

<p><b>PI name:</b> Vikas Dharnidharka, MD (Rutgers); David Axelrod, MD (Cleveland MC)</p> <p><b>Co-I (Data Coordinating Center):</b> Xing Song, PhD (MU)</p> <p><b>Working Title:</b> Choosing Immune Suppression in Renal Transplantation by Efficacy and Morbidity (CISTEM2)</p> <p><b>Sponsor:</b> NIH/NIDDK</p> <p><b>Performance Period:</b> 2024 - 2029</p>
---

## Study Overview

Recognizing the data continuum gaps in CISTEM [e.g., lack of longitudinal information on more granular clinical observations and outcome measures, such as serum creatinine levels (for computation of estimated glomerular filtration rate, eGFR), tacrolimus drug levels, measures of viremia and viuria, biopsy reports (for defining acute rejection (AR) and other forms of graft injury), malignancy diagnoses, rehospitalization events], we will establish a novel, robust and curated database (CISTEM2 database) integrating transplant registry data with multi-site electronic medical records (EMRs), administrative claims, and social determinants of health data for renal transplant patients leveraging the PCORnet infrastructure.

## Site Scope of Work

The GPC Coordinating Center (GCC) located at University of Missouri will operate as the Data Coordinating Center (DCC).

### *Contracting and Regulatory*

#### **1. Institutional Review Board**

Given that the nature of the project is using Datavant technology to generate de-identified hash token which will be shared with DCC and there is no way for DCC to re-identify the individuals, this project has been determined as IRB exempt by the leading site. As a result, each participating site will obtain separate IRB (or exempt IRB) independently to ensure compliance with ethical standards and regulatory requirements relevant to their institutional policies.

#### **2. Data Transfer and Use Agreement**

A CISTEM2 study-specific Data Transfer and User Agreement (DTUA), in line with the master GPC Data Sharing Agreement (DSA), will be developed, reviewed and signed by interested study sites. Both Scientific Registry of Transplant Recipient (SRTR) registry and the approved study sites will forward hashes to DCC, where data linkage and integration will be performed. For overlapped patients, the honest broker at DCC will request GPC EHR data in CDM format following the GPC Data-as-a-Product (DaaP) model. Separate DUAs will be established between leading site and SRTR and DCC, as well as between DCC and participating sites. A new CMS DUA will be further established between the leading site and CMS for Medicare claims integration.

#### **3. Datavant License Agreements and Order Forms**

The Datavant corporation will serve as the Honest Broker for CISTEM2 SRTR-GPC linkage. All GPC sites have an existing Site License Agreement (SLA) in place with Datavant and already have access to the Datavant De-ID software. To supplement the extant Site License Agreement, a CISTEM2 study-specific Order Form will be provided by Datavant to request basic details about the SLA at your site, such as the license number. This is a NO COST, NO SIGNATURE FORM that must be returned to the GCC and the contact at Datavant regardless of who serves as the Honest Broker. There is no cost to your site to utilize Datavant software for the purposes of CISTEM2.

## Data Linkage and Integration Workflow

### 1. Utilization of Datavant Software

The Datavant software solution enables Privacy-Preserving-Record Linkage (PPRL) through the use of de-identified, encrypted tokens that can be linked across data sources. CISTEM2 sites with existing SLAs have already installed the Datavant De-ID software to generate multiple, encrypted identifiers (hash tokens) based on different permutations of personally identifiable information (PII) held within source systems. SRTR has also installed this software. The underlying PII is processed by the Datavant De-ID software, where it is irreversibly hashed (i.e., cannot regenerate the PII from the hash value) into a series of Master Tokens using the Datavant Master Seed. The key-based hashing process means that the same PII processed at one site will result in different hashes from those produced from another site. In order to match tokens from different sites, the site-specific Master Tokens are then transformed into Transit tokens that are specifically directed at a third-party site using a site-specific key. This means that if the transit tokens are accidentally sent to the wrong location, the transit tokens cannot be processed by the incorrect site. The receiving site can then transform the transit tokens from different sites into tokens that can be used to match the same patients across different sites. The same PII always generates the same set of Master Tokens, but it is never present in any output or log stream from the De-ID software. Only the site-specific encryption tokens are written to the output file.

### 2. Tokenization and CISTEM2 Cohort Identification

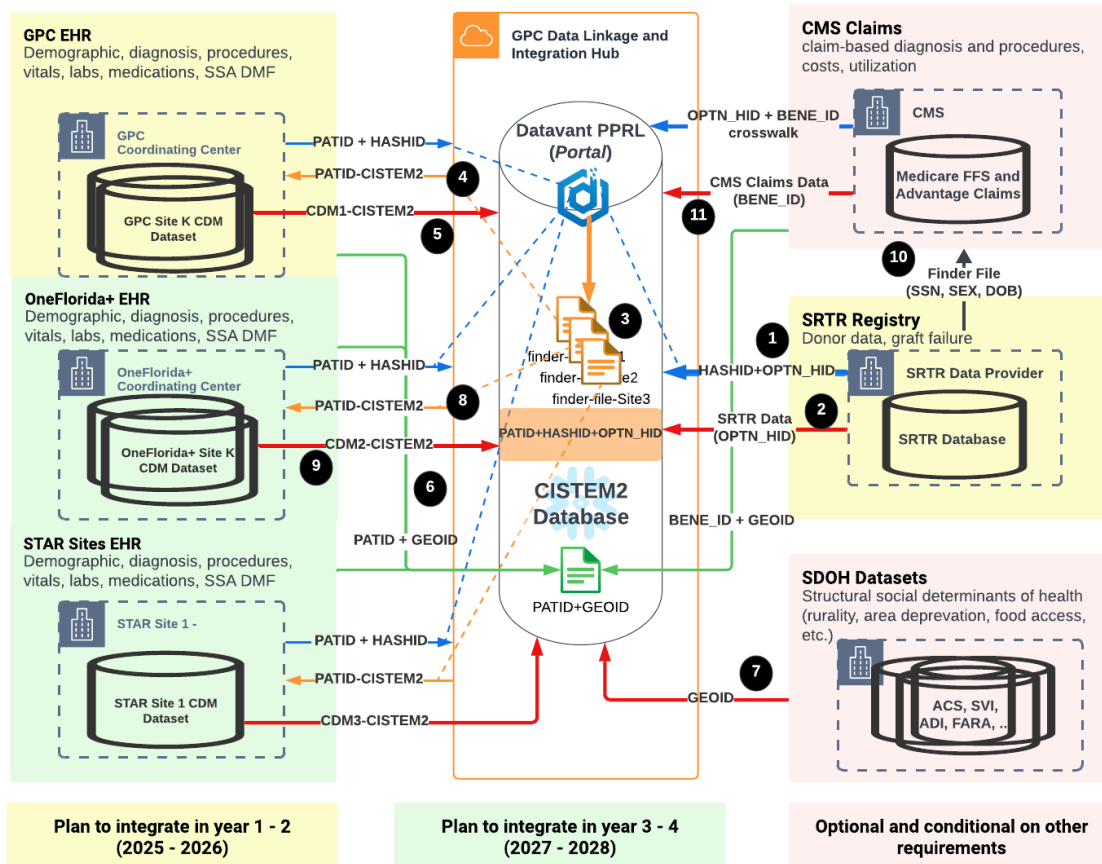


Figure 1. Federated Data Linkage and Integration Model

CISTEM2 cohort includes all kidney transplant (Kx) patients who are included in the SRTR registry with an established linkage between SRTR registry and USRDS database. As shown in Figure 1, DCC will first leverage Datavant Portal to identify the crosswalk population between GPC EHR cohort and SRTR registry to generate site-specific finder files (1). The finder files will then be disseminated to participating sites for clinical data extraction (i.e., CDM-CISTEM2) on the crosswalk Kx patients (3).

### **3. CDM Data Submission**

For patients identified in the distributed finder file (i.e., the crosswalk Kx patients), the data teams at each participating site will extract the following PCORnet CDM tables and submit to DCC:

- DEMOGRAPHIC
- ENCOUNTER
- ENROLLMENT
- CONDITION
- DIAGNOSIS
- PROCEDURES
- PRESCRIBING
- DISPENSING
- MED\_ADMIN
- VITAL
- LAB\_RESULT\_CM
- OBS\_CLIN
- OBS\_GEN
- DEATH
- PRO\_CM
- PROVIDER
- IMMUNIZATION
- LDS\_ADDRESS\_HISTORY
- HASH\_TOKEN – this table will include the mapping between PATID and HASHID generated from Datavant tokenization software.

As required in CDM v7.0, the following tables may also be populated and will be requested from sites if available:

- EXTERNAL\_MEDS
- PAT\_RELATIONSHIP

The CISTEM2 database will be established on the DCC cloud-based data infrastructure leveraging Amazon Web Services (AWS) services and Snowflake Data cloud technologies. DCC will utilize AWS S3 bucket or Snowflake secure share for data transfer and storage. DCC will create separate S3 bucket for each participating site and site-specific role with “least privileged” so that it can only access site-specific bucket. Each extracted CDM-CISTEM2 clinical dataset will be submitted to these site-specific AWS S3 buckets following the TLS/SSL secure protocol. Participating sites can also choose to adopt additional layer of data encryption following AES-256 with password protection.

### **4. Data Integration**

Upon receiving data from participating sites, SRTR, and CMS. GCC will perform data linkage and integration using the mapping among four source-specific, masked identifiers: PATID, HASH\_ID, OPTN\_HID, and BENE\_ID. HASH\_ID is the study-specific transit token generated by Datavant software.

### GPC Site Budget

Budgeted Items	Each Performing Site
1. Infrastructure Cost Recovery (ICR)	\$ 10,962
2. Administrative and Startup costs (including small PI oversight effort)	\$ 5,000
3. Hash token generation and linkage	\$ 7,800
4. Site CDM Data-as-a-Product	\$ 5,000
<b>Total</b>	<b>\$ 28,762</b>