# All-Payer Claims Database (APCD) Technical Build Guidance Document

**July 2011**

## Background

States that are in the planning and early implementation phases of building an All-Payer Claims Database (APCD) often seek guidance regarding the technical build and management of these large-scale databases. The impetus for technical assistance may be new legislation authorizing the establishment of an APCD reporting program (but prior to specific rule making for data submission and release) or it may be an expansion to existing reporting laws and requirements. In either case, states benefit from the lessons learned in other states with established APCD systems.

This overview was developed by the All-Payer Claims Database Council (APCD Council) which is comprised of the National Association of Health Data Organizations (NAHDO) and the New Hampshire Institute for Health Policy and Practice (NHIHPP) at the University of New Hampshire. Funding for this paper was provided by The Commonwealth Fund. This is a living document, and as more states gain experience in implementing APCDs this document will evolve to include additional insights and lessons learned.

*Common APCD Technical Build Considerations*

- *Conduct Inventory of the Insurance Market*
- *Develop Submission Rules for APCD*
- *Data Management-System Development and Maintenance*
- *Data Consolidation and Validation*

The information in this document is intended to guide the planning and implementation discussions in states, based on actual APCD state experiences. This information address common questions state data agencies ask the APCD Council to address during the technical build stages of their APCD initiatives. Because states vary in their computing environments, technical capacity, and reporting laws, this document addresses issues at a general level. Specific technical assistance questions can be directed to the APCD Council by individual states (apcdcouncil.org).

## Conduct Inventory of the Insurance Market

States embarking on an APCD initiative must inventory and assess the state's insurance market as an essential first step. This information will guide the planning, budgeting, and technical build decisions that follow. Each state's market differs in its structure (number of commercial payers, size of the state's population, and mix of carriers). Consultation with state Insurance Department (which licenses/regulates commercial carriers); the Medicaid and State Children's

Health Insurance Program (SCHIP) program, which contracts with carriers and Pharmacy Benefits Managers (PBMs); and other public payers for enrollment estimates will help define the scope and staging of the state's APCD development. In addition, obtaining enrollment numbers for Medicare and the estimated uninsured population provides a more robust market view, and an understanding of populations that are not included if the APCD targets commercial and Medicaid data.

One of the most important determinants of the cost of a statewide APCD is the number of data sources and data feeds that are expected to supply information to the state agency. The cost of the system increases as the number of data sources (payers) increases, especially in the early phases. Each data source and platform must be assessed, normalized or mapped into a common uniform format across all sources, and tested for accuracy. One payer can maintain multiple computing platforms, which multiplies the intensity of the effort.

The inventory generally gathers the following information from each carrier:

- Total number of platforms for each data source (medical, dental, pharmacy, eligibility)
- Types and numbers of carve-out services (e.g. pharmacy, mental health) and other contractual relationships
- The data formats/layouts for each file type for each platform

**Where to Begin**

Most states begin the development of an APCD with the commercial carriers, because those data represent the largest percentage of the population (compared to public payers) and the carriers can be identified through licensing/regulatory documents. Estimating the size of the enrollment/eligibility population of the commercial carriers will help guide decisions about the scope of the APCD. In most states, commercial carriers with larger enrolled population sizes are the initial submitters as a starting point for data aggregation. Figure 1 below is designed as a worksheet to assist in estimating the scope of the insurance market in a state.

**Figure 1. Insurance Market Estimate Worksheet**

| | | |
|---|---|---|
| Commercial Carrier 1 | Total past year enrollment | _____ |
| Commercial Carrier 2 | Total past year enrollment | _____ |
| Commercial Carrier X | Total past year enrollment | _____ |
| Commercial Carrier Y | Total past year enrollment | _____ |
| Third Party Administrators | Total past year enrollment | _____ |
| Pharmacy Benefit Managers | Total past year enrollment | _____ |
| | **Total Commercial Insured** | _____ |
| Medicaid FFS | Total past year eligibility | _____ |
| Medicaid HMO | Total past year eligibility | _____ |
| SCHIP | Total past year eligibility | _____ |
| | **Total Medicaid/SCHIP** | _____ |
| Medicare FFS | Total past year eligibility | _____ |
| Medicare Advantage | Total past year eligibility | _____ |
| Medicare Part D | Total past year eligibility | _____ |
| | **Total Medicare** | _____ |
| Estimated Other Carriers in past year | | _____ |
| **Total Insured Population Estimate** | | _____ |
| **Uninsured (estimate from surveys/state data)** | | _____ |

**Adding Public Payers**

In assembling an APCD, many states have incorporated Medicaid; some have included Medicare claims data. Having both private and public payers contributing to the APCD allows for comparison across payers. Including Medicare data allows for comparisons more comprehensively across all age groups.

Each state will need to evaluate their plans for public payer data. States have taken several approaches, but all have had to make decisions based on the following considerations:

- Will the APCD merge commercial with Medicaid and SCHIP claims data? If so, are resources available for the mapping of the Medicaid data to commercial claims data?
- Will the APCD incorporate Medicare data? Are resources available for the purchase and mapping of the Medicare claims data to the commercial claims data?

**Medicaid.** States have learned that first aggregating the data from commercial payers is relatively straightforward, as most payers follow national standards to some extent. State Medicaid agencies, many extracting data from their legacy systems, may have more unique/non-standard formats and definitions than commercial payers. Additional challenges are faced if Medicaid contracts with Medicaid Managed Care Organizations (MCOs) with capitated reimbursement arrangements, because claims for individual encounters may not be available. If possible, detailed encounter data, without charges, will be another source of Medicaid data. Each state Medicaid agency will vary in their MCO data reporting, requiring additional assessment on the part of the APCD data agency.

After the assessment and discovery stages, if Medicaid data are available, data elements must be mapped to the commercial claims structure to be incorporated into the APCD. Obtaining Medicaid data layouts for medical, dental, and pharmacy and working with the Medicaid information systems personnel to understand the definitions for each data elements will be essential. While many of the Medicaid data elements will map directly to the commercial elements, some will not. These distinctions must be made clear to the users of the data.

Merging Medicaid data with commercial claims data will also require approval from CMS, if it is to be released to the public. APCD data agencies must be aware that federal law requires that the use of Medicaid data must be for purposes directly connected with the administration of the program such as: establishing eligibility; determining the amounts of medical assistance; providing services for recipients; and/or conducting or assisting an investigation or prosecution related to the administration of the plan. Acceptable uses under these requirements can include: monitoring the movement of individuals from commercial coverage to Medicaid; conducting comparative analyses to determine if Medicaid beneficiaries are receiving the same level of quality and cost-effective care as the commercially insured; and analyzing utilization and cost trends for specific procedures.

**Medicare.** Obtaining and merging Medicare data with commercial claims poses challenges for APCD data agencies. The process can be lengthy and is typically accomplished through two different processes. For Medicare Part A and B data (Standard Analytical Files and ancillary files), a state must enter into a Data Use Agreement (DUA) with CMS. All requests for Part A and B data must go through an application process at the CMS funded Research Data Assistance Center (ResDAC) http://www.resdac.org/. An applicant must specify the intended use of the data and provide assurances for the privacy and security of the data. Upon completion of the application, the proposed use of the data will be reviewed and approved or denied by the CMS Division of Privacy Compliance. Federal law currently restricts the use of Medicare data to

research related purposes. This may limit how a state can re-release the Medicare data with other APCD data.

As with Medicaid data, it will be necessary to map the Medicare Standard Analytical Files, the Denominator File, and other ancillary files to the commercial format. This has already been accomplished by two states, Maine and Minnesota. Ideally, if a standard format is adopted for the commercial claims files and more states request the Medicare Part A and B data, it would be cost effective for the CMS Research Data Distribution Center to provide a data extract to the states in the commercial format. This approach would eliminate the time and effort spent by the states to transform the raw data.

**Medicare Data Acquisition: The State Experience**

Maine, Minnesota and Massachusetts have successfully applied for and received Medicare data for use with their respective APCDs. At the time of this writing, Vermont and New Hampshire are in the process of applying for RESDAC Medicare data files. Massachusetts has posted their RESDAC application for Medicare data on their APCD website[1].

Although Medicare Part D data can also be obtained from CMS through ResDAC, it is much more efficient and timely to obtain it, and the Medicare Part C data, directly from the commercial carriers to which those programs have been delegated. The participating carriers typically have the data stored in their warehouses in a parallel manner to the commercial data. In order to segregate the Medicare C and D data from the commercial claims data, it is necessary for the data submission specifications to include codes for the Medicare Advantage Part C program and the Medicare Part D drug program under the "Insurance Type/Product" elements of the eligibility, medical, and pharmacy claims files.

## Develop Submission Rules for the APCD

Once states have conducted an inventory of the insurance market and have identified the major carriers, they typically are ready to move onto the documentation phases that develop reporting rules and submission specifications. Critical to this step are discussions and technical workgroup meetings with all key stakeholders, including and especially payers, to define the reporting requirements for carriers that will be submitting their claims data to the authorized APCD agency. In addition, when defining submission rules a state needs to consider their authority to require/compel Third Party Administrators (TPA) and Pharmacy Benefits Managers (PBMs) to end them claims data, and to work with these groups as well.

**Accessing claims Data from PBM's; The State Experience**

States with existing APCDs report the importance of getting claims data directly from PBMs (opposed to MCOs that manger Medicaid in some states). It's important to first review the states statue definition of "health insurance issuer" to see whether it includes PBMs. It's a significant issue as PBMs have a large percentage of claims.

During the rulemaking process, states will need to:

- Define scope (thresholds) for initial carrier reporting, typically defined by the number of people covered by the insurer. Most states set thresholds so that carriers with very few covered members are not required to submit data to the system. The market structure and mix, including the number of carriers, will drive the decisions about which carriers will be required to report, which will be exempt, and under what criteria (enrollment size, annual premium total).
- Define file structure/file layout/formats.
- Define which platforms each payer must report and from which sources (eligibility, medical, pharmacy, dental).
- Define the schedule (monthly, quarterly, and annually) for submissions.
- Determine penalties for non-compliance of submissions. Compliance to reporting requirements is essential to all APCDs and will require continuous monitoring by designated compliance staff.

**Data Submission Rules.** Rules vary from state to state but many of the data submission rules contain similar components. Examples of APCD state regulations and their related submission rules as well as data element specifications can be found at www.apcdcouncil.org.

**Data Submission Specifications.** The APCD Council and APCD states have worked with the national insurance carriers, America's Health Insurance Plans (AHIP), the Agency for Healthcare Research and Quality (AHRQ) and other stakeholders to define a core set of APCD data elements (http://apcdcouncil.org/standardization), based on what is collected by several states and what is available in payer information systems. The core data set will help states in the early planning process by providing a common data format across all carriers and platforms.

National payers have programmed extracts for a growing number of state APCDs. To the extent that states adopt the common core data set for their APCD, payers incur less cost and require less time to create a unique state extract. A common format will also allow for the comparability of that state's data with other state APCD benchmarks in the future. In addition, using a common set of data may reduce the analytic costs for a state, as more tools/software are developed that use the common core data set.

## Data Management—System Development and Maintenance

Data management is the foundation of the APCD and includes the development and execution of architectures, policies, practices, and procedures that properly manage the full data lifecycle needs of an enterprise[1]. In building an APCD, every state should consider:

- Does the APCD agency have internal capacity to manage the database and relationships with carriers, or should this function be outsourced?
- Are there resources available to ensure high quality data, and/or improve data quality? Should these functions be supported in-house, or outsourced?

Data management activities can be conducted in-house by the data agency or they can be outsourced to a contracted vendor[2]. In both cases, the data management infrastructure requires hardware (APCDs are large-scale databases that require large data disc arrays and powerful servers), software, security protocols, and a trained technical workforce to build the databases and generate reports. States will vary in their approaches, but the essential components of a data management system include the following:

1. File Transfer Protocol (FTP)/internet system to receive submissions for each file type. The data agency (or its contracted vendor) must build a data system that can receive submissions from the data sources, for each file type, in the structure/format that has been approved in the rulemaking/specifications development process.

2. Edit functions/rules for every data element.

    a. **Load edits** are the basis to assuring the proper file structure and formats of data submissions to the data agency. The load edits are basic checks to assure that files conform to the approved structure, fits into the receiving portal, and contains the following:

        - Detailed header and trailer records (designating number of records in submission that equal the actual received record number)
        - Correct field lengths in the required sequential order

[1] Data Management Association International, available at: http://www.dama.org/i4a/pages/index.cfm?pageid=1

[2] The APCD council maintains a list of vendors at http://www.apcdcouncil.org/vendors

- Each required data element field is present and filled with the correct data type (integer/alpha) and to a level of completeness, or threshold (100%, 98.5%, 50%, etc.), determined by the data collection entity based upon the characteristics of each data submitter. The data collection agency or vendor must work with its data suppliers (carriers, TPAs, PBMs) through test submissions and two-way conversations to establish thresholds for accepting or rejecting a submission. Initially, thresholds may be set low and then raised as baselines are established and data content improves over time. High thresholds for patient demographic data elements (date-of-birth, gender, patient numbers) and financial and diagnostic/procedure coding fields can be set. Other fields may not be uniformly collected/retained across payers (e.g. race/ethnicity, market codes). Once the permanent thresholds are established, elements with incomplete or invalid data are rejected and must be resubmitted. ***NOTE: ONGOING COMMUNICATION WITH EACH DATA SUPPLIER IS ESSENTIAL. THE DATA MANAGEMENT SYSTEM SHOULD HAVE A MECHANISM TO NOTIFY PAYERS OF PROBLEMS AND SUPPORT PAYER RESPONSES TO EXPLAIN VARIANCES, FAILED RECORDS, OR OTHER PROBLEMS WITH THE DATA.***

b. Establish a process for identifying and maintaining homegrown code lists. **Homegrown codes** are those codes that are developed and used by a single payer for their specific purposes. Even with national industry-based standards, local and homegrown diagnosis and procedure codes still exist.

c. **Quality edits** are important to assess and improve the content of the data submitted to the data agency. The data agency must work closely with each carrier to understand their data for each file type, troubleshoot problems, and set thresholds for errors. As with the load edits, over time these thresholds can also be increased for most data elements as the data agency identifies priorities for improvement and data submissions become standardized across payers.

d. **Trend edits** are a mechanism agencies us for monitoring data consistency within and across payers. One example of a trend edit is Per Member per Month (PMPM) trends which can be flagged if they vary by more than a set percentage by payer source by submission.

3. Data field transformation—recoding. In some cases, direct patient identifiers will need to be encrypted. In addition, new data elements can be created from the raw data submitted if there are fields that need to be converted before data release (e.g. - date-of-birth to age categories, assign county codes).

4. Automated flags for outliers and rare codes/services. Outliers are data that appear to be inconsistent with the balance of the collected data in the database. Identifying records with unusual patterns (high or low values) should be flagged for further study, as outliers may indicate rare events or potential errors in the data. For example, a hospital stay of 100 days or negative charges or payments may be actual occurrences or reporting errors. For data consistency and quality, these extreme values should be researched.

5. Management of provider and physician fields. Most APCDs will collect and use physician and provider identifiers for a range of analysis and applications that include health system assessment, pricing and outcomes comparisons, and consumer information. States with APCDs are struggling with provider identification. Payers vary in how they code and report provider fields. One of the most problematic issues currently impacting the accuracy of provider identification is the substitution of billing provider for the rendering provider on the claim. While many health care facilities can be both the billing and service provider, this is not the case with practitioners. Many practitioners are part of group practices, and the group practice, rather than the individual practitioners, is identified as the rendering provider. Although the National Provider Identifier (NPI) can be required, the NPI does not include information about employment relationships (participation in a group practice) or site of services. Additionally, examples of a provider having multiple NPIs have been seen.

**Possible Future Collaboration to Develop Provider Directories**

With the advent of Health Information Exchanges (HIEs), we can expect overlapping interests and resources to create single, uniform provider directories. With a robust and accurate provider directory, ideally, all the data agency would need is the provider name and the NPI---not multiple fields for specialty codes, service/office location, and other identifying information. Using the NPI as the linkage mechanism, these data would be found in the National Plan and Provider Enumeration System (NPPES). However, some states may want to assign their own specialty codes based on CPTs during the data management process, to assess specialty service mix and compare with NPI or state-assigned specialty codes.

States will typically require the payer to provide data from their internal provider files. The state can then consolidate these files into a directory that maps various payer numbers to the individual provider using the NPI, state license number, and names. This is an expensive and time-consuming process that not all states are able to deploy. Additionally, it does not solve the previously mentioned problem of distinguishing the primary (rendering) physician from the secondary (billing) entity. States must work with their payers and providers to standardize these coding practices over time.

6. Data integrity. Compliance to reporting requirements is essential to all APCDs. The system itself must be capable of tracking data supplier submissions and indicate to each data supplier relevant submission failures. Documentation of the files (eligibility, pharmacy, medical, dental), the platforms, and the relationship of the various data submitters (e.g., contractual carve outs for claims processing) is critical and will vary with each payer. Ideally, compliance requires continuous monitoring by a designated staff in the agency to keep track of each data supplier's submissions for all types of data for which the data supplier is required to report. Agencies with the authority to levy penalties for non-compliance to reporting requirements have leverage to compel accurate reporting. Data quality issues may require a more iterative approach between the agency and data suppliers to identify priority data and coding problems and correct these going forward.

**Data Consolidation and Validation**

Once editing and cleaning of the data are complete, the data agency will combine edited data and create analytic files and data output reports. Typically this output includes the following:

1. Analytic master files are created for each aggregated data type
2. Grouping category assigned (for states using software groupers)
3. Suppression of restricted fields (such as all line item charges, in some states)
4. Frequency/output reports are generated for each payer for each data source and total
5. Payers receive their frequency/output totals for key fields to review, verify, and update as needed/required. Many states will have provisions in law that require a data supplier review and validation period (e.g. 45 days). During this time, each data supplier will have the opportunity to review the frequency/output reports for their submissions against the total database and identify discrepancies in their data. ***NOTE: EVEN IF A REVIEW AND VALIDATION PERIOD IS NOT REQUIRED BY LAW, THIS STEP IS***

***RECOMMENDED AS A WAY TO IMPROVE THE DATA AND BUILD TRUST WITH DATA SUPPLIERS. PROBLEMS WITH THE DATA CAN BE IDENTIFIED PRIOR TO CONDUCTING ANALYTICS AND REPORT PRODUCTION.***

6. Codebooks/data dictionaries/data specification manual are created for each analytic file/product

**Summary**

Collecting and aggregating claims data files across payers is a complex process, with technical and political challenges. States with established APCD reporting programs have proven that these challenges can be overcome. Factors that are key to a successful APCD implementation include an understanding of the state market structure and robust stakeholder involvement in the planning, design, and implementation of an APCD initiative. States have learned that the reporting specifications must be aligned with payer system capabilities and that data quality improves over time with consistent feedback and direct consultation with each data supplier's technical staff. Because states vary in their computing environments, technical approaches, and reporting requirements, no single approach will work in every state. States must design the processes and specifications that address their unique situations. The APCD Council and NAHDO will work with individual states to provide more detailed solutions as needed. We invite APCD data agencies and others to share their lessons learned with other states through the APCD Council and NAHDO.

*This document was developed by the All-Payer Claims Databases Council (APCD Council) which is comprised of the National Association of Health Data Organizations (NAHDO) and the University of New Hampshire's Institute for Health Policy and Practice. Lead authors, Denise Love, Executive Director of NAHDO and Alan Prysunka, Executive Director of the Maine Health Data Organization.*