

Agenda Item VI: Data Access and Release: Standard Limited Datasets

Chris Krawczyk, Chief Analytics Officer, HCAI

Jill Yegian, HPD Consultant, HCAI

For Today

- Context for and purpose of standard limited datasets
- Considerations in defining standard limited datasets
 - Market segments
 - Personal identifiable information
 - Entity and financial indicators
- Proposed approach to HPD standard limited datasets

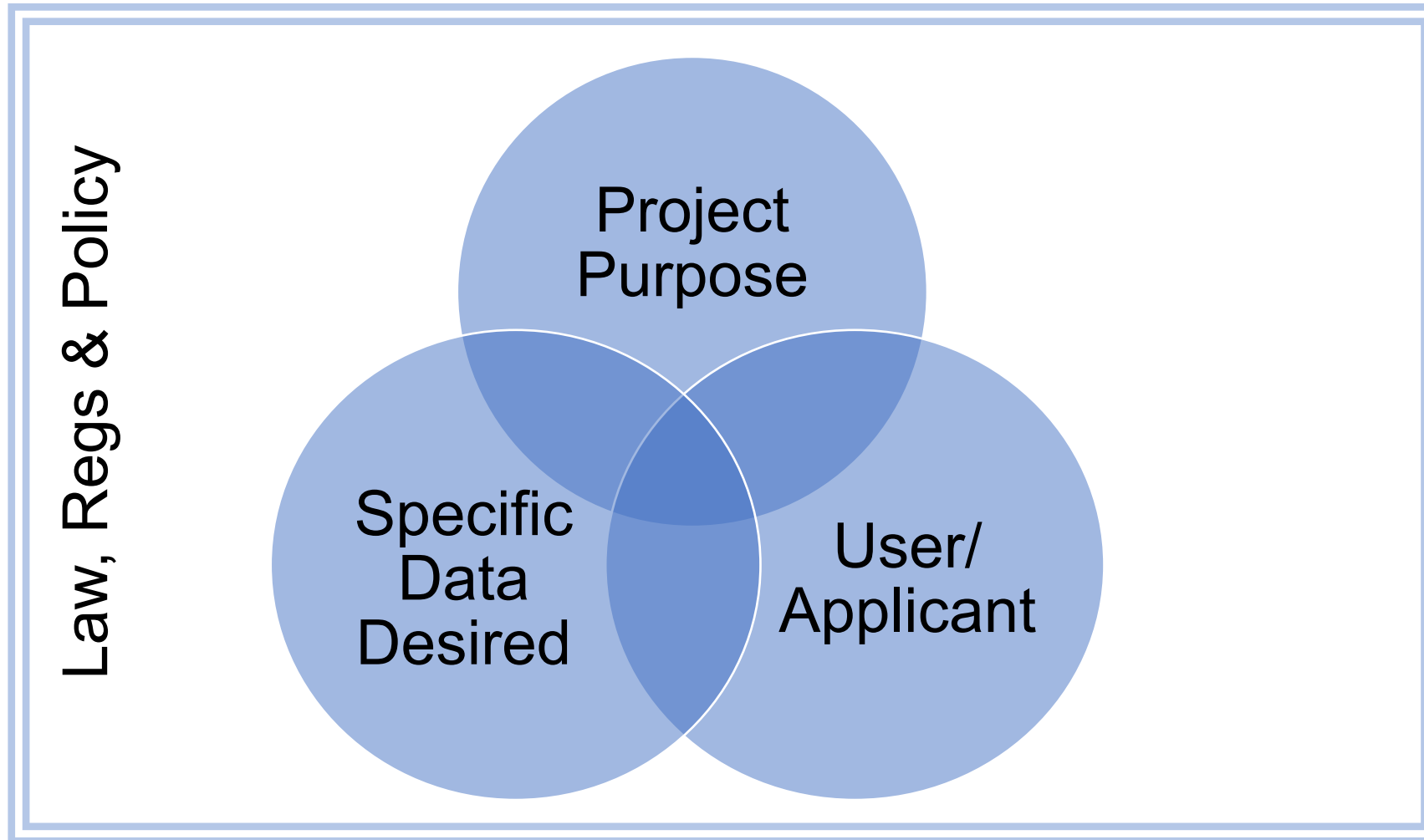
HPD Program Goals

1. Provide **public benefit** for Californians and the state **while protecting individual privacy**.
2. Increase **transparency** about health care costs, utilization, quality, and equity.
3. Inform **policy decisions** on topics including the provision of quality health care, improving public health, reducing disparities, advancing health coverage, reducing health care costs, and oversight of the health care system and health care companies.
4. Support the development of **approaches, services and programs** that deliver health care that is cost effective, responsive to the needs of Californians, and recognizes the diversity of California and the impacts of social determinants of health.
5. Support a **sustainable health care system** and more **equitable access** to affordable and quality health care for all.

HPD Data Access & Release Objectives

1. Protect patient privacy
2. Support program goals
3. Ensure appropriate data users, uses, methodologies, and compliance with all requirements
4. Provide timely data access to qualified applicants and make decisions in a consistent manner
5. Enable & support diverse users, representing multiple audiences & levels of expertise
6. Complement the HPD public reporting program
7. Support sustainability through user fees that balance program support and affordability

Intersecting Factors in Data Access Approvals



Framework for HPD Data Access

	Aggregated	Limited Data	Research Identifiable Data
Standardized	<ul style="list-style-type: none"> • Shortest approval process • Designed to meet needs of multiple applicant categories, use cases • Could rely on interactive tool that supports a query-based approach • DRC provides guidance on process • HCAI approves applications 	<ul style="list-style-type: none"> • Shorter approval process • Designed to meet needs of multiple applicant categories, use cases • HCAI-defined, no direct identifiers • DRC review required for transmission outside enclave • HCAI approves applications 	
Custom		<ul style="list-style-type: none"> • Longer approval process • Requestor-defined dataset • Content of dataset negotiated during review, no direct identifiers • DRC review required for transmission outside enclave • HCAI approves applications (tailored dataset + purpose + requestor) 	<ul style="list-style-type: none"> • Longest approval process • Requestor-defined dataset • Content of dataset negotiated during review • Direct identifiers rarely released • HCAI, CPHS, & DRC approve application (tailored dataset + purpose + requestor)

What is a limited dataset?

- The term “limited data set” comes from the HIPAA Privacy Rule – it is a data set that meets the criteria defined in the rule related to patient identifiers.
- **HCAI and HPD are not subject to HIPAA**, but the terminology associated with limited data sets has been widely adopted and is useful in the context of APCD and HPD data offerings.
- A limited data set **may contain some indirect** elements of personally identifiable information (PII) but **excludes 16 “direct” identifiers** (listed at right).

16 “direct” identifiers

1. Names.
2. Street Addresses.
3. Telephone numbers.
4. Fax numbers.
5. Electronic mail addresses.
6. Social security numbers.
7. Medical record numbers.
8. Health plan beneficiary numbers.
9. Account numbers.
10. Certificate/license numbers.
11. Vehicle identifiers and serial numbers, including license plate numbers.
12. Device identifiers and serial numbers.
13. Web universal resource locators (URLs).
14. Internet protocol (IP) address numbers.
15. Biometric identifiers, including fingerprints and voiceprints.
16. Full-face photographic images and any comparable images.

[Summary of the HIPAA Privacy Rule](#) available at HHS.gov

What is the purpose of standard limited datasets (SLD)?

- Protects patient privacy (no direct identifiers)
- Standard, pre-made limited datasets are easier to produce, while managing privacy risk
 - Custom limited datasets and custom research datasets take more effort to produce
- Shorter approval process, streamline access to the data for users whose needs are met with standard limited datasets
 - Requesters needing variables not available in the standard limited dataset may apply for a custom or research identifiable dataset

Considerations in Defining Standard Limited Datasets

Market Segment

What market segments should be included?

Personally Identifiable Information

How should indirect identifiers be handled?

Entity and Financial Information

How should allowed amounts and payer/provider identifiers be handled?

Market Segments

Market Segment	Restricted to State-Supported Research Users/Uses	DHCS Approval Required
Commercial		
Medi-Cal		
Medicare Advantage		
Medicare Fee-For-Service	X	

DHCS = Department of Health Care Services

Personally Identifiable Information (PII)

- Limited data sets **exclude all direct identifiers**.
- Limited data sets may include some **indirect identifiers** that are not included in de-identified data sets:
 - **geographic information** such as member city and 5-digit ZIP (in addition to 3-digit ZIP)
 - **day and month date elements** (in addition to year)
- Demographic information that is not considered to be PII (either direct or indirect) is included, e.g. race/ethnicity

Proposed approach balances between patient privacy and usefulness (granularity) of the data.

- Include 5-digit ZIP
- Date of service: provide full date (DD/MM/CCYY)
- Age in years (up to 90) or months (up to 24); not date of birth

Entity and Financial Information

- EFI includes:
 - Payer/Health Plan identifiers
 - Provider identifiers – health systems, hospitals, physician organizations, physicians and other billing clinicians
 - identifiers include name, address, NPI
 - Detailed payment information for specific health care services
 - allowed and plan paid amounts
 - patient responsibility (copay, coinsurance and deductible)
 - charged amounts
- HCAI will consider the potential for anticompetitive behavior
- EFI is not likely to impinge on patient privacy

Proposed approach provides detailed financial data, replaces payer and provider identifiers with codes that enable analysis.

- Include payment data (e.g., allowed amounts)
- Replace payer names/identifiers with codes
- Replace provider names/identifiers with codes

Key Elements of Proposed Approach to Standard LD

Content

- Meets a wide range of use cases
- PII: balance between patient privacy and usefulness of the data
- EFI:
 - Balance between potential for anticompetitive behavior and usefulness of the data
 - Release full financial detail, consistent with policy environment shifting toward greater transparency
 - Replace plan/provider identifiers with codes as guard against anticompetitive behavior, protect individual clinician identity/privacy

Structure

- Single version simplifies, reduces resource requirements to create and review
- Enclave environment will contribute to ensuring appropriate use
- Multiple process steps support appropriate use: application and review, DUA

Use Cases: Likely Possible Using Standard Limited Dataset

User	Purpose	PII Required	EFI Required
Reseacher, Policy makr	Analysis of variation in cost, utilization, quality, access to care by geography	5-digit ZIP, age, date of service (DD/MM/CCYY)	Allowed amounts
Policmaker, Multle	Analysis of population-level disparities in care based on available patient demographics	5-digit ZIP, age, date of service (DD/MM/CCYY)	May require allowed amounts
Reseacher, Policy makr	Analysis of cost drivers - identify medical and pharmacy services contributing the most to health care cost/spending growth and analysis of drivers	None	Allowed Amounts
Purcaser, Payer	Obtain geography-specific benchmarks for comparison to rates paid by purchaser for common costly procedures	May require ZIP	Allowed Amounts
Multle	Identify top 10 high-cost procedures, drugs, etc.	None	Allowed Amounts

Use Cases: Likely Would Require Custom Request

User	Purpose	PII Required	EFI Required
Publichealth invesigator	Analyze connection between environmental pollution and asthma diagnoses at the neighborhood level, requiring granular geographic data	Census-tract geographic data (not just 5-digit ZIP)	None
Reseacher	Link HPD data to additional data sets using date of birth as an identifier (as a prerequisite to additional analysis)	Date of birth (not just age)	None
Paye	Develop new, cost-effective health insurance products by identifying and contracting with high value (low cost and high quality) hospitals/providers.	None	Allowed Amounts Hospital/Provider Identifiers
Multle	Hospital Reference Pricing/Medicare Benchmark Analysis - requires data for named hospitals by payer type.	None	Allowed Amounts Hospital/Provider Identifiers
Start up compy	Use payment amounts to support analytics identifying opportunities for self-funded purchasers to reduce their costs	None	Allowed Amounts, Payer and Provider Identifiers

2 Standard Limited Datasets:

1. Commercial / Medicare Advantage

2. Add Medi-Cal (requires DHCS approval)

Both standard limited datasets include:

- Complete, record-level data on diagnoses, procedures, medical/pharmacy services
- Complete information on allowed amounts and other financial data
- Payer and provider data (codes replace names and other identifiers)
- No direct patient identifiers (e.g., name, medical record number)
- Limited indirect patient identifiers (e.g., ZIP, age)

What are your thoughts on HCAI's proposed approach to HPD Standard Limited Datasets?

Public Comment