reports.

Database Management System (DBMS)

Computer software used for putting data into databases, storing it electronically for computer access, and extracting and processing it for authorized users.

Enterprise Information System (EIS)

EIS uses integration technology to combine information from a number of different systems in “real time”. There is no storing of information on a disk or in a physical data warehouse. EIS is a form of a “virtual” data warehouse.

Extract, Transform, Load (ETL)

ETL is a process used in data warehousing that takes information from different sources (extract), cleans and manipulates it to fit the requirements of the receiving databases (transform), and then stores this information in a data warehouse (load).”

Field

An element of a database record in which one piece of information is stored. In a data set, it is a single column. Fields are also called variables because they can “vary” for each observation in the data set. In a consumer survey example, each question on the survey is a field or variable.

File

A program, document, or data set physically stored on a network or local drive.

Operational Data Store (ODS)

A database containing limited amounts of current data which is continually updated in the course of daily operations, can be accessed to extract information, and may be used as an interim storage area for data that will eventually be stored in a data warehouse.

Query

Computer software code, most commonly written using Structured Query Language (SQL), that is used to extract and process data from a relational database. Queries are frequently used to generate data for reports and in some applications may be created using software query tools to generate the code.

Record

A single observation/row in a data set. It is often an individual person but depending on the data set, it could also be a medical claim, a geographic unit, an organization or other unit of analysis. For example, in a data set of consumer survey results, each consumer survey is a record. A record represents an entity with certain field values.

Relational Database

A collection of data stored in two-dimensional tables with rows (records) and columns (fields) in which key fields are used to relate or link the data stored in separate tables so that it can be retrieved and processed using queries.

Star Schema

A star schema is one way organizing data warehouse information. Like a star, the center has the basic factual information. The points of the star are different ways you can look at the data.

Value

The numeric or categorical contents of a single cell in a data set. For example, “12” is the value of an individual person’s highest year of school if they have graduated from high school. In a consumer survey, the answer to a question on the survey (e.g., yes, no, unsure) is the value in the field.