# Buried Treasure:
## Freeing California Data to Improve Health Care

## Introduction

It was considered revolutionary when the National Oceanic and Atmospheric Administration (NOAA) began releasing weather data to the public for free more than 30 years ago. Today, NOAA data are used by researchers, weather forecasters, the insurance industry, and application developers to fuel services used by the public daily, such as The Weather Channel.

Following NOAA's example, the US Department of Health and Human Services (HHS) has embarked on a campaign to increase the amount of new data in the public domain, make data more accessible and usable, and publicize the availability of its vast supply of health information. HHS hopes to spur development of a health information ecosystem that will inform health care decisionmaking by patients, providers, payers, and policymakers.

Like HHS, California can improve the quality and efficiency of its health care system by leveraging its wealth of information. The state plays different roles in the health care arena and collects diverse sets of data:

- **Payer.** Claims data generated under Medi-Cal and other public health programs provide valuable information about the costs of providing health care services to California's most vulnerable populations.

- **Insurance regulator.** California receives a bounty of data from the different health plans that it oversees, which are captured and housed in different agencies, including the Department of Managed Health Care and the Department of Insurance.

- **Overseer of licensing, utilization, and seismic safety.** California's Department of Public Health and the Office of Statewide Health Planning and Development receive data from various health care providers related to professional certification, patient use of health care services, and seismic safety compliance.

The state also receives data from health care providers on specific measures, such as hospital-acquired infections and other adverse events, infectious diseases, and other reportable conditions. These and other data can help California evolve its health care system from a "pay for volume" methodology, which reimburses providers based on the volume of care they provide, to a "pay for value" model, which takes into account the quality of services provided. This shift is motivating health care providers to leverage data in their efforts to reduce errors, cut avoidable hospital admissions and readmissions, improve care coordination, and improve care quality in general.

This issue brief highlights federal and state efforts to "free the data," and outlines benefits to California were the state to follow suit.

## Background

On December 8, 2009, the White House issued an Open Government Directive, which calls upon federal agencies to be more transparent and collaborative.[1] HHS released its own plan — the HHS Open Government Plan Version 1.1 — in

ISSUE BRIEF

June 2010, in which it defined open government as one that is "transparent, publishing government data that generates significant benefit for citizens and which helps the public hold the government accountable."[2] Under that plan, which was updated in April 2012, HHS committed to the following:

- Publishing more government information online in ways that are easily accessible and usable

- Developing and disseminating accurate, high-quality, and timely information

- Fostering the public's use of the information

- Advancing a culture of data sharing at HHS[3]

Health care policy experts suggest that measuring and publicly reporting health care provider performance and community health information, and providing patients with better access to their own health information, will be critical to improving care quality while reducing costs.

## HHS Data Sharing Efforts

While HHS health care data has long been publicly available for research and quality improvement efforts, it has generally been difficult to access and manipulate.[5] Recent efforts by HHS to free the data are focused on making previously available data more accessible and usable, and on releasing new data. At the same time, HHS is encouraging innovative new uses of the data by the public.

**Excerpt from HHS Open Government Plan Version 1.1:**

[T]ransparency and data sharing are of fundamental importance to our ability to achieve HHS's strategic goals of advancing the health and well-being of the United States. HHS's vast stores of data are a remarkable national resource which can … help the public hold the private sector accountable, increase awareness of health and human services issues, generate insights into how to improve health and well-being, spark public and private sector innovation and action, and provide the basis for new products and services that can benefit the American people.[4]

## Health.Data.Gov

The Health.Data.Gov website is the centerpiece of HHS's larger Health Data Initiative, a public-private collaboration led by the Institute of Medicine and HHS to build a network of community health data suppliers (e.g., HHS) and data appliers (e.g., researchers, policymakers, application developers). The Health Data Initiative enables data appliers to use community health and other data to create applications that (1) raise awareness of the health of a community, such as tracking rates of diseases like obesity or diabetes; (2) increase pressure on decisionmakers to improve health system performance; and (3) help facilitate and inform action to improve health system performance.

Under the initiative, HHS provides to the public — through the Health.Data.Gov website — national, state, and county level health care data, and groups it by age, gender, race/ethnicity, and income. Data are harvested from across the agency and provided in formats that are easy to use; the data are standardized, structured, downloadable, and accessible via application programming interfaces (APIs). The data consist of hundreds of measures of health care quality, cost, access, and public health that are in machine-readable formats, such as Microsoft Excel files.

HHS is encouraging the following potential uses of the data, among others:

- Social networking applications that allow health improvement leaders to connect with each other, compare performance, and share best practices

- Online games that incorporate community health education messages

- Use of community health data to help improve the usefulness of consumers' online health searches

- Use of data from the Blue Button Initiative — a federal program enabling veterans and Medicare beneficiaries to download their health information — by mobile application and software developers

to create health management tools, such as reminders to get preventive services or refill a prescription

Under the Health Data Initiative, innovators have already developed a number of applications using HHS data, including an interactive web-based tool, also known as a dashboard, which allows civic leaders and citizens to see a report card of their community's health.[6] HHS publicizes the availability of this data through application competitions and code-a-thons, where participants are challenged to use the data to develop new programs and projects over a set period of time.

## Centers for Medicare and Medicaid Services Dashboards

The Centers for Medicare and Medicaid Services has developed a series of dashboards to help researchers, policymakers, and the public evaluate the quality and efficiency of Medicare:[7]

- The **Medicare Inpatient Hospital Dashboard** uses Medicare inpatient prospective payment system data to help users analyze Medicare spending on inpatient hospital care.

- The **Medicare Part B Physician/Supplier Dashboard** provides statistical views of outpatient health care services and spending under Medicare Part B, which covers outpatient services.

- The **Medicare Prescription Drug Benefit Dashboard** uses prescription drug event (PDE) data collected under the Medicare drug benefit to help users analyze prescription drug use and spending under Medicare.

- The **Medicare Enrollment Dashboard** provides detailed information on recent Medicare enrollment patterns at both the national and state levels as well as historical enrollment trend data dating back to the program's inception in 1966.

For example, using the Medicare Inpatient Hospital Dashboard, individuals can track and graph Medicare spending on inpatient hospital services by state. Users can also see how much Medicare is spending on the provision of medical education, additional funding to hospitals that care for a disproportionate share of low-income patients, and additional payments to rural hospitals.

## Medicare Claims and Prescription Drug Event Data for Qualified Entities

To evaluate the performance of providers that care for Medicare patients, the Affordable Care Act (ACA) requires CMS to make standardized extracts of claims data and PDE data available to qualified entities — public or private organizations that meet a set of CMS requirements.[8] Qualified entities can combine the Medicare data with private sector claims data to develop and disseminate public reports measuring the performance of physicians and other health care providers.[9] According to CMS, qualified entities are likely to resemble community quality collaborative program participants, such as those that were involved in the CMS Better Quality Information for Medicare Beneficiaries pilot.

Qualified entities will have access to Medicare claims and PDE data broken down by region. Data will be de-identified and encrypted with a unique beneficiary identifier that will allow for the linking of claims without divulging the identity of the beneficiary. However, under certain circumstances, patient-identified data will be available to help Medicare providers verify and correct performance data.

To date, private sector entities have had few opportunities to access patient-level data associated with specific health care providers. The availability of Medicare data will increase the public's ability to evaluate and compare health care provider performance.

## Medicare Claims Data for Accountable Care Organizations

CMS is also encouraging data-driven health system performance improvement by making data available to accountable care organizations (ACOs), which are groups of providers jointly held accountable for improving quality and reducing the cost of care. The ACA is testing new health care payment and delivery models through two ambitious ACO programs — the Medicare Shared Savings Program and the Pioneer ACO Demonstration Program.

Information about beneficiaries' health and health care is central to ACO efforts to evaluate provider performance and quality improvement activities. To help ACOs understand the breadth of care provided to their beneficiaries, CMS will provide ACOs with access to a variety of Medicare claims and other data.

De-identified, aggregated data on beneficiary use of health care services will help ACOs better manage their beneficiary population's use and spending patterns. For example, if data analyses show that an ACO's beneficiary population has a high rate of hospital readmissions, the ACO could work to improve discharge coordination among its attending physicians, hospitals, and post-acute care providers. CMS will also provide ACOs with certain patient-identifiable claims data generated under Medicare Parts A, B, and D. These data will enable an ACO to establish baseline levels of beneficiary use, identify specific populations for proactive care coordination, and track progress against defined performance measures.[10]

## Early State Efforts to Free the Data and Recommendations for California

State government efforts to free the data are just beginning, and some have made early strides.

### New York — Maximizing Essential Tools for Research Innovation and Excellence

In New York, the State Department of Health (DOH) launched a project to provide researchers, entrepreneurs, and community-based organizations with access to DOH's health care data — the METRIX (Maximizing Essential Tools for Research Innovation and Excellence) Project.[11] The project's goals are to:

- Increase transparency of government and improve access to DOH data

- Identify high priority areas where analytics can support the state's health reform strategy

- Create a streamlined process to make DOH data available for innovation, economic development, and research with appropriate protections relating to privacy, security, and data use

- Expand policy analysis and research in key public health and health services delivery areas

- Contribute findings to the scientific evidence base, which will be used to improve public health and health services delivery programs' performance and improve health outcomes.[12]

Under the project, DOH has released a number of data sets. For example:

- The New York Adult Tobacco Survey, which assesses adults' tobacco purchasing and cessation behaviors and attitudes toward and exposure to secondhand smoke.

- The state's nursing home bed census, a weekly report on nursing home bed availability.

- The Behavioral Risk Factor Surveillance System data set, which can be used to explore the relationships between risk factors, such as poor mental health and inadequate sleep, on outcomes like obesity in adults.

Along with the data sets, DOH provides information about how the data are collected, how DOH uses the data, how frequently it will be updated, and suggests uses by the public.

## New Hampshire's Comprehensive Health Care Information System

New Hampshire is also taking steps to share its health care data. New Hampshire operates an all-payer claims database (APCD) called the Comprehensive Health Care Information System (CHIS) through which it makes claims and eligibility data from New Hampshire Medicaid available to the public.[13] CHIS aggregates claims data across public and private payers to provide a systemwide view of health care cost, quality, and access.

CHIS provides information in a variety of formats, from enrollment and claims payment reports to more detailed and complex reports on disease cohorts and use of related preventive and other health services. The reports and the CHIS website are produced by Onpoint Health Data (formerly the Maine Health Information Center) under a contract with the New Hampshire Department of Health and Human Services, Office of Medicaid Business and Policy.

Other states — among them Maine, Tennessee, and New York — are also operating or have plans to develop an APCD. APCDs can provide a wealth of information, including which hospitals have the highest prices, how far people travel for which services, and which parts of the state have greater numbers of specialists.

## How Freeing the Data Can Help California

California will make great strides towards its health policy objectives — to improve care and reduce costs — if patients, providers, payers, and policymakers can make evidence-based, data-driven health care decisions. While various state agencies do release health care claims data for research and analysis, further efforts can be made to free the state's valuable data for broader use.

Below are five ways California can improve how it shares health care data for the benefit of the state's health care system.

## 1. Make health care claims data more accessible for research.

Research based on DHCS claims data can support analyses of the effectiveness of DHCS's health care programs. It can also identify health care use and cost trends across providers and support public reporting of providers' performance against specified quality measures.

DHCS already provides some aggregated, de-identified data to the public through its Research and Analytical Studies Section website, which includes high-level Medi-Cal enrollment and utilization statistics.[14] DHCS also grants researchers' requests for health care claims data on a one-by-one basis. Requests are evaluated by the DHCS Data and Research Committee, and data sets are custom-made to fit each approved request.[15]

Developing large, aggregated, de-identified data sets and posting them on a public website would allow researchers to access data without making a formal request and having to wait for a response. This would free up valuable DHCS resources spent reviewing and triaging data requests from individual researchers. DHCS has begun to explore development of a portal that will make these data more readily accessible for research. In this process, DHCS will need to consider how to provide de-identified claims and encounter data that is meaningful to the research community, and how to create linkages between these and other data sets that are of value to researchers. Creating these matched data sets will be challenging as some data sets may reside in other systems and therefore not be easily accessible, may be managed by other departments, or may have dissimilar beneficiary identifiers.

## 2. Make health care claims data more accessible for demonstration programs.

California is currently supporting a number of demonstration programs to test alternative health care delivery and reimbursement mechanisms. The goal of these programs is to identify ways to save money

and improve the quality of care for recipients of state-funded or managed health care services. In this process, the state has begun to provide aggregated data files to demonstration participants.

For example, under the state's Section 1115 Medicaid waiver, DHCS is supporting five California Children's Services (CCS) demonstrations. By piloting the redesign of the CCS program, California aims to identify delivery models, including pediatric ACOs, medical homes, and case management programs for high-need children, that will result in a well-integrated, coordinated, and value-based health care system. Making up-to-date agency health care data available would enable the state's pediatric ACOs, for example, to better identify high risk patients, proactively engage and manage their care, and better coordinate care with other health care providers that are treating this population. DHCS has provided a data book, in Microsoft Excel format, for demonstration providers that aggregates data by county, condition services, and other variables.

Under another program, DHCS will implement integrated care pilot programs for dual-eligible Medi-Cal beneficiaries in at least four counties.[16] These pilots will integrate health care services for individuals who are eligible for both Medicare and Medicaid benefits — ensuring that appropriate services are delivered to the right people at the right time. The pilots will be evaluated to determine if they have increased the appropriate use of services by beneficiaries, increased beneficiary satisfaction, improved health outcomes, and reduced the total cost of care. Managed care organizations and health care providers participating in these pilots could benefit from access to administrative data, including hospital, outpatient, pharmacy, and other claim types. These data would provide a comprehensive view of each beneficiary's care and utilization history, which would in turn enable better care coordination, medication reconciliation, discharge planning, and follow-up care for this high-risk population.

DHCS currently shares health care data with county participants in the Low Income Health Program (LIHP), another Section 1115 waiver program. LIHP offers county-operated health coverage to uninsured, childless adults in preparation of the forthcoming Medi-Cal expansion mandated by the ACA. The University of California, Los Angeles, is using DHCS claims data to develop dashboard reports for each county and for the statewide program.[17] The reports include enrollment data, enrollee demographics, enrollee utilization data (e.g., average inpatient visit length of stay), and care quality measures. The dashboard reports allow each county to evaluate its performance against the performance of the other participating counties.

While the state has made significant progress in releasing some of the data at its disposal to county, provider, and health plan demonstration participants, a number of obstacles remain that prevent the state from providing more comprehensive information.

First, in many cases, relevant data are managed or housed by multiple departments. For example, while DHCS is in charge of most claims data, the Department of Mental Health is responsible for mental health encounter and claims data. To provide a complete and accurate picture of a beneficiary's utilization or total cost of care, the state needs to devise a consistent set of data release policies across its health departments so that program participants can more effectively understand and act on a beneficiary's needs.

Second, different data types are governed by different state laws and regulations, which in some cases are ambiguous and difficult to interpret. The state needs to define a clearer and more transparent set of policies regarding the privacy and security of the protected health information it maintains, while taking into account existing state and federal laws and the privacy preferences of its beneficiaries.

Finally, the state should provide individually identifiable information to demonstration participants for all beneficiaries that have enrolled in their pilot programs. Providing regular updates in a timely manner will allow participants to better understand utilization patterns, identify high-risk and high-need beneficiaries, and develop targeted interventions to effectively manage their care. An investment in resources may be required to ensure that data can be delivered in a consistent, timely manner that respects the privacy preferences of beneficiaries participating in state-supported demonstration programs.

## 3. Create an APCD to support new health insurance risk adjustment programs.

All-payer claims databases, which typically include claims data from medical, eligibility, provider, pharmacy, and dental files from private and public payers, are increasing the transparency of health care cost and quality information and informing health policy decisionmaking in states across the country.[18] According to a project funded by the California HealthCare Foundation, stakeholders support the development of an APCD in California but are looking for a compelling business or policy rationale to justify its costs.[19]

As part of its larger provisions to expand health insurance coverage, the ACA authorized the creation of state-based risk adjustment programs to mitigate health plan risk for providing coverage to higher cost enrollees — typically people in poor health or those with pre-existing conditions. Risk adjustment ensures that insurance plans can support the people who need care the most by providing more funds to plans that cover people who are likely to have high health care costs. The ACA authorized states to establish a risk adjustment programs that apply to health plans in the small group and individual markets both inside and outside of their developing health benefit exchanges — the two markets where health benefit exchanges have authority to initiate operations.

Risk adjustment programs require extensive data on enrollees' medical conditions, demographics, and other health information — much of which could be provided through an APCD — in order to estimate each enrollee's cost to the health plan.

The federal government recently released a final regulation indicating that data for risk adjustment calculations should be collected at the state level.[20] It also released a companion impact analysis suggesting that states with APCDs will likely be able to leverage that infrastructure to reduce the administrative costs of collecting and analyzing the data necessary to perform risk adjustment calculations.[21]

California should determine how to bring the best available data to bear on its risk adjustment activities, including determining which agencies should contribute which data sets to the effort and examining how an APCD could help. For example, development of an APCD would reduce the administrative burden to Medi-Cal, other public programs, and private health plans that may be required to comply with the ACA's state-based risk adjustment program provision. Data from an APCD could also be used to support a host of other reporting activities focused on pay-for-performance, quality, utilization, and costs.

## 4. Create a robust, publicly available health care data portal.

The federal government's Health.Data.Gov website provides hundreds of data sets as well as descriptive information for each data set (e.g., user views and downloads), and makes the data sortable along a number of dimensions. Health.Data.Gov is easily navigated and has sparked many creative uses of government data.

In contrast, California's Data.Ca.Gov website lists only 47 data file items, most of which link to data sources or state agency websites. Users must sift through and locate the

information they seek from a variety of sources. California should develop a more user-friendly portal to contain its disparate data sets.

## 5. Support linked data and APIs to enable better access to and use of data.

Simply providing access to data is not sufficient if it cannot be readily transformed into usable and actionable information. To effectively support a broad variety of uses (e.g., research, analysis, and application development) the data must be machine-readable (e.g., not as a PDF or image file) and readily accessible through a stable distribution environment, such as the Internet. The data must also be in a format that is structured and allows for simple extraction, translation, and loading into other applications where it may be manipulated in various ways. Data that conform to these principles, which have been adopted by the federal government, are collectively referred to as "linked data."

California can make its publicly available data more accessible and user-friendly by adopting linked data principles. This would include providing metadata — a detailed description of the data's content. Metadata explains the nature of each data field, reduces ambiguity as to the data's purpose, and allows both humans and computers to interpret the data's content. Finally, the state's approach should be platform independent; the data should be accessible and presentable from a variety of systems (e.g., computers and mobile devices) and software (e.g., applications and operating systems). This can be accomplished by supporting APIs, which allow information to be directly accessed and presented in a mobile phone application, for example. The federal government's Health.Data.Gov website uses lightweight, standard APIs that can serve as a model for California.

## Conclusion

The federal government is working to spur development of a health information ecosystem that will inform health care decisionmaking by patients, payers, providers, and policymakers. Several states are beginning to follow suit, including California, which has already taken steps to free its valuable health care data for the public's benefit. The data sharing proposals outlined in this issue brief build on this forward momentum and suggest ways the state can leverage its health care data to improve existing and future state-based and private health care initiatives.

### AUTHORS

Jonah P.B. Frohlich, managing director, Manatt Health Solutions

Susan R. Ingargiola, director, Manatt Health Solutions

### ABOUT THE FOUNDATION

The California HealthCare Foundation works as a catalyst to fulfill the promise of better health for all Californians. We support ideas and innovations that improve quality, increase efficiency, and lower the cost of care. For more information, visit us online at www.chcf.org.

## Endnotes

1. Peter Orszag, director, Office of Management and Budget. Memorandum 10-06: "Open Government Directive." December 8, 2009.

2. US Department of Health and Human Services. "HHS Open Government Plan, Version 1.1." Revised June 25, 2010. See also US Department of Health and Human Services. "HHS Open Government Plan, Version 2.0." Released on April 9, 2012.

3. Ibid.

4. Ibid.

5. Ibid.

6. Todd Park and Ahman Bandari. "Harnessing the Power of Information to Improve Health." June 4, 2010. For more examples of innovative applications that use HHS data see http://www.whitehouse.gov/blog/2010/06/04/harnessing-power-information-to-improve-health.

7. CMS dashboards are available on the Health.Data.Gov website and at http://www.cms.gov/Dashboard/01_Overview.asp#TopOfPage.

8. Effective January 1, 2012, Section 10332 of the ACA requires CMS to make available standardized extracts of Medicare Parts A and B claims data and Part D prescription drug event data to qualified entities.

9. 76 Fed. Reg. 76542 (Dec. 7, 2011).

10. Beneficiaries may opt out of having their patient-identifiable data shared. See 76 Fed. Reg. 67802 (Nov. 2, 2011).

11. See New York State Department of Health METRIX project website at http://health.ny.gov/METRIX.

12. Nirav R. Shah, MD, MPH, commissioner of the New York State Dpartment of Health. "METRIX Overview." December 2011.

13. See New Hampshire Comprehensive Health Care Information System (CHIS) website at http://www.nhchis.org/index.html.

14. The DHCS Research and Analytical Studies Section website is available at http://www.dhcs.ca.gov/dataandstats/statistics/Pages/RASS_Default.aspx.

15. For more information about the process for requesting DHCS data for research or public health purposes, see "Accessing DHCS Protected Data for Research and Public Health." Available at http://www.dhcs.ca.gov/dataandstats/data/Pages/AccessingProtectedData.aspx.

16. DHCS is currently considering providing participants in these integrated care pilots with data that would enhance their ability to manage high-risk enrollees.

17. UCLA Center for Health Policy Research. "Low-Income Health Program Evaluation Frequently Asked Questions & Appendix: Proposed Evaluation Publications and Products." Version 13. December 5, 2011.

18. See the APCD Council's website at http://www.apcdcouncil.org/state/map for a list of states that either have or are implementing APCDs.

19. California HealthCare Foundation. "Briefing — All Payer Claims Databases: Lessons from Other States." Available at http://www.chcf.org/events/2011/briefing-allpayer-claims-databases-lessons-from-other-states.

20. 77 Fed. Reg. 17220 (March 23, 2012).

21. Centers for Medicare and Medicaid Services, Center for Consumer Information & Insurance Oversight. "Patient Protection and Affordable Care Act; Establishment of Exchanges and Qualified Health Plans, Exchange Standards for Employers (CMS-9989-FWP) and Standards Related to Reinsurance, Risk Corridors and Risk Adjustment (CMS-9975-P)." Regulatory Impact Analysis. March 2012.