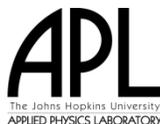


Prepared for the
Administration for Children and Families (ACF)

National Human Services Interoperability Architecture
Performance Information Repositories (PIRs)
DRAFT Version D0.3
September 2012

Prepared by:
The Johns Hopkins University
Applied Physics Laboratory (JHU/APL)



Draft Issue

It is important to note that this is a draft document. The document is incomplete and may contain sections that have not been completely reviewed internally. The material presented herein will undergo several iterations of review and comment before a baseline version is published.

This document is disseminated in the interest of information exchange. The Johns Hopkins University Applied Physics Laboratory (JHU/APL) assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation. JHU/APL does not endorse products or manufacturers. Trade and manufacturer's names appear in this report only because they are considered essential to the object of this document.

Note: This document and other NHSIA-related documentation are available for review from the Administration for Children and Families (ACF) Interoperability Initiative website. The URL for the site is currently: <http://transition.acf.hhs.gov/initiatives-priorities/interoperability>. When ACF completes the migration to their new website the URL is expected to be <http://acf.hhs.gov/initiatives-priorities/interoperability>.

Review and comments to this document are welcome. To comment, please contact Joseph Bodmer at joseph.bodmer@acf.hhs.gov or 202-690-1234.

Christine Salamacha
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723
Phone: 240-228-4976
E-Mail: christine.salamacha@jhuapl.edu <mailto:Christine.S>

Revision History

Version	Date	Description of Change	Reference	Edited Location	Executor
D0.3	2012-09	Original version published.			COS

This page intentionally blank

Table of Contents

List of Figures.....	ii
1 Introduction	1
2 NHSIA Performance Information Repository Concept Overview.....	1
3 High Level Requirements for Integrated PIRs	4
4 Integrated PIR Considerations.....	5
4.1 Common Data Elements Stored in Integrated PIRs.....	5
4.2 Client-Specific Information.....	6
4.3 Operational Data Bases and Integrated PIRs.....	6
4.4 Data Ownership and Managing Access to Data.....	8
4.5 Establishment and Management of PIR	9
4.6 PIR Interfaces and NIEM-based Standards.....	9
4.7 Users of Integrated PIRs	9
5 Analytics	9
6 Integrated PIR Vignettes.....	10

List of Figures

Figure 2-1. Reference Model: Performance Information Repositories 2

Figure 2-2 Basic Steps in Data Analysis (source: Leveraging Data in Federal Organizations)
..... 3

1 Introduction

The National Human Services Interoperability Architecture (NHSIA, pronounced niss'-e-a) is being developed by the Administration for Children and Families (ACF) as a framework to support comprehensive eligibility determination and information sharing across programs and agencies, improved delivery of services, prevention of fraud, and better outcomes for children and families.

The focus of this paper is Performance Information Repositories (PIRs). The PIR is a NHSIA core concept and a key component of the NHSIA Performance Management strategy. Implementation of PIRs involves concepts outlined in the NHSIA Systems and Infrastructure Viewpoints. The overarching purpose for establishing PIRs is to improve data analytics and thereby support the President's challenge to make evidence-based policy decisions.

Since taking office, the President has emphasized the need to use evidence and rigorous evaluation in budget, management, and policy decisions to make government work effectively. This need has only grown in the current fiscal environment. Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future. ¹

The NHSIA architecture supports the collection and management of data required to support several types of analyses, both descriptive and prescriptive. This paper describes the PIR concepts in more detail so that those considering NHSIA can better understand the value of implementing PIRs. This paper describes how PIRs might be implemented and used.

2 NHSIA Performance Information Repository Concept Overview

Figure 2-1 depicts the NHSIA Performance Information Repository (PIR) concept. The figure shows a repository that supports performance management requirements at each level of government.

¹ M-12-14 "Use of Evidence and Evaluation in 2014 Budget", Executive Office of the President Office of Management and Budget, Memorandum to the Heads of Executive Departments and Agencies, May 18, 2012.

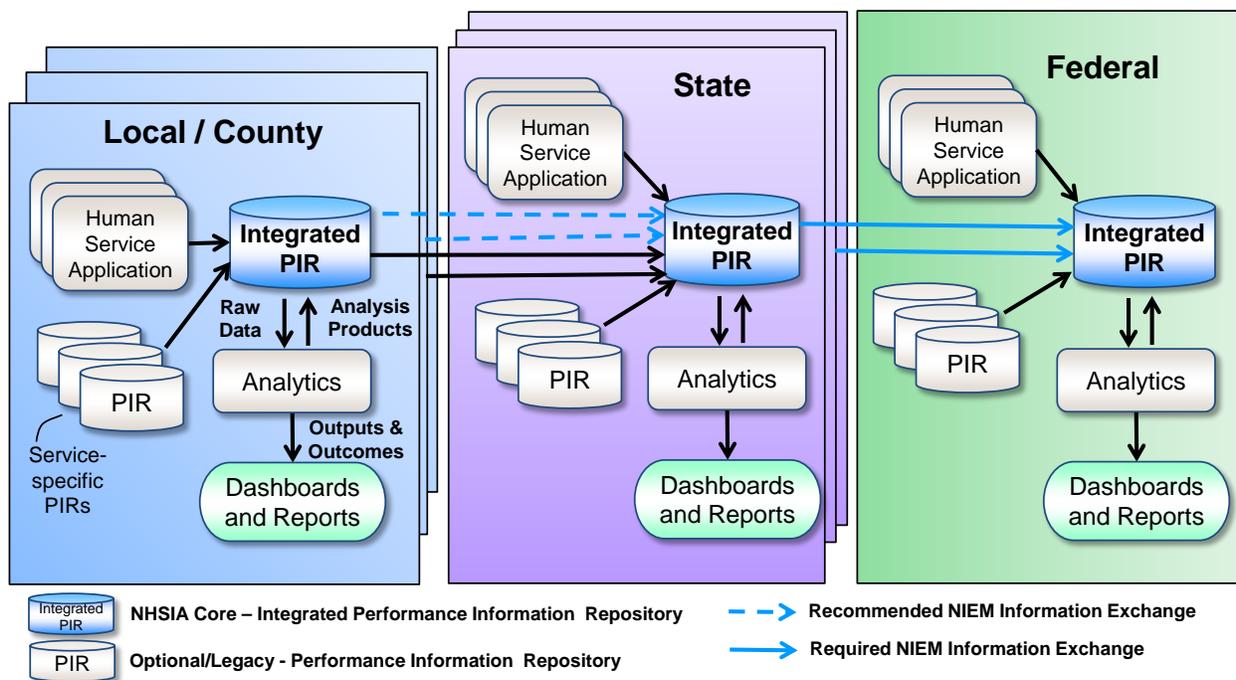


Figure 2-1. Reference Model: Performance Information Repositories

Human Service Applications: are applications used in human services operations. An example of a human services application is a case management tool. These applications typically interface with operational data bases that store transaction information (e.g., a data base for case records). Performance related information can be collected from these applications and their associated data bases.

Service-specific PIRs: are local repositories that store performance information related to delivery of services. They may be maintained by the local jurisdiction or perhaps by service providers. These may or may not exist in a given jurisdiction or for a given program.

Integrated PIRs (local, state, federal level): are essentially data warehouses. At each specified level, information is collected across programs into a common environment. IT services replicate performance information stored in operational data bases and service-specific PIRs to the integrated PIR at the same level. IT services are also employed to copy performance information to higher-level integrated PIRs. Standard interfaces (primarily NIEM-based) are used to exchange information with integrated PIRs.

Local jurisdictions (county, city, and tribe) may elect to consolidate multiple jurisdictions' data into a single repository. Local jurisdictions may also share a repository with the state. Likewise, states in a region may share an integrated PIR. Integrated PIRs are stored at hubs to enable ready access to the information. Integrating the PIRs at the state and federal levels should reduce duplication of effort and data reporting, and also enable more meaningful analysis. A PIR may be implemented as a "virtual" structure that is physically several different databases.

Analytics: processes, techniques and tools that collect, clean up and analyze raw performance-related data as well as products from analytics to generate outputs and outcomes including statistical comparison over time. Figure 2-2, an excerpt from a Corporate Partner Advisory Group report (sponsored by Advancing Government Accountability (AGA)), outlines the basic steps of data analytics. Analytics can be enhanced by the integration of performance information across human services programs and agencies/organizations. Analysis products are stored in integrated PIRs and are also provided to higher-level integrated PIRs. Figure 2-1 depicts output from data analysis being available through dashboards and reports.

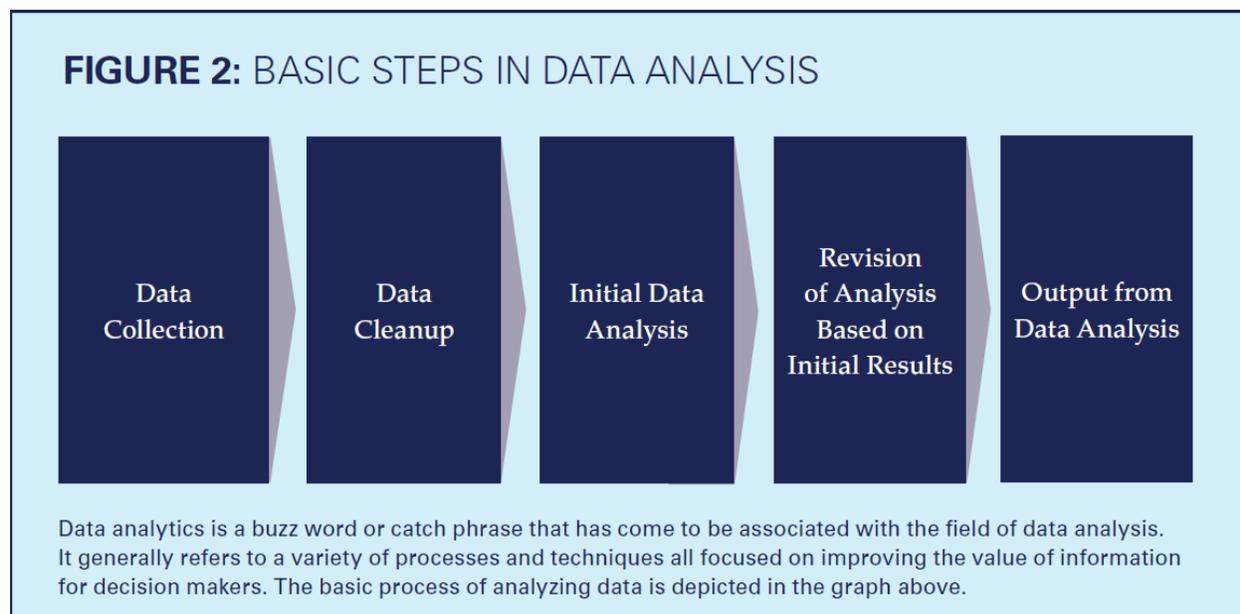


Figure 2-2 Basic Steps in Data Analysis (source: Leveraging Data in Federal Organizations²)

The connections in Figure 2-1 between the local/county-level integrated PIRs and the state-level integrated PIR represent sharing whatever information the state needs to perform its analysis. Similarly, the connections between the state-level integrated PIRs and a federal-level integrated PIR represent sharing whatever information the Federal Government needs to perform its analysis.

Data analytics can be done in real-time, producing reports without storing results in the PIR. Other analytics may generate summary results required for reports; these data products would be stored in the PIR if the PIR is the source for generating reports. Aggregated and disaggregated data reporting requirements are addressed in section 5.

Dashboards and reports: are mechanisms used to present analysis results to stakeholders. Dashboards would typically be provided through an on-line application or web portal.

² Corporate Partner Advisory Group/ Advancing Government Accountability (AGA), “Leveraging Data Analytics in Federal Organizations” Report No. 30, May 2012

Interfaces: The arrows in Figure 2-1 represent interfaces between PIRs and other components. The blue arrows represent suggested NIEM-based exchanges. Though NIEM is being explored as the standard of choice, the important point is that interfaces are implemented using common standard(s) and standardized interfaces are shared and re-used.

3 High Level Requirements for Integrated PIRs

This paper focuses on integrated PIRs and the interfaces to these PIRs. The other components of the performance management strategy (human services applications, service-specific PIRs, analytics, and dashboards) are established at the local, state, or federal level as appropriate.

Integrated PIRs are established to enable the following:

- Aggregation of performance indicators by county, jurisdiction, state, and program.
- Aggregation of performance data over multiple years, to include lifetime for select segments of the population (e.g., individuals who were in foster care).
- Aggregation of any other performance-related data required to compute outputs.
- Aggregation of client service information associated with multiple programs, over years of receiving services. Aggregated client-related information is used to analyze change in a client's situation against program criteria and assess program success for that client.
- Generation of standard periodic performance reports in the format mandated by federal programs, to include performance indicator reports. PIRs support uniform and robust reporting capabilities across jurisdictions and programs.
- Sharing of performance metrics and performance data across human services programs.
- Use of analytics software to conduct descriptive and prescriptive analyses.
- Tracking program-related changes that impact performance such as regulation changes. At the simplest level, an integrated PIR stores performance data over time and is a source of performance data prior to the specified change and after the change.
- Creation of dynamic and interactive dashboards by state, county, zip code and other categorization parameters.
- Discovery of information collected by/for programs and the specification of new reports.
- Conduct of longitudinal studies (e.g., by academic and research institutions).
- Correlation of data and anomaly detection required for fraud detection.

PIRs should:

- Support program data retention requirements
- Support program cleansing/ redaction requirements
- Comply with program data sharing policies
- Comply with Personal Identifiable Information (PII) regulations
- Comply with client confidentiality and privacy authorizations

4 Integrated PIR Considerations

This section discusses several topics related to integrated PIRs:

- The collection and management of data elements common to different programs (agencies)
- The availability of client-specific information and/or aggregate (e.g., program, population) analysis products
- How operational data bases relate to an integrated PIR
- Data ownership and managing access to data
- Establishment and management of integrated PIRs
- PIR interfaces and NIEM-based standards
- Who are the users of integrated PIRs

4.1 Common Data Elements Stored in Integrated PIRs

Human services programs collect information pertaining to client status and conditions, agency operations, etc. The same information may be collected by different programs or, very likely, similar information (comparable but not exactly the same) information is collected. Integrated PIRs host (“integrate”) information for multiple health and human services programs. One objective in establishing integrated PIRs at the state and federal levels is to reduce duplication of effort and data reporting. To realize this, it is necessary to understand what each program collects as “performance” data and determine if there are common data elements.

In the case of common data elements, partners contributing to the PIR need to establish policies and practices that address whether there will be a single instance of these common data elements and if so, how data are managed and updated. Furthermore, interfaces established to transfer data from operational data bases and service-specific PIRs to integrated PIRs will need to mediate among data definitions used in these repositories and the definitions adopted for the integrated PIR.

The expectation is that all counties provide the same data elements to the state PIR and all states supply the same data elements to a federal PIR.

4.2 Client-Specific Information

If an integrated PIR stores information for individual clients (across programs), MPI services are needed to aggregate information per client. Common data elements essential to mapping service and performance data to a specific client are considered in Master Person Index Services.

A foundational question is at what level (local, state, federal) are client-specific data stored in the integrated PIR? For example, are client-specific data replicated from the local PIR to the state PIR and then to a federal PIR? Or is the practice to retain client data at point of origin (e.g., county) and allow for analytics at a higher level to access the lower level PIR? If data hosted at a lower-level PIR is accessed at a higher level, then the lower-level PIR must provide a common interface / data service to expose this data. Furthermore, this data service will have to be common across all PIRs hosting this type of data.

Currently, the yearly Online Performance Appendices generated by the Office of Planning, Research and Evaluation (OPRE), a unit within the Administration for Children and Families, report program-wide measures derived from the roll-up of state outcome and output data. Temporary Assistance to Need Families Program (TANF), however, requires both aggregated and disaggregated data related to disbursement of services. One example of the “disaggregated” data that must be reported is social security number. The following Federal Registry excerpt (February 5, 2008) provides insight into the regulation related to disaggregated data:

Collecting social security numbers is an existing requirement. We have been collecting the social security numbers for TANF family members since October 1999. This information is protected by the safeguards under the Privacy Act. The TANF recipient social security numbers are encrypted during data transmission, maintained in a secure location and use and access to them is limited. We have found them very useful in our research, especially as it relates to the use of the National Directory of New Hires database to assess the impact of welfare reform on TANF recipients using such measures as job entry, job retention, and earnings gain. We do not have statutory authority to collect additional data on the wellbeing of children; the statute limits the data that the Department can collect.³

4.3 Operational Data Bases and Integrated PIRs

Integrated PIRs are data warehouses. A data warehouse is a database of unique data structure that allows relatively quick and easy performance of complex queries over large amounts of data.

³ Federal Registry Part II Department for Health and Human Services, Administration for Children and Families, 45CFR Parts 261, 262, 263, and 265, Reauthorization of the Temporary Assistance for Needy Families (TANF) Program, Final Rules, February 5, 2008

Per the integrated PIR concept, data from numerous operational databases is stored in a single PIR. Data warehouses usually aggregate and denormalize operational data. Where an operational database would have one table with a person record in it plus one record per transaction, denormalized case data for a person would have person data replicated in each transaction record.

The following extract from the [NHSIA Infrastructure Viewpoint D0.1⁴](#) addresses business intelligence systems. For the purposes of this discussion, analytics systems are a type of business intelligence systems and an integrated PIR is a type of common data warehouse.

Business intelligence systems focus on improving the access and delivery of business information to both information providers and information consumers. This section addresses the architecture patterns relevant to implementing a business intelligence infrastructure to support reporting and analysis of information combined from multiple organizations as well as individual reporting requirements that are shared among organizations.

Business intelligence systems are typically designed to process and analyze large volumes of information using a variety of different tools. A business intelligence system must, therefore, provide scalability to support growing information volumes and be able to support and integrate products from multiple vendors.

A business intelligence system may also provide access to business information through the use of an information catalog that documents decision support objects that can be employed by information consumers to answer the main business questions that arise in everyday business operations. These objects may be combined and presented via a user interface called a dashboard. Some systems also provide information consumers with the ability to subscribe to the information they require, and the system delivers it to them at predefined intervals.

To support reporting of information combined from multiple organizations, the Business Intelligence and Analytics pattern presupposes the existence of a common data warehouse to store the required information. The information stored in a data warehouse will be sourced from operational databases most likely via the Population pattern and will be structured in such a way as to facilitate inquiry, reporting and analysis. This will thereby reduce the impact on an organization's operational transaction systems.

⁴ NHSIA Infrastructure Viewpoint, Draft Version D0.1, September 2011.

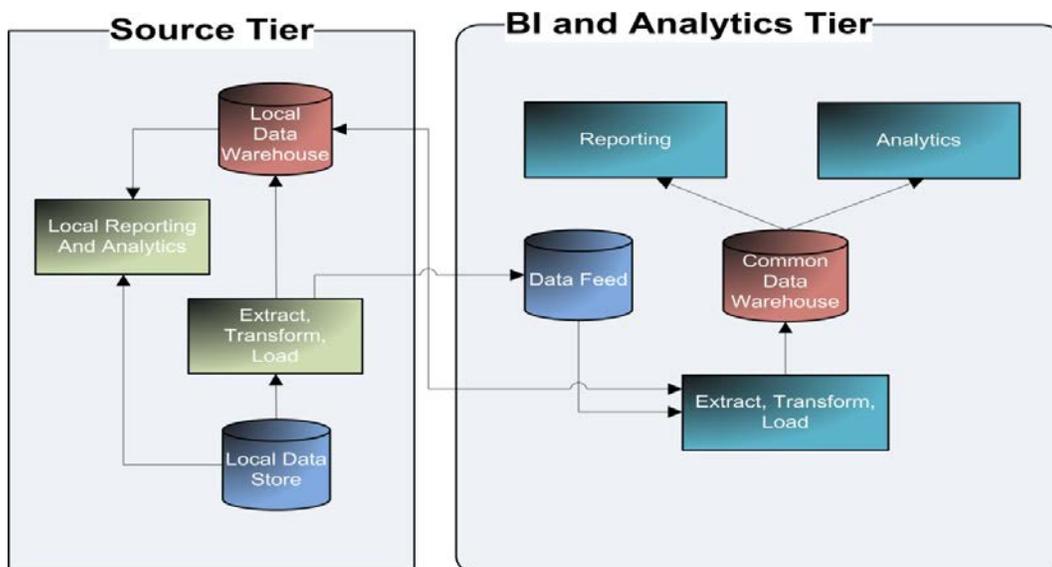


Figure 4-1 Business Analytics Patterns (from NHSIA Infrastructure Viewpoint)

The process of moving data from its location in an operational system involves transforming it into the required format for the target data warehouse, and loading it into the target data warehouse. Even though the three steps may be performed on different systems or in different locations, there are often referred to together as extract, transform and load, or ETL.

Individual organizations will most like assume responsibility for providing data feeds to the central data warehouse. These may be created as a part of the organization's own ETL process or by some other means that will depend highly on the application software and database technology of the organization. The format of the data and the communications means will again be up to the individual organization.

The transform process takes the incoming data feed and processes that data according to predetermined criteria. For example, name and address data may be normalized before being loaded. In addition, data may be aggregated or analyzed prior to loading. Finally, the load phase writes the data to the local data warehouse.

A single integrated PIR will host data from multiple sources – multiple programs and multiple jurisdictions. The “Transform” step discussed above transforms and matches or maps comparable data from different sources.

4.4 Data Ownership and Managing Access to Data

Rules and privileges established for operational databases and service-specific PIRs should be supported in the PIR environment. The PIR environment is simpler in that most users of the data require “read only” access. By default, guidelines established for sharing of data amongst programs, providers, etc. apply whether data are in an integrated PIR or not.

One strategy for dealing with personally identifiable information (PII) is to support anonymous “views” in the PIR environment, where the information that really identifies the individual client is redacted. Real client identity information in the PIR could be separated from the transaction-level data. Also, PII data could be obfuscated in the PIR database.

In terms of data ownership, the two key considerations are how PIR data can be used (addressed above) and who is responsible for ensuring data in the PIR are current. The entity that provides the data in the PIR is responsible for ensuring its accuracy and timeliness. Likewise, entities that invoke analytics to generate outputs and outcomes data stored in the PIR are responsible for managing those analytics data products.

4.5 Establishment and Management of PIR

At each level, a department or agency will be designated to establish and manage the PIR for the entire county (local PIR), the entire state (state PIR) or the federal programs (federal PIR).

4.6 PIR Interfaces and NIEM-based Standards

There is growing support for developing NIEM-based information exchanges. If well-defined legacy interfaces exist, there may be little value in expending the effort to develop a NIEM-based exchange. However, if there is a new requirement or a significant change to an existing interface, it is recommended that a NIEM-based information exchange be defined. There is value in mapping data elements used in legacy interfaces to NIEM in order to develop a commonly understood vocabulary.

If information provided by the county to the state is then replicated in the exchange between the state PIR and a federal PIR, a common (possibly NIEM-based) information exchange is recommended.

Finally, analytics provide the main interfaces to PIRs; these interfaces are used to both extract data and load results. A NIEM interface is not necessary for the analytics software to extract data from the PIR.

PIR interfaces need to manage access privileges per the rules defined for the information they contain.

4.7 Users of Integrated PIRs

Users of integrated PIRs include staff persons who want to check the status of their data in the higher level PIR (state, federal), social scientists doing longitudinal studies, state and federal program analysts, and the public (for aggregated information). As stated previously, integrated PIRs are designed to share performance data across programs and foster better-informed performance management.

5 Analytics

As shown in Figure 2-1, analytics are employed at each level to support the development of outcomes and outputs. Program outputs are program products that result from the internal

operations of the program, such as the delivery of planned services. Outcomes measure the effect on client state resulting from the receipt of services. Intermediate outputs are measured (for each client) at the end of program participation and are hypothesized to be requisite steps towards the desired longer-term outcomes. Strategic outcomes are the longer-term effects (benefits) of the services (programs) and address changes in client population state relative to specific needs.

Ideally, analytics are implemented using a common set of tools that can be tailored to generate outputs and outcomes specific to programs. The tool should be separate from the repository and sufficiently agile to accommodate requests for new analysis of existing program data or new analysis of data across programs. As new data are collected in the integrated PIR, analytics are extended to generate reports based on this new data.

A Corporate Partner Advisory Group/ AGA report⁵ that examined data analytics identified the following descriptive (i.e., describe the population being examined) and predictive (i.e., use patterns discovered by examining historical data to identify previously unidentified risks and opportunities) analytic techniques:

- Descriptive
 - Standard Reporting
 - Custom Reporting or “Slicing and Dicing: the Data”
 - Queries/ Drilldowns
 - Dashboards/ Alerts (Business Intelligence)
 - Statistical Analysis
 - Clustering (Unsupervised Learning)
- Predictive
 - Predictive Modeling
 - Optimizations and Simulations
 - Next Generation Analytics: Text Mining and Link Analysis

Understanding the analytic techniques that may be employed will help PIR architects to better understand how the repositories will be used. Additional discussion of operational use is provided in Section 6.

6 Integrated PIR Vignettes

The following vignettes highlight different users of integrated PIRs.

- Standard program reporting: State program administrators manage the upload of program performance data stored in their state integrated PIR into a federal PIR. Administrators are able to retrieve and review data that has been previously loaded. Information required by different programs is only loaded once. An automated monitoring capability utilizes the PIR to detect anomalies that could be indicators of fraud.

⁵ Corporate Partner Advisory Group/ Advancing Government Accountability (AGA), “Leveraging Data Analytics in Federal Organizations” Report No. 30, May 2012.

- Program performance analysis: Analysts, directly supporting a federal program office, conducting grant research, etc., access a federal PIR to obtain performance data from different states. Analysts may require data from all states or they may focus on selected states or a specific region. Analysts may also obtain additional data directly from state PIRs.
- Analysis of cross-program interactions: A 3rd-party research entity is funded to evaluate the cumulative impact of human services programs on a target population. Researchers are granted access to performance data stored in the PIRs. Researchers need to associate information from different sources with specific clients, although they do not require actual personally identifiable information; retrieval services perform matching for users. This effort is aligned with the following OMB guidance:

*Agencies should also consider seeking authority from Congress, through the FY 2014 budget process, to allow new waivers linked to evaluation or to **establish cross-agency "performance partnerships" that enable blending of multiple funding streams to test better ways to align services and improve outcomes.** Several agencies are seeking such authority in 2013 for initiatives supporting distressed communities and disconnected youth".⁶ [emphasis added]*

- Federal funding decisions:

Federal agencies must decide which projects to fund. Increasingly, they are looking for proposals with strong, evidence-based support. Per OMB guidance M-12-14:

*"OMB invites agencies to identify areas where research provides strong evidence regarding the comparative cost-effectiveness of agency investments. The research may pertain to the allocation of funding across agency programs (e.g., research showing that some funding streams have higher returns on investments) or within programs (e.g., research showing that some types of grantees or programmatic approaches have higher returns). Agencies should describe the body of research and then apply its results to support a proposed resource reallocation. **OMB is more likely to support an existing resource allocation or a request for new resources supported in this way, and may feature the agency's reasoning in the 2014 budget.**" [emphasis added]*

"Several agencies - ranging from the Department of Education to the U.S. Agency for International Development - have implemented evidence-based grant programs that apply a tiered framework to assess the evidence supporting a proposed project and to determine appropriate funding levels. Under this approach, programs supported by stronger evidence, as established in a rigorous agency process, are eligible for more funding. All programs are expected to evaluate their results. Examples of tiered-evidence programs include the Department of Education's Investing in Innovation program and

⁶ M-12-14 "Use of Evidence and Evaluation in 2014 Budget", Executive Office of the President Office of Management and Budget, Memorandum to the Heads of Executive Departments and Agencies, May 18, 2012.

the Department of Health and Human Services' Teen Pregnancy Prevention and Home Visiting programs.

*Even without creating tiers, **agencies can provide points or significant competitive preference to programs that the agency determines are backed by strong evidence, and can build the evidence base by embedding evaluation into programs.** Because running evidence-based programs requires more resources, agencies may wish to combine multiple smaller programs into larger, evidence-based efforts.” [emphasis added]*

- State budget planning:
State administrators facing funding shortfalls need to evaluate program performance to determine how to best utilize available funds.

References

M-12-14 “Use of Evidence and Evaluation in 2014 Budget”, Executive Office of the President Office of Management and Budget, Memorandum to the Heads of Executive Departments and Agencies, May 18, 2012.

“Leveraging Data Analytics in Federal Organizations” Report No. 30, Corporate Partner Advisory Group/ Advancing Government Accountability (AGA), May 2012.

Federal Registry Part II Department for Health and Human Services, Administration for Children and Families, 45CFR Parts 261, 262, 263, and 265, "Reauthorization of the Temporary Assistance for Needy Families (TANF) Program", Final Rules, February 5, 2008.

"NHSIA Infrastructure Viewpoint", Johns Hopkins University Applied Physics Laboratory, Draft Version D0.1, September 2011.