Agenda Item VI: Data Access

Data Release Program Launch

Christopher Krawczyk, Chief Analytics Officer, HCAI

Data release program earliest launch: end of Q2 2024

- Regulations have gone out for a third round of public comment, extending the launch date
- HPD is using the time to:
  - Hold a closed beta test of the enclave with synthetic data
  - Improve the data documentation resources, including:
    - Data dictionary and completeness report
    - How-to documentation for using the Enclave and the HPD dataset
  - Iterate upon and refine the data request form
  - Hire and train staff to review requests and support data access
  - Collaborate and coordinate with DHCS, CPHS, and the HPD Data Release Committee on the joint review and approval processes

Data Release Ramping Up

- HPD is a new program with new processes; data release will start slowly
  - The review process will speed up as:
    - Processes are refined
    - Precedents are set
    - The teams gain experience
  - Reviewing HPD requests will also be a new experience for DHCS and CPHS
    - May need DUA adjustments after launch
    - May need data release process refinement after experiencing it in action, to better align with the cadence of external reviewers
- Review time will vary based on the complexity of each request

Data Release Program Components

Christopher Krawczyk, Chief Analytics Officer, HCAI

Jasmine Neeley, Research Data Analyst II, HCAI

Standard Limited Datasets:

1. Commercial / Medicare Advantage
2. Add Medi-Cal (requires DHCS approval)

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)
- Data Enclave or Direct Transmission

Standard limited datasets+ include:

- Payer and provider data (actual values, not codes)
- Only accessible via Data Enclave
- Application review criteria under development

The Standard Limited Datasets

- Version 1 of the Standard Limited Datasets (SLDs) is in production and testing
- The SLDs will be available as commercial-only datasets (including Medicare Advantage) or commercial-plus-DHCS datasets
  - If a requestor wants only DHCS data, they will go through DHCS' established data request process
- Requests for SLDs will be the fastest for HCAI to approve, and will have the lowest price

Details

- **The "Standard Limited Dataset (SLD)" has all direct identifiers removed for patients, providers, and payers**
- **The "SLD Plus" includes identifiers for providers and payers**
- **Neither SLD nor SLD Plus will include Medicare FFS data from CMS**

The HPD Enclave

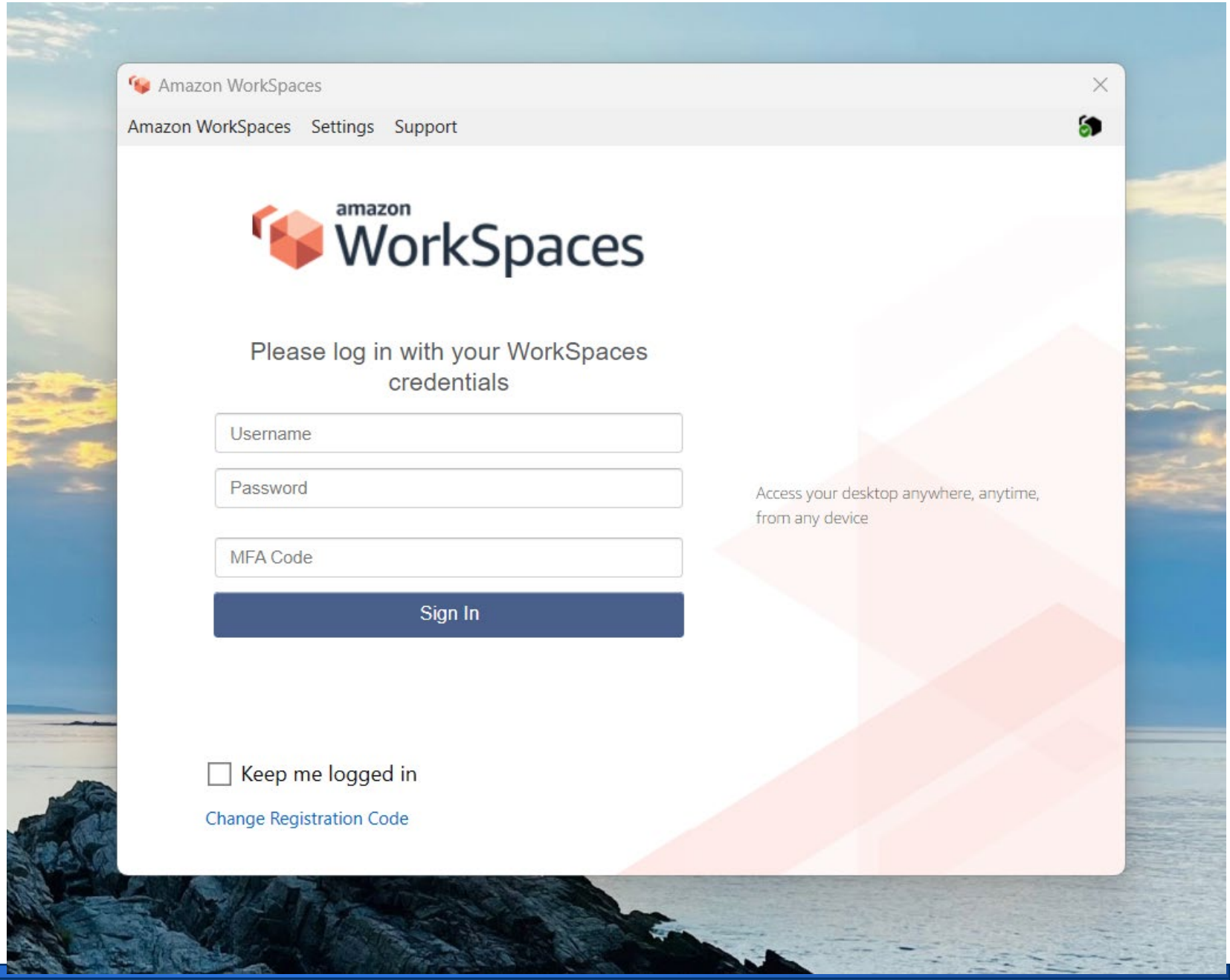
- Enclave development is complete!
- HCAI successfully tested all Enclave functions, including integration with the HPD data request process
- The HPD program is making use of the time between Enclave development and its active use to refine the Enclave documentation and user experience

Details

- **The Enclave is a secure platform where users can remotely access and analyze record-level HPD datasets**
- **Importing and exporting files will require HCAI review and approval**
- **Analytic tools – plus customer service and support from HPD experts – are included in the Enclave access price**

Logging In

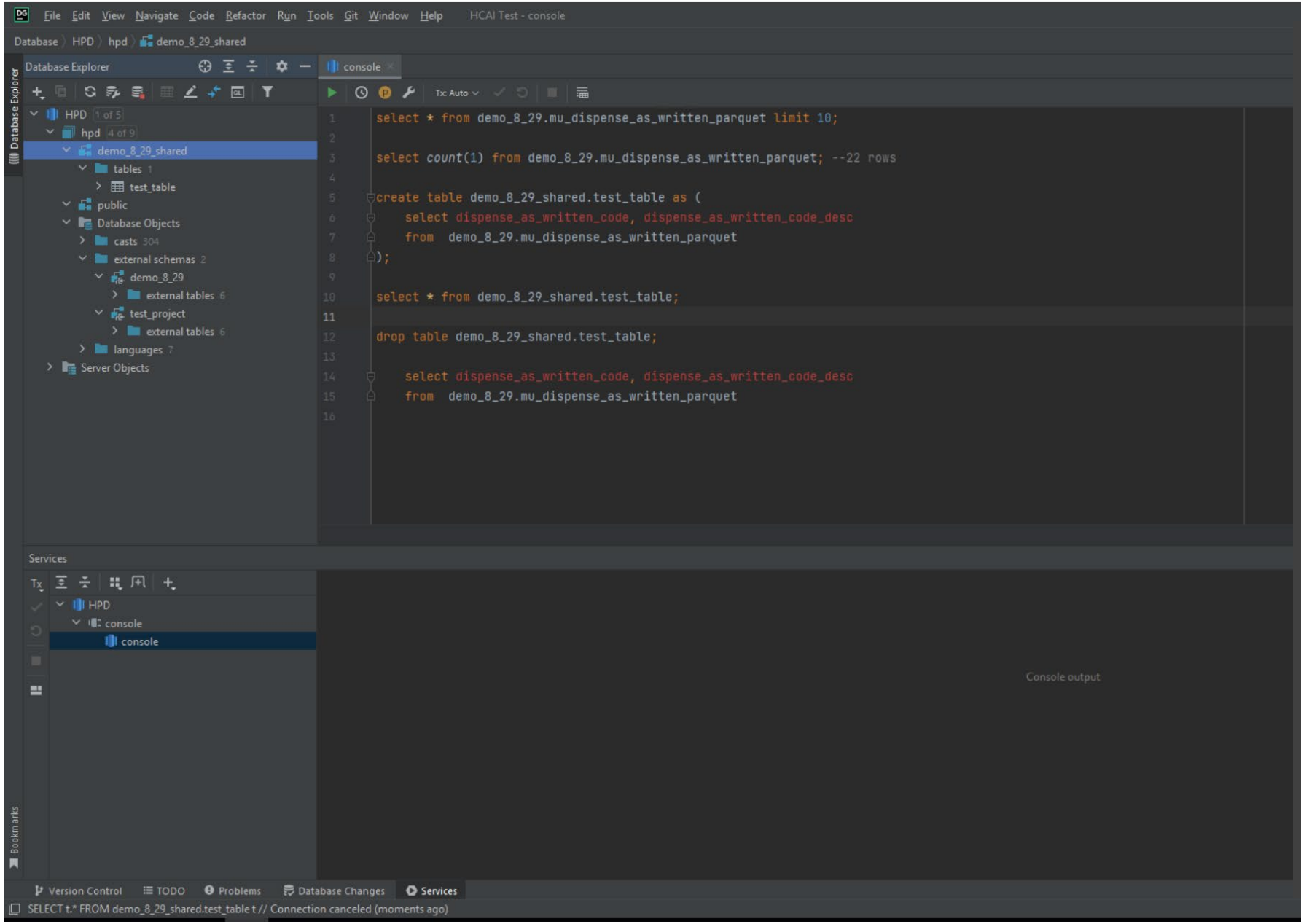
This shows the user logging into the virtual machine they will use to access the HPD Enclave.



Virtual Machine

This shows the desktop of the virtual machine on which they will use tools like DataGrip, R, Tableau, and SAS Viya.

Note: Browser applications in the Enclave cannot connect to the internet.



Writing Code

This is an example of DataGrip, where users can write SQL to analyze the data. Data sources are on the left. Queries are written in the top-center box.

The HPD Data Request Form

- Data request forms are in penultimate version, pending final regulation approval.
- Final request form and workflow functionality are on-track to be tested and deployed in time for the data release program launch.
- The HPD program is making use of the time between form implementation and its active use to refine the documentation and user interface.

Details

- **The data request form is an online form available through the HPD website**
- **The request form is connected to the review workflow, so requestors can stay engaged with reviewers and monitor their request's status**
- **The request form is aligned with the data release regulations, and that will continue for any changes in the future**

Application Sections (high-level)

Overview

- Data Applicant information, qualifications, and prior experience
- Alignment of request with HPD purpose, goals, requirements
- Submit data to HPD program?
- Requesting a Price Reduction?
- New or Supplemental Request?

Project Specific

- Description of project purpose, goals, and data use/methods
- Requested HPD data product, timeframe of data, and desired delivery format
- List of data elements with justification/reason based on project purpose
- DHCS data requested?
- Proposed Linkages to other sources with justification/reason

Application Sections (high-level)

Data Security

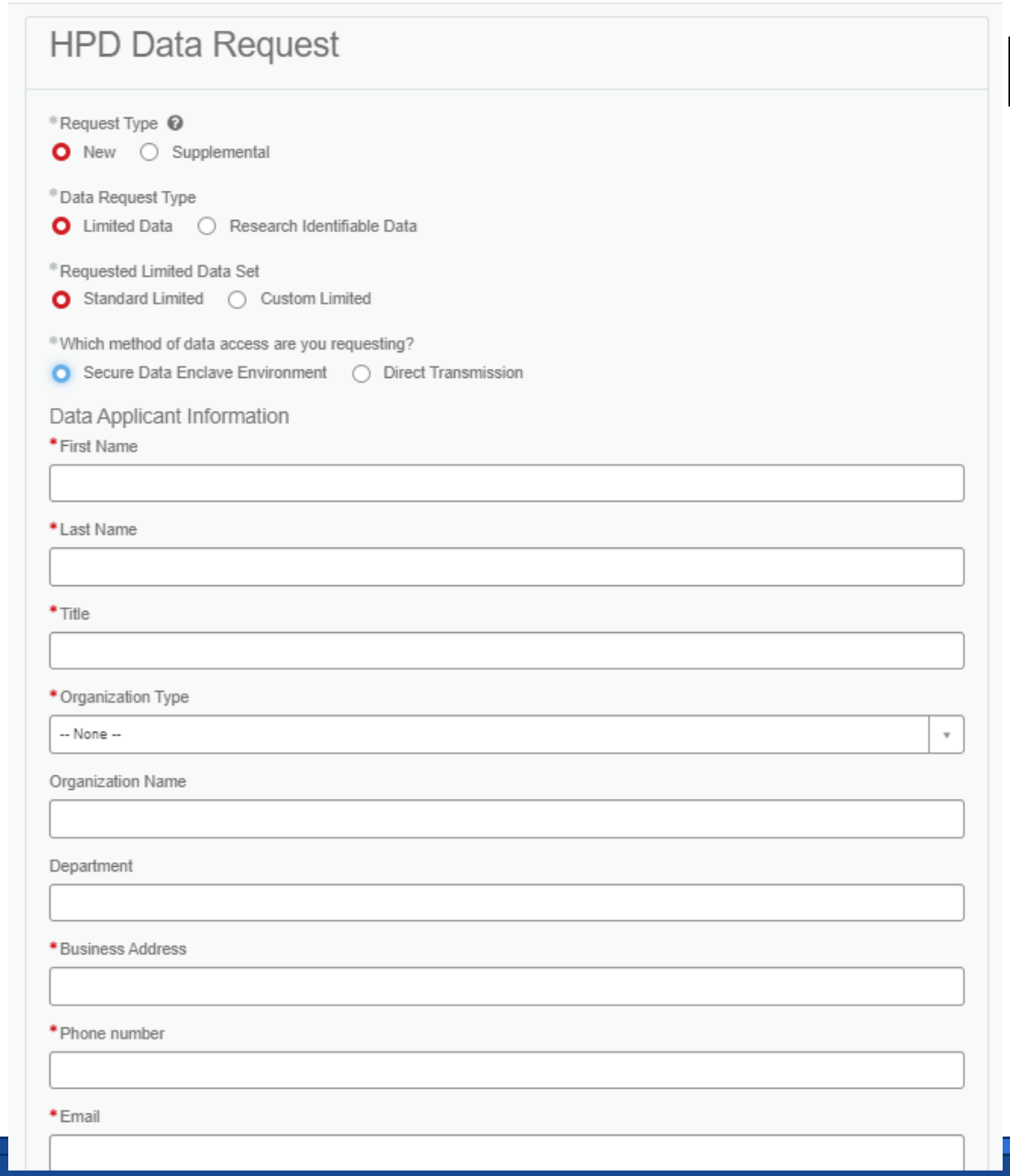
- Method of Data Access – Enclave or Direct Transmission
- Data User information – Contractors?
- Data Security Plan
- Description of how results and key findings will be shared
- Small cells

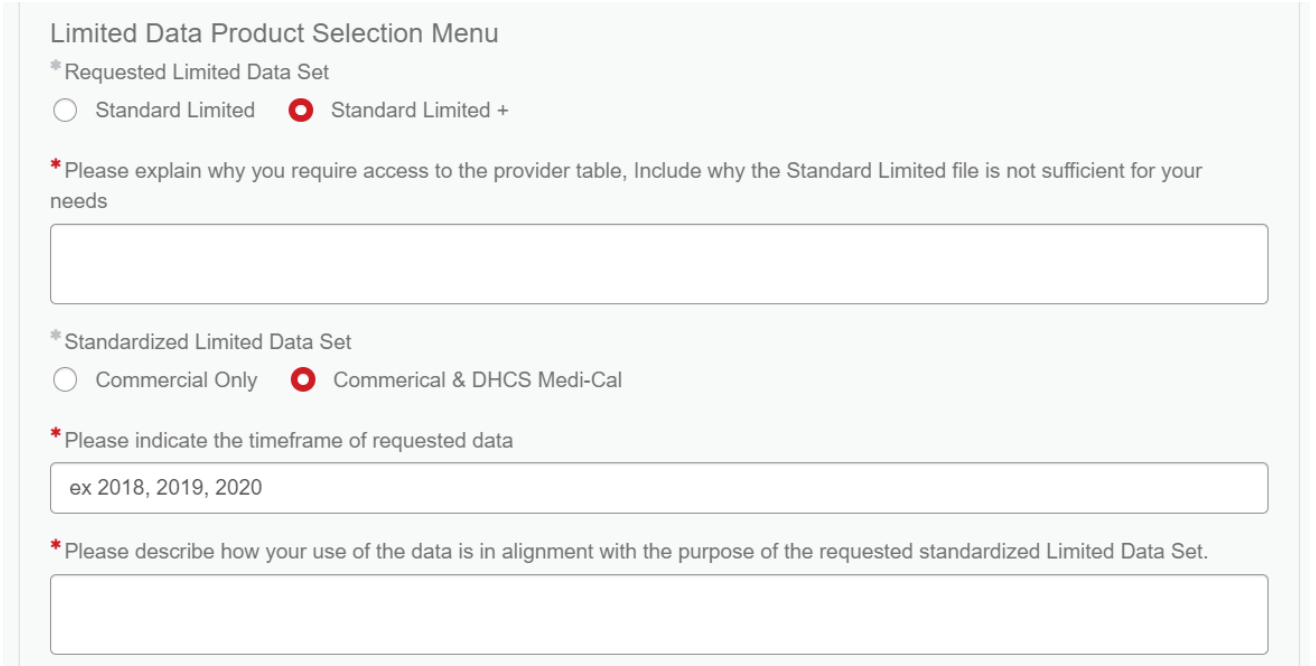
Additional Information

- Does the project offer significant opportunities to achieve HPD goals?
- All records vs. Subset of Records
- Minimum Data necessary
- Planned Data Products
- Applicant History, including with Breaches/Convictions/Civil Actions

HPD Limited Data Request Form

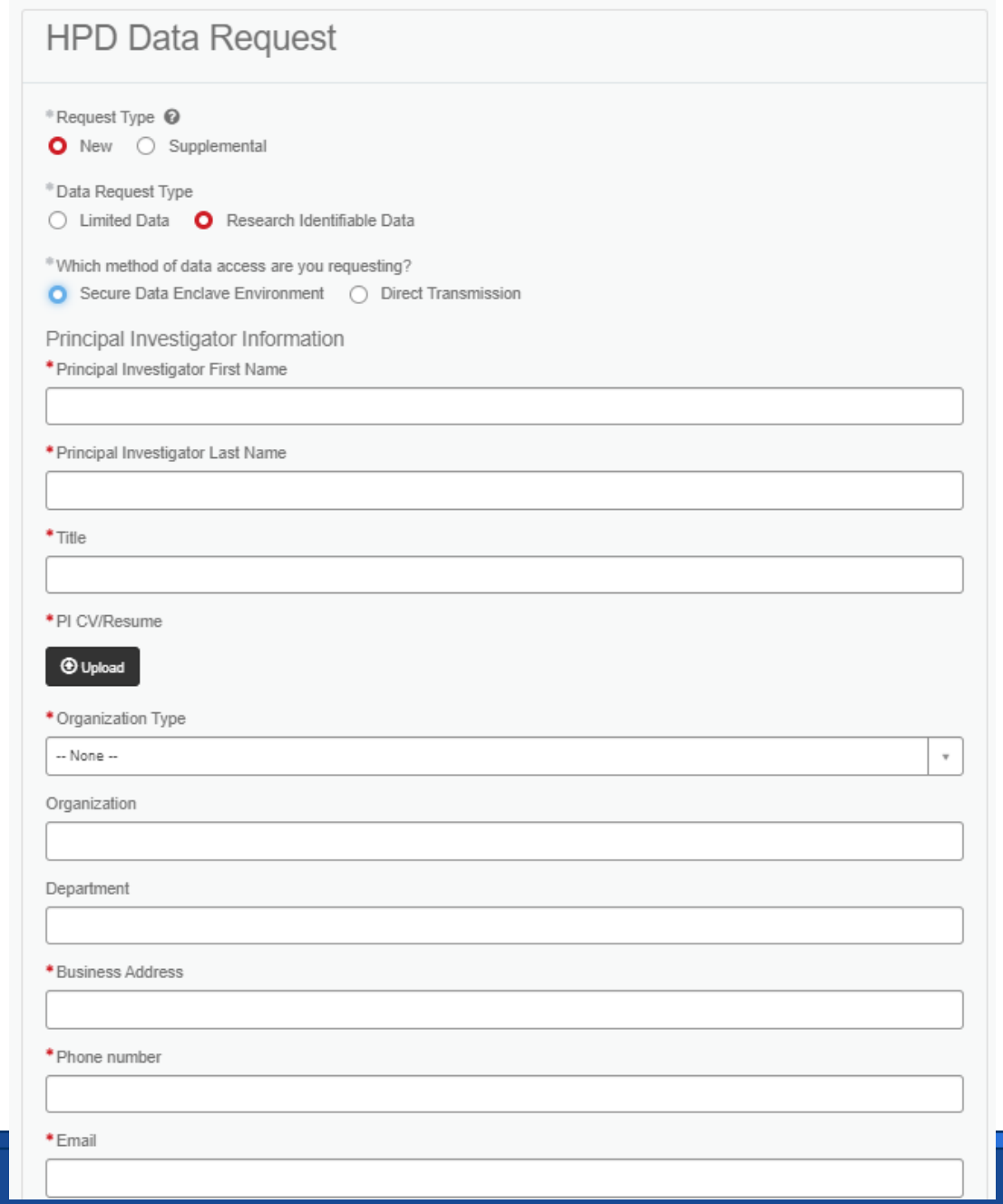
(external view)





HPD Research Data Request Form

(external view)



HPD Data Request

*Request Type
New Supplemental

*Data Request Type
Limited Data Research Identifiable Data

*Which method of data access are you requesting?
Secure Data Enclave Environment Direct Transmission

Principal Investigator Information

*Principal Investigator First Name

*Principal Investigator Last Name

*Title

*PI CV/Resume

Upload

*Organization Type

-- None --

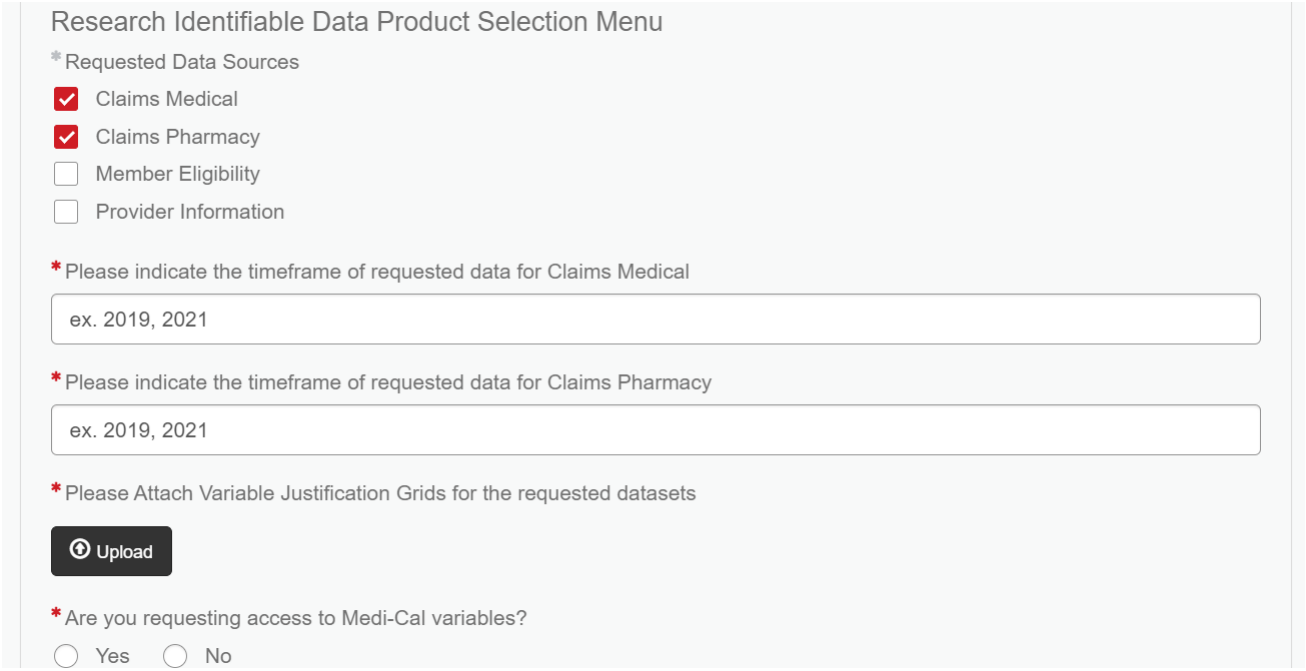
Organization

Department

*Business Address

*Phone number

*Email



Research Identifiable Data Product Selection Menu

*Requested Data Sources

Claims Medical
Claims Pharmacy
Member Eligibility
Provider Information

*Please indicate the timeframe of requested data for Claims Medical

ex. 2019, 2021

*Please indicate the timeframe of requested data for Claims Pharmacy

ex. 2019, 2021

*Please Attach Variable Justification Grids for the requested datasets

Upload

*Are you requesting access to Medi-Cal variables?

Yes No

HPD Data Request Form

*Describe how use of the requested data aligns with the goals of the HPD Program. Identify the specific goals your request aligns with below.

*HPD Program Goals

- [ ] Provide public benefit for Californians and the state while protecting individual privacy.
- [ ] Increase transparency about health care costs, utilization, quality, and equity.
- [ ] 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.
- [ ] 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.
- [ ] Support a sustainable health care system and more equitable access to affordable and quality health care for all.

Data Access

*Please list all individuals who are anticipated to observe, use, or access confidential data, including yourself

Add | Remove All

Actions	Data User Name	Data User Organization	Data User Email Address	Data User Phone Number	Dat
					No data to displa

*Will a contractor be used for this project?

○ Yes ○ No

Project Security

*Have you previously requested HPD data?

○ Yes ○ No

*Has the data applicant been involved in a data breach, or other similar issue, in which Personally Identifiable Information (PII) and/or Protected Health Information (PHI) was misused or improperly disclosed, within the last 7 years?

○ Yes ○ No

*Has the data applicant been involved in any criminal convictions or substatiated violations of California, or other state, or federal laws regarding fraud, theft, data breach, data misuse or related offenses in the past 7 years? This includes civil or admistrative penalties, civil judgements, or disciplinary actions.

○ Yes ○ No

*Please describe your experience with data security and the protection of large sets of confidential information.

Additional Documents

During Review

- Variable Justification Grids
- Draft CPHS Protocol
- Data Requestor CV*
- Contract*

Final Approval

- CPHS Approval letter
- Approved CPHS Protocol
- Signed Data Use Agreement (DUA)
- Signed Request Form

HCAI Provided Documents

- Unsigned Data Use Agreement (DUA)
- Invoice

*If the application requires

Data Access Pricing

Christopher Krawczyk, Chief Analytics Officer, HCAI

Nitisha Patel, Research Data Specialist I, HCAI

Advisory Committee Feedback - Summary

Price Schedule

- Appropriate to have a user price schedule and application fees
- Important that user payments are not a primary funding source for HPD – focus should be on supporting access not maximizing revenue
- Supportive of connection between the cost of fulfilling data request and the price (more complex = higher price)
- User price schedule could be designed to help achieve program goals based on purpose, e.g., distinguish between companies that want to monetize HPD data and those working in the public interest to improve care

Price Reductions

- Purpose of the request is important to consider; greater public benefit = more worthy of price reduction
- Assessing purpose could rely on program goals or HCAI could identify priority topics (e.g., Oregon's identification of disparities research)
- Incorporating purpose as a criterion should not create undue burden for review staff
- Consider budget size of requester alongside other criteria
- Broaden eligibility for consumer organizations beyond those with DMHC designation
- Consider actionability of the project results – academic research can take years

DRC Committee Feedback - Summary

Price Schedule

- Noted concerns about the affordability of the pricing structure, suggesting there may be challenges for researchers at the current rates.
- Requests for additional clarity on approach to determine data costs for applicants
- Requests for clarity on definitions for students and academic fellows
- Requests for clarity on pricing for collaboration between consumer groups, nonprofits and academic institutions
- Taking into account the complexity of a request and making it easy to add additional variables or additional team members.

Price Reductions

- Discussed the need for clear criteria for eligibility for a price reduction and reduction amounts
- Use price reductions as a way to have more equity in access.
- Emphasized the importance of predictability in pricing for researchers seeking funding, particularly for graduate students and fellows.
- Identified limitations with first come, first serve reductions.
- Suggested prioritizing projects based on public health goals, funding availability for researchers, especially considering early-career investigators.

Purpose, Goal, and Result of Access Prices

- **Purpose:** to contribute to HPD Program sustainability through a user price schedule that balances program support and affordability of access.
- **Goal:** As demand for access to HPD data goes up, the number of projects the program can support increases.
- **Result:** The money HPD collects from users is added to the HPD Fund and will be used to help the data access program support future projects.

Developing HPD Data Access Prices

- Pricing information was gathered for comparison from:
  - Other states' APCDs
    - States Included: Arkansas, Colorado, Connecticut, Delaware, Maine, Maryland, Massachusetts, Oregon, Rhode Island, Utah, and Washington
      - The following information was collected from these states to establish the HPD price Schedule: types of data, application price, price reduction/waiver process, pricing information, number of requests accepted and processed per year, type of data access (Direct Transmission vs. Data Enclave)
    - States with an Enclave equivalent included Washington and Maine
  - CMS' data access program
    - Access through the Research Data Assistance Center
  - The schedule of costs HCAI pays to the HPD Enclave vendor

Developing HPD Data Access Prices Cont.

- HCAI developed assumptions and scenarios about things like:
  - The number and variety of data access requests that will arrive each year
  - The number of simultaneous projects HPD can support in the Enclave
  - How many data requests HCAI has the resources to process
    - HCAI is increasing staff resources in preparation for HPD Data Release
  - The amount of cost recovery needed to meet program goals and steadily grow the data release program
    - Permanent funding has yet to be established for the HPD program

Developing HPD Data Access Prices Cont.

- HCAI tested various pricing models
  - Created many price model projections to compared data prices over the next 1-5 years
  - Priority focused on sustainability of the program while maintaining the lowest costs
  - Prices were adjusted until they represented the most feasible option for applicants while sustaining the HPD program

Enclave Annual Pricing Schedule (accessing up to 10 years of data)

Components - choose space + seat(s)	Annual Price
Project space with SLD or SLD+	$5,000
Project space with custom/research dataset	$14,000
Analyst seat	$5,000
Researcher seat	$6,000

Additional Options (per seat)	Annual Price
SAS Viya License	$5,000
Stata/MP License	$1,100
Tableau Creator License	$930

Details

- Smallest possible request is $10,000 (an SLD project with 1 analyst seat and no optional licenses)
- A “researcher” seat in the Enclave has 4vCPU + 16GB, 500 GB storage, 10 TB/month of compute, and access to SQL, R, Python, and Office tools. An “analyst” seat has half that storage and compute capacity
- A “project” has a data source (standard or custom) with up to 10 years of data included, 500 GB of database storage, and 1 TB of S3 storage for imports, code, or exports

Direct Transmission Pricing Schedule

One plan year of the SLD	Price
Commercial (includes Medicare Advantage)	$20,000
Commercial + Medi-Cal	$22,000

Notes:

- Enclave pricing is incentivized and subsidized by the HPD program; direct transmission of data to a user's system is not.
- Medicare fee-for-service data is only available to California state agencies, per the CMS DUA.
- SLD+ is not available for direct transmission.

One plan year of a Custom Dataset	Price
Medical Claims	$8,500
Pharmacy Claims	$7,000
Dental Claims	$7,000
Member Eligibility	$4,000
Provider Information	$3,000
All Tables	$29,500

Example Pricing Scenarios

"Lone Researcher" Enclave $11,000

- One-year project
- 1 "researcher" Enclave seat
- Standard Limited Dataset
- 5+ plan years of data
- Workspace stats:
  - 4vCPU + 16GB
  - 500 GB storage
  - 10 TB/month compute

"Custom Team" Enclave $30,000

- One-year project
- 3 enclave user seats
  - 1 researcher
  - 2 analysts
- Custom Dataset
- 5+ plan years of data

"Direct Transmission" 5 years of data $110,000

- Standard Limited Dataset
- Sent to the user's system
  - More security requirements
  - More DUA components, including time limits and data destruction requirements
- Commercial (including Medicare Advantage) and Medi-Cal data

Enclave Pricing Comparisons

	CMS Centers for Medicare & Medicaid Services	Health Care Cost Institute	Washington State Health Care Authority	HCAi
Annual Cost	$15,000-$35,000 per project. +$20,000-$22,000 per user	$45,000 for a team of two; $15,000 for a student. +$10,000 each additional user	$100,000 for Premium, $60,000 for Standard, $40,000 for Reduced. +$5,000 each additional user	$10,000+ per project. +$5,000-$6,000 each additional user. +$9,000 for custom dataset.
Tools	SAS, Python, R, Microsoft Suite	SAS, Python, Stata	SQL Workbench, R-Studio	Python, SQL, R-Studio, Microsoft Suite Additional cost for: SAS Viya, Stata/MP, Tableau Creator

Enclave Price Reductions

Wade Iuele, Consultant, HCAI

Price Reduction – Statute and Regulations

HPD Statute HSC 127674(f)(2):

"The department shall work with the advisory committee to make considerations for state agencies, data submitters, and consumer organizations…"

HPD Data Use, Access, and Release Regulations §97414:

"The Department may reduce program data prices…if it determines there is good cause for reduction, supported by documentation. Good cause includes, but is not limited to:

(1) the financial hardship of data applicants, such as students with needs-based financial aid working toward completion of required academic milestones, or government or nonprofit organizations whose funding sources for their projects do not cover data fees prices; or

(2) whether reduction will encourage the use of program data in high priority areas, or will lead to innovations that will benefit the public at large."

"Price reductions will be considered on a per project basis."

"Price reduction requests will be considered in the order received until available funds for price reductions are exhausted or price reductions are no longer compatible with program sustainability."

Starting Place - Assumptions

1. An annual price reduction program is contingent on long-term funding for the HPD program.
2. A finite pool of funds will be available to HCAI each year to pay for data access projects.
3. HPD project costs will vary substantially, so a fixed discount percentage regardless of project price may not be compatible with program sustainability.
4. Price reduction requests must be considered on a per-project basis, in the order received, as stated in the regulations.
5. The price reduction program can be refined over time as HCAI gains experience with the level of demand and requests for reductions.

Recap: Feedback on Price Reductions

Advisory Committee:

- Purpose of the request is important to consider; greater public benefit = more worthy of price reduction
- Assessing purpose could rely on program goals or HCAI could identify priority topics (e.g., Oregon's identification of disparities research)
- Incorporating purpose as a criterion should not create undue burden for review staff
- Consider budget size of requester alongside other criteria
- Broaden eligibility for consumer organizations beyond those with DMHC designation
- Consider actionability of the project results – academic research can take years

Data Release Committee:

- Discussed the need for clear criteria for eligibility for a price reduction and reduction amounts
- Use price reductions as a way to have more equity in access.
- Emphasized the importance of predictability in pricing for researchers seeking funding, particularly for graduate students and fellows.
- Identified limitations with first come, first serve reductions.
- Suggested prioritizing projects based on public health goals, funding availability for researchers, especially considering early-career investigators.

Consideration #1: Managing Uncertainty

Decisions about price reductions must be made on a case-by-case basis in the order received. While the HPD prices are publicly available, applicants who request a price reduction will not know in advance:

a) if they will be granted one, or
b) how large the reduction will be

Discussion:

- How can HPD help grant writers and other applicants manage the uncertainty of the price reduction program?

Consideration #2: Balancing Affordability and Sustainability

HPD Enclave access prices have been set as low as possible while maintaining program sustainability. While HPD prices are competitive, the cost to access HPD data is high enough to create a challenge for some applicants, which will reduce initial demand and limit the variety of HPD use cases.

Discussion:

- What guidance do you have for HCAI on striking a balance between affordability of data access and sustainability of the program?

Consideration #3: Direct Transmission Price

The price reduction program only applies to Enclave access; direct transmission prices are not planned to be reduced.

HPD's data access program takes an Enclave-first approach to best control and protect the data. Direct transmission prices are set above-cost, and not incentivized with price reductions, to help mitigate the privacy and security risks direct transmission adds to the program.

Discussion:

- Tradeoffs around adding direct transmission requests to the price reduction program, and prioritizing these factors: sustainability, data release program expansion, privacy/security, statutory requirements to promote enclave use.

Consideration #4: Length of Review Time

In the first year of the data access program, multiple unknowns will affect application review times, including:

- How many data access applications will HPD receive in the first month, quarter, and year?
- How complex will the applications be? (custom datasets? direct identifiers? DHCS data?)
- How many applications will require the evaluation of a price reduction request?

After a couple of years of experience, HCAI will be in a better position to predict and manage the flow of applications and the review process timeline.

Discussion:

- Given these multiple unknowns, would you prioritize a more streamlined and predictable review timeline or lower prices and more reduction opportunities?

Could Below-Cost Pricing in Year One Address the Considerations?

- What if HPD had 50% lower prices for all year-one applications, ensuring the widest possible user pool in the first year?
- Further price reductions could be requested, and would consider:
  - Documented cause for reduction
  - The purpose of the request
  - The total cost of the request
  - The remaining funds available in the price reduction pool

Components: choose space + seat(s)	50% Price First year only
Project space with SLD or SLD+	$2,500
Project space with custom/research dataset	$7,000
Analyst seat	$2,500
Researcher seat	$3,000

Additional Options (per seat)	50% Price First year only
SAS Viya License	$2,500
Stata/MP License	$550
Tableau Creator License	$465

Discussion: Would below-cost pricing in Year One address the considerations in a meaningful way?

Additional Thoughts?

Advisory Committee feedback on the HPD price schedule and price reduction program is welcomed.

Enclave: requires space + seat(s)	Annual Price
Project space with SLD or SLD+	$5,000
Project space with custom/research dataset	$14,000
Analyst seat	$5,000
Researcher seat	$6,000

Enclave options (per seat)	Annual Price
SAS Viya License	$5,000
Stata/MP License	$1,100
Tableau Creator License	$930

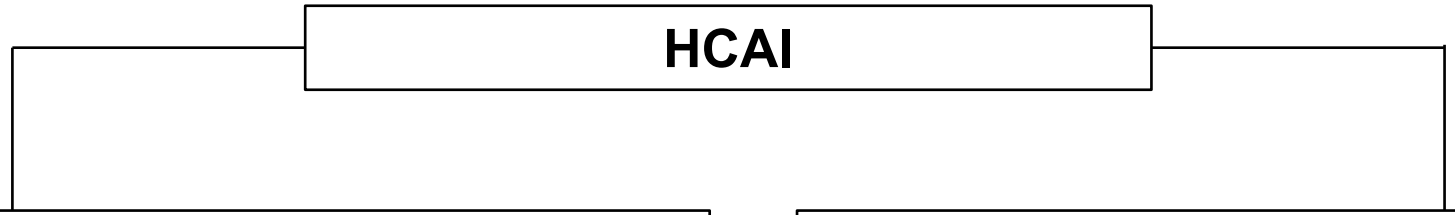
Direct Transmission: SLD	1 Year of Data
Commercial (includes Medicare Advantage)	$20,000
Commercial + Medi-Cal	$22,000

Direct Transmission: Custom	1 Year of Data
Medical Claims	$8,500
Pharmacy Claims	$7,000
Dental Claims	$7,000
Member Eligibility	$4,000
Provider Information	$3,000
All Tables	$29,500

Data Release Committee Updates

Nuriel Moghavem, HPD DRC Chair

Roles: Advisory Committee, Data Release Committee



Advisory Committee

- Assist and advise HCAI director in formulating program policies regarding data collection, management, use and access, and development of public information to meet the goals of the program
- Provide a forum for stakeholder and public engagement
- Provide input on priorities for the public information portfolio
- **Advise on criteria, policies, and procedures for access to and release of HPD data**
- Provide input on establishing the user fee schedule and fee waivers

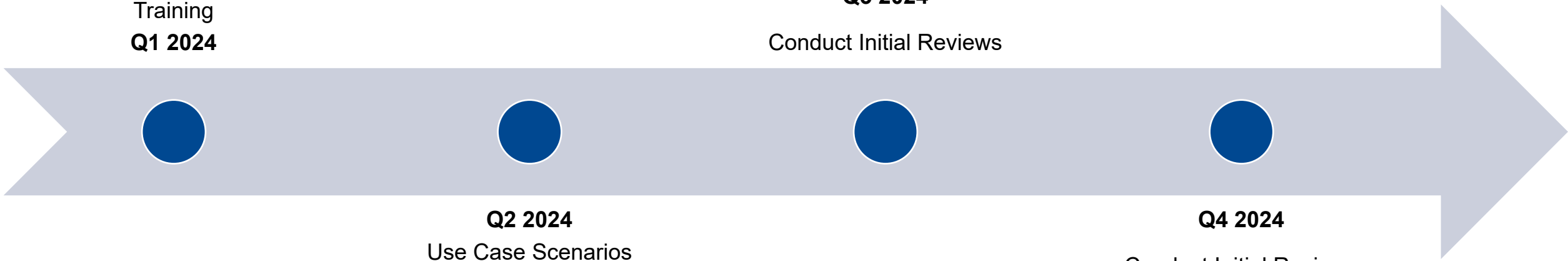
Data Release Committee

- **Advise on criteria, policies, and procedures for access to and release of HPD data**
- Make recommendations about applications for HPD data
- Consider whether the use of HPD data is consistent with the goals of the program, whether it provides greater transparency regarding health care costs, utilization, quality, or equity, or how the information may be used to inform policy decisions regarding the provision of quality health care, improving public health, reducing health disparities, advancing health coverage, or reducing health care costs
- Advise the HCAI director about privacy and security matters related to the program
- Provide feedback on the program's data application review processes and other matters

Source: Committee roles described in HSC Section 127671-127674

Previous HPD DRC Meeting Topics

Anticipated 2024 DRC Topics



Q1 2024

Pricing Schedule

Adopt DRC Manual

Use Case Scenarios

Training

Q2 2024

Use Case Scenarios

Conduct Initial Reviews

Q3 2024

Conduct Initial Reviews

Q4 2024

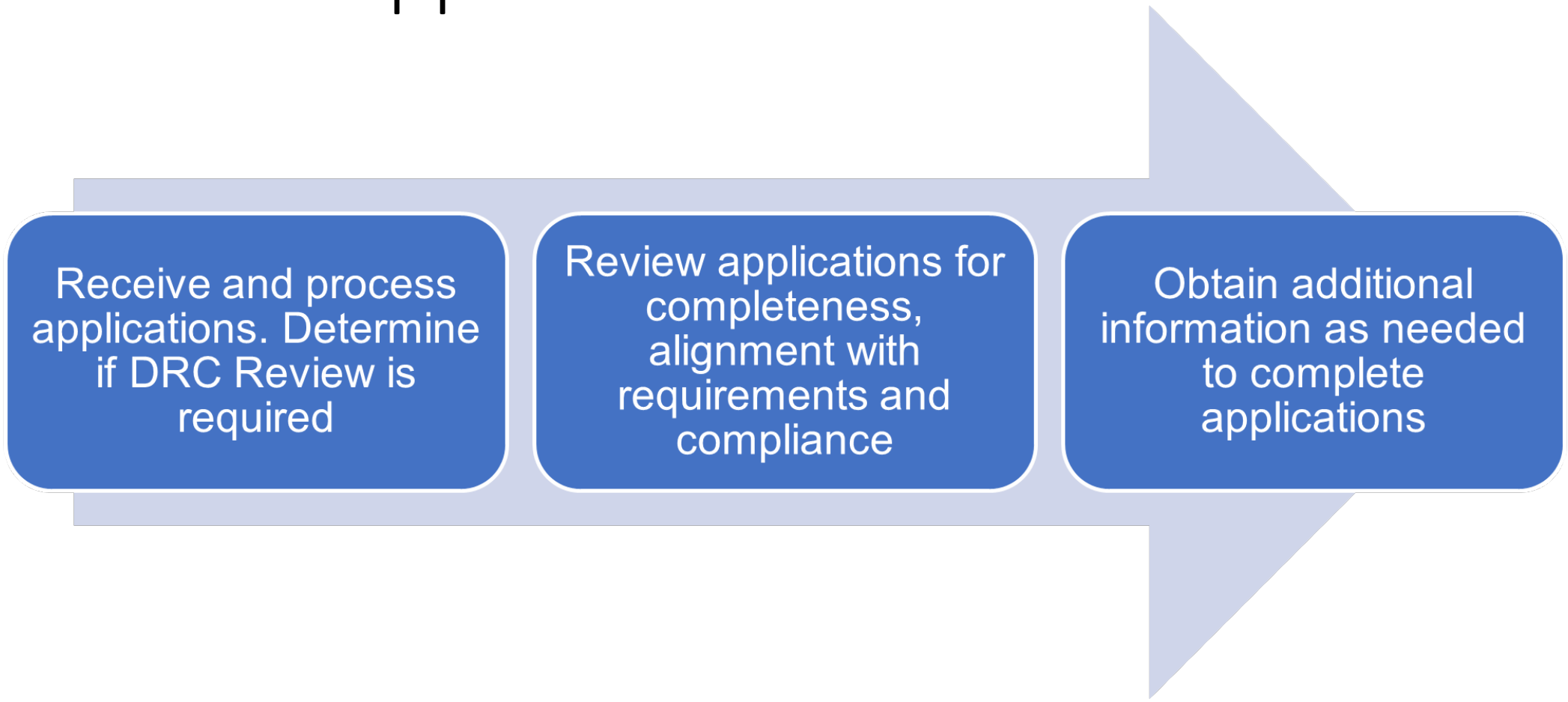
Conduct Initial Reviews

Conduct HPD DRC Evaluation

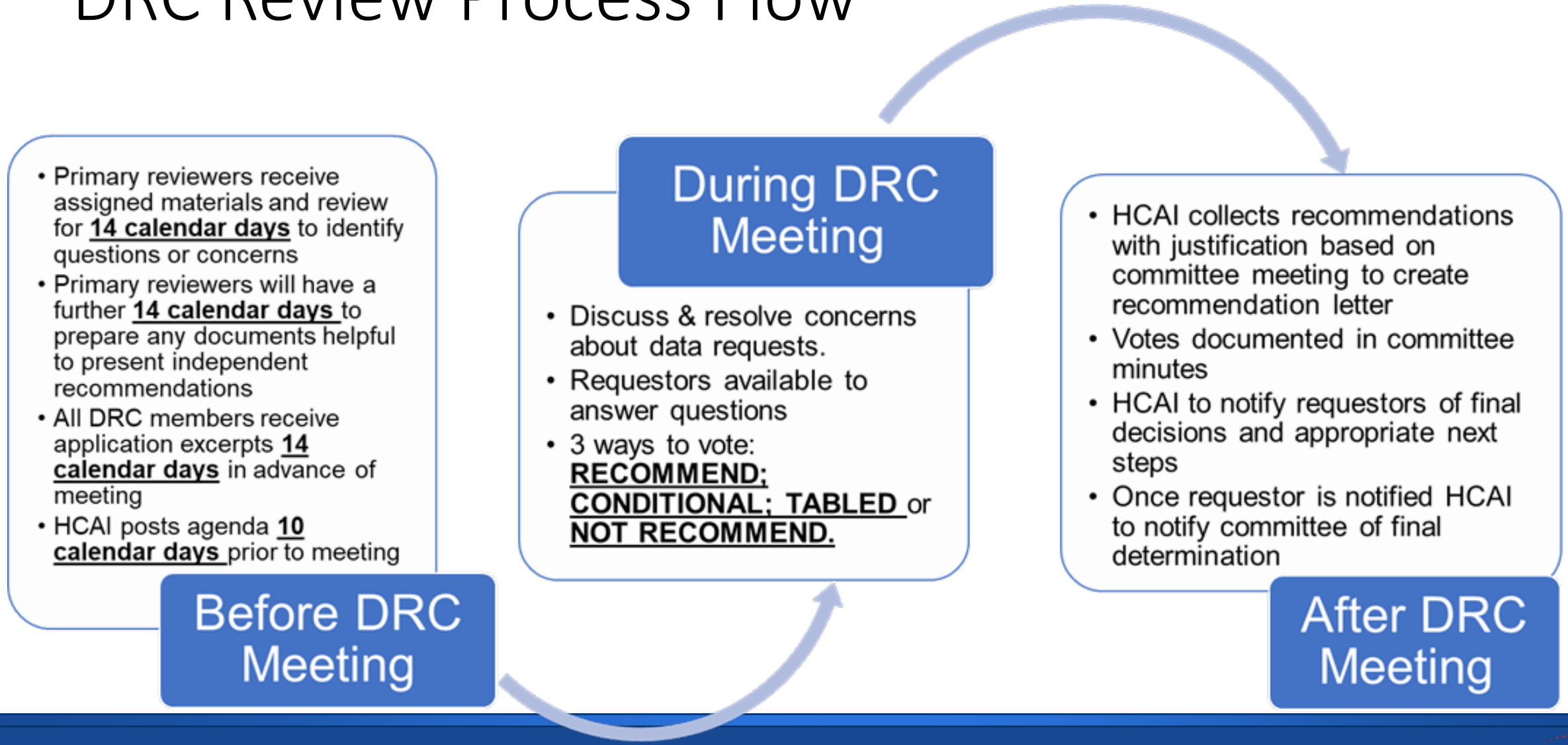
Proposed Next DRC Meeting Topics

- Review of Previous DRC Discussions
- Use Case Scenarios
- Application Excerpts
- DRC Considerations for Review

Initial HCAI Application Review



DRC Review Process Flow



Additional DRC Considerations

In addition to statutory and regulatory requirements, the DRC may consider other factors in the review of applications related to their experiences as health care payers, providers, purchasers, researchers, consumers, and other stakeholders.

The DRC may also consider whether the proposed use of the data is “consistent” with the goals; the requestor’s ability to balance the possible anticompetitive and/or commercial risks; and for requests that include direct personal identifiers, whether there are “significant opportunities to achieve Program Goals.”

Additional DRC Considerations (cont.)

Privacy

- What are the reasons or justifications provided for all confidential or sensitive data elements requested? These include direct identifiers, indirect identifiers (dates, 5-digit ZIP), and sensitive diagnoses and procedures.
- How does the request meet minimum data necessary standards?

Additional DRC Considerations (cont.)

Results and Information Products

- What are the expected findings/outputs? Will they be made available to policymakers or the public? What level of aggregation will be applied to reports, presentations, and other information products intended for public dissemination?
- How does the proposed project generate a public benefit for California?

Anticompetitive Use & Commercialization of HPD Data

- What are the risks for commercialization or marketing created by this request? And what is the state's involvement if any?
- What are the risks for anticompetitive use of the data?

Proposed DRC Recommendation Options

Four recommendation options are available to the DRC:

- **Recommended:** The committee votes to recommend approval of the application. If recommended, HCAI will make a final determination and provide information on data user fees, the method of data access (Enclave, direct transmission), and the Data Use Agreement (DUA) with the written notification. However, even if the committee recommends approval, HCAI may still deny the application.
- **Conditional:** The committee can recommend conditional approval for applications that require additional information to address minor issues or deficiencies. In these cases, conditional approval letters listing the issues and deficiencies will be provided to the requestor and HCAI will work with the requestor to modify the application or obtain additional information to address the DRC's concerns. These applications will not need to return to the DRC for review.

Proposed DRC Recommendation Options

- **Tabled:** The committee can recommend tabling applications that require substantive changes or if a recommendation does not receive a majority vote. In these cases, tabled letters listing the substantive changes will be provided to the requestor and HCAI will work with the requestor to modify the application to address the DRC's concerns prior to the application returning to the primary reviewers and subsequently the committee.
- **Not Recommended:** The committee does not recommend approval of the application. In these cases, denial letters listing the reasons for denial will be provided to the requestor.

There is no appeal process for recommendations. HCAI will make a final determination on approval on all applications and shall notify the DRC of its final determinations.

2024 HPD DRC Committee Meetings

- Meetings occurring monthly on the third Wednesday from 1 – 4 pm

Wednesday, June 19	Wednesday, July 17	Wednesday, August 21
Wednesday, September 18	Wednesday, October 16	Wednesday, November 20
	Wednesday, December 18	

Public Comment